U.PORTO

UNIVERSIDADE DO PORTO

3.º CICLO

D 2023

FMUP
2023

U.PORTO

Unveil *CDH1* regulatory network, beyond coding mutations in HDGC

CELINA SÃO JOSÉ

U.PORTO

FACULDADE DE MEDICINA
UNIVERSIDADE DO PORTO

Unveil *CDH1* regulatory network, beyond coding mutations in HDGC

CELINA SÃO JOSÉ
TESE DE DOUTORAMENTO APRESENTADA
À FACULDADE DE MEDICINA DA UNIVERSIDADE DO PORTO EM
BIOMEDICINA

Unveil *CDH1* regulatory network, beyond coding mutations in HDGC

Doctoral thesis

Doctoral programme in Biomedicine

Faculty of Medicine of the University of Porto

2023

Celina Beatriz Teixeira São José

Host institutions









Financial support

This study was supported by Fundação para a Ciência e Tecnologia (FCT) through a PhD grant (SFRH/BD/140796/2018, COVID/BD/152989/2022) financed by Programa Operacional and União Europeia.

Artigo 48º, § 3º - A Faculdade não responde pelas doutrinas expendidas na Dissertação.

(Regulamento da Faculdade de Medicina do Porto – Decreto-Lei nº 19337, de 29 de janeiro de 1931).

Supervisor

Doctor Carla Oliveira, PhD, Affiliated Professor

Instituto de Investigação e Inovação em Saúde da Universidade do Porto (i3S), Porto, Portugal

Instituto de Patologia e Imunologia Molecular da Universidade do Porto (IPATIMUP), Porto, Portugal

Faculdade de Medicina da Universidade do Porto (FMUP), Porto, Portugal

Co-supervisor

Doctor Stefan Mundlos, MD, PhD

Max Planck Institute for Molecular Genetics, RG Development & Disease, Berlin, Germany

Institute of Medical and Human Genetics, Charité Universitätsmedizin Berlin, Berlin, Germany

Berlin-Brandenburg Center for Regenerative Therapies (BCRT), Charité Universitätsmedizin Berlin, Berlin, Germany

Aos meus heróis,

“Insight must precede application.”

Max Planck

Published manuscripts

Under Art. 8 of Decree-Law nº 388/70, the following publications are part of this dissertation:

São José, C., Pereira, C., Ferreira, M., *et al.* 3D chromatin architecture re-wiring at the *CDH3*/*CDH1 loci* contributes to E-cadherin to P-cadherin expression switch in gastric cancer. (2023). *Biology*, 2, 803.

São José, C.*, Garcia-Pelaez, J.*, Ferreira, M., *et al*. (2023). Combined loss of *CDH1* and downstream regulatory sequences drive early-onset diffuse gastric cancer and increases penetrance of Hereditary Diffuse Gastric Cancer. *Gastric Cancer*.

*Equal contribution

Tedaldi, G., **São José, C.**, & Oliveira, C. (2023). Other Syndromes and Genes Associated with Gastric Cancer Predisposition. In *Hereditary Gastric and Breast Cancer Syndrome: CDH1: One Genotype with Multiple Phenotypes* (pp. 99-116). Cham: Springer International Publishing.

Garcia-Pelaez, J.*, Barbosa-Matos, R.*, **São José, C.***, *et al*. (2022). Gastric cancer genetic predisposition and clinical presentations: Established heritable causes and potential candidate genes. *European Journal of Medical Genetics*, 65(1), 104401.

*Equal contribution

Tedaldi, G., Molinari, C., **São José, C**., *et al*. (2021). Genetic and Epigenetic Alterations of *CDH1* Regulatory Regions in Hereditary and Sporadic Gastric Cancer. *Pharmaceuticals (Basel, Switzerland)*, 14(5), 457.

Manuscripts in preparation

São José, C., Ferreira, M., Pelaez-Garcia J., *et al*. Combined analysis of chromatin accessibility and genome-wide loss of function variants reveals missing heritability in cancer. *In preparation*.

São José, C.*, Barbosa-Matos, R.*, Lobo, S., *et al*. Combined disruption of two *CDH1* intronic enhancers outperform the effect of loss of function *CDH1* coding mutations and modulate lung metastases in mice. *In preparation*.

*Equal contribution

Publications in conference proceedings

Abstracts from the 55*th* European Society of Human Genetics (ESHG) Conference: Oral Presentations. (2023). *European journal of human genetics: EJHG*, 31(Suppl 1), 3–90.

Proceedings of the 26th Annual Meeting of the Portuguese Society of Human Genetics (SPGH - Sociedade Portuguesa de Genética Humana) Coimbra, 17-19 November 2022. (2023). *Medicine*, 102(13), e33154.

Proceedings of the 25th Annual Meeting of the Portuguese Society of Human Genetics (SPGH - Sociedade Portuguesa de Genética Humana). (2022). *Medicine*, 101(30), e29313.

Abstracts from the 53rd European Society of Human Genetics (ESHG) Conference: Oral Presentations. (2020). *European journal of human genetics: EJHG*, 28(Suppl 1), 1–140.

Thesis commitee

Presidente

Doutora Filipa Abreu Gomes de Carvalho

Professora Catedrática da Faculdade de Medicina da Universidade do Porto (Porto, Portugal)

Vogais

Doctor Elfride De Baere

Full Professor, Center for Medical Genetics, Ghent University Hospital (Ghent, Belgium)

Doctor Patrick Benusiglio

Professor Associado, Hôpital Pitié-Salpêtrière, Assistance Publique – Hôpitaux de Paris, Clinical Oncogenetics Unit, Sorbonne University (Paris, France)

Doutor Sérgio Manuel Madeira Jorge Castedo

Professor Associado da Faculdade de Ciências da Universidade do Porto (Porto, Portugal)

Doutora Ana Paula Soares Dias Ferreira

Professora Associada da Faculdade de Ciências da Universidade do Porto (Porto, Portugal)

Doutor José Bessa

Investigador Principal do Instituto de investigação e inovação em saúde (Porto, Portugal)

Doutora Carla Isabel Gonçalves de Oliveira

Professora Afiliada da Faculdade de Medicina da Universidade do Porto (Porto, Portugal)

Em cumprimento com o disposto no Decreto-Lei nº 388/70, declara que participou ativamente na recolha e estudo do material incluído e redigiu todos os trabalhos.

Esta Dissertação inclui resultados de trabalhos não publicados.

Agradecimentos/acknowledgments

Durante a realização desta dissertação contei com o apoio de várias pessoas, essenciais para a sua concretização.

Gostaria de expressar o meu agradecimento à Carla Oliveira, minha orientadora. Ao longo desta jornada ensinou-me a discutir, questionar e pensar cientificamente, e um pouco mais para levar na bagagem. Sinto-me privilegiada por ter uma orientadora que me desafia e me concede a liberdade de experimentar e crescer, tanto a nível profissional como pessoal.

À Silvana, Liliana e Marta quero agradecer todos os momentos dentro e fora do laboratório. Foi uma jornada muito facilitada convosco.

Ao José, Rita, Ana, Carla, Mafalda e Nelson, obrigada por facilitarem todo o trabalho que desenvolvemos. Obrigada por vossa contribuição inestimável.

Quero expressar meu agradecimento a todo o grupo ERiC, um enorme obrigada por me receberem tão bem, pela incrível entreajuda e pela dinâmica inspiradora.

Rita, a minha metade da laranja, obrigada por teres aparecido no momento certo. Obrigada por me ouvires e me mostrares que posso sempre superar-me, colação. A tua presença é fundamental para o meu crescimento.

I would like to extend my gratitude to my co-supervisor, Stefan Mundlos, for consistently welcoming me back to the MPI. Thank you for all the invaluable scientific insights.

To all the Development and Disease group, thank you for always making me feel at home and a valued member. A special thanks goes to Juliane and Fiona, who consistently went out of their way to assist me and engage in discussions about our work. I am also grateful to Silvia, Alicia, Gabi, Silvia, Dani, Jana, Josh, Konrad, Andreas, and Alessa for making my time in Berlin incredibly enjoyable. Your collective efforts have made a significant impact on my experience.

Gostaria de expressar a minha gratidão à Mafalda, Raquel, Vassalo, Ana Teixeira, Luis Meireles, Xana, Mafalda, Inês, Mariana, Ricky e Nuno. O vosso apoio, companheirismo e alegria são fundamentais para o meu sucesso.

Um especial agradecimento aos meus pais por tudo que fizeram por mim e principalmente por me incentivarem a atingir todos os meus objetivos. Obrigada por me ouvirem sempre explicar o meu trabalho e por serem sempre os meus guias. Aos meus avós quero agradecer todo o apoio e por me mostrarem sempre o melhor caminho. Ao meu herói, obrigada pelo orgulho, tenho-te como o melhor exemplo.

Ao João, com quem tive a sorte de me cruzar no início desta jornada, quero expressar meu profundo agradecimento. Obrigada por despertares o melhor que existe em mim e por seres um apoio incondicional e indispensável. És a minha inspiração.

Table of contents

THESIS OUTLINE ..... **III**

ABSTRACT ..... **V**

SUMÁRIO ..... **VII**

ACRONYMS AND ABBREVIATIONS ..... **XI**

CHAPTER 1 ..... **1**

GENERAL INTRODUCTION **1**

1. Genomic variation *2*
2. 3D genome organization *3*
3. Gastric cancer *11*
4. Genetic predisposition to gastric cancer *17*
5. References *32*

CHAPTER 2 ..... **2**

RATIONAL AND AIMS **2**

CHAPTER 3 ..... **6**

GERMLINE REGULATORY ELEMENTS **6**

REVIEW PAPER 1 7
ORIGINAL PAPER 1 22
ORIGINAL PAPER 2 48

CHAPTER 4 ..... **74**

SOMATIC REGULATORY ELEMENTS **74**

ORIGINAL PAPER 3 75
ORIGINAL PAPER 4 93
ORIGINAL PAPER 5 118

CHAPTER 5 ..... **137**

GENERAL DISCUSSION **137**

APPENDIX A **143**

APPENDIX B **162**

APPENDIX C **175**

APPENDIX D **188**

APPENDIX E **207**

Thesis outline

This thesis is composed of five chapters, portraying one book chapter, one review article and five original publications, three of them published.

Chapter one covers a general introduction with pivotal aspects for the development of this research work. The main players of this thesis are regulatory elements and their roles in disease. Thus, the current knowledge in chromatin architecture, gastric cancer and the landscape of its hereditary counterpart are discussed.

Chapter two introduces the general and specific aims in the scope of this research work.

Chapter three provides an insight into the hereditary component of gastric cancer, covering the genes whose genetic defects may cause hereditary gastric cancer. Zooming in on hereditary diffuse gastric cancer, regulatory elements were explored to solve the missing heritability in families and the high penetrance and aggressiveness of the disease.

Chapter four covers *CDH1* regulatory elements in sporadic gastric cancer and metastasis development.

At last, **chapter five** provides a general conclusion of the research work, considering the abovementioned chapters.

Abstract

Enhancer-promoter contacts occurring within the 3D chromatin folding are crucial for gene expression in time and space. If disrupted, regulatory landscapes may lead to gene misexpression and disease, a context where hereditary diffuse gastric cancer (HDGC) is no exception. Genetic defects in *CDH1* and *CTNNA1* genes, encoding epithelial adhesion molecules E-cadherin and α-E-catenin, respectively, predispose to HDGC. However, more than half of HDGC families remain genetically unsolved, even though *CDH1* mRNA monoallelic downregulation is observed in the germline. Altogether, these data suggest *CDH1* as the major gene triggering HDGC, pinpointing defects in regulatory elements controlling *CDH1* expression, as potentially causative in unsolved HDGC families.

Herein, we explored the *CDH1* regulatory network, its impact in *CDH1*/E-cadherin expression and at the genome-wide level, using HDGC and sporadic gastric cancer as disease models. For this, we 1) dissected the (non)*bonafide* genes whose genetic defects may cause hereditary gastric cancer, 2) disclose the *CDH1* regulatory network to solve the missing heritability in HDGC families, 3) unravelled the impact of germline defects in *CDH1* regulatory elements co-occurring with *CDH1* in HDGC high penetrance and aggressiveness, 4) evaluated potential enhancers of *CDH1* expression in cell lines and gastric cancer patients, 5) unveiled the role of intronic *CDH1* enhancers in E-cadherin expression and metastasis development in gastric cancer, and 6) unravelled the impact of *CDH1* upstream regulatory elements in E-cadherin to P-cadherin switch.

Given that most HDGC and all familial intestinal gastric cancer (FIGC) families remain genetically unexplained, several gastro-intestinal cancer syndromes-associated genes have been suggested as disease causative. Thus, in **review paper 1** we performed a search for individuals with family history compatible with gastric cancer associated syndromes, and bearing pathogenic or likely pathogenic variants in those genes. *BRCA2* and *PALB2* were the most frequently mutated candidate genes in individuals with family history compatible with FIGC and HDGC, respectively.

Original papers 1 and 2 were focused on solving the missing heritability in HDGC families, resourcing to whole genome/exome sequencing of HDGC-like families, respectively. Briefly, in **original paper 1** we found a *MLH1* coding deletion in a family with a *bias* for GC development, overlapping a potential normal stomach regulatory element, and with no apparent activity in normal colon. Further, we found two deletions overlapping non-coding regulatory elements, that downregulated both *CDH1* and E-cadherin expression. We would find that young-affected individuals from HDGC-like families present a germline

landscape involving downregulation of immune-associated pathways due to either deletions or truncating variants, that may impact genome-wide stomach regulatory elements and increase immune evasion. While the **original paper 2** focuses on the diagnostic odyssey of a HDGC family unsolved for over 20 years and presenting very early onset diffuse gastric cancer. The high penetrance, aggressiveness and early development of the disease may be influenced by a deletion encompassing the *CDH1* gene, a non-coding regulatory region identified in **original paper 1**, and the downstream gene *TANGO6*. This *CDH1-TANGO6* deletion led to a nearly two-fold *CDH1* mRNA expression downregulation and impacted adhesion- and immune-associated pathways, such as response to virus and interferon-α/-β.

The role of *CDH1* regulatory elements in gastric cancer were covered in **original papers 3, 4** and **5**. In **original paper 3,** we dissected the role of genetic variants and DNA methylation in *CDH1* promoter and its enhancers in gastric cancer cell lines, sporadic gastric cancers, HDGC-like patients and healthy controls. We found that genetic variants and DNA methylation patterns in *CDH1* promoter and enhancers correlate with *CDH1* mRNA expression, pinpointing 7 potential *CDH1* enhancers. **Original paper 4** focuses on the impact of deletions in *CDH1* regulatory elements in E-cadherin expression and at the genome-wide level, in comparison to a coding deletion. We found that within the *CDH1 locus,* combined perturbation of two hypomorphic regulatory elements (CREs) produce complete *CDH1/* E-cadherin expression loss, similar to that produced by a coding deletion, and 50% *CDH1* expression loss, if deleted independently. Edition of both CREs has a stronger impact in adhesion-associated pathways, that translates in reduced lung metastasis, when cells were injected in the mice tail vein, and increased survival in breast cancer patients from the METABRIC cohort. Altogether, our results show 1) that the deletion of two hypomorphic regulatory elements produce a similar impact in *CDH1* expression, as a coding deletion in *cis*. Yet, act as protectors against metastasis formation throught transcriptome and chromatin re-wiring, rather than the coding deletion. **Original paper 5** explores the *CDH1* regulatory network to unveil the role of *CDH1* regulatory elements in E-cadherin to P-cadherin switch. P-cadherin misregulation occurs in most gastric cancer tumours, from which nearly half display a E- to P-cadherin switch. *CDH1* knockout triggered *CDH1*/E-cadherin complete loss and *CDH3*/P-cadherin expression increase at the plasma membrane. This switch, likely rescuing adherens junctions, increased cell migration/proliferation, commonly observed in aggressive tumours. E- to P-cadherin switch accompanied increased *CDH1* promoter interactions with a *CDH3*-eQTL, absent in normal stomach and parental cells. Deletion of this *CDH3*-eQTL promotes *CDH3*/*CDH1* reduced expression.

From the clinical standpoint, we found novel germline causes of HDGC to improve diagnosis and clinical management in carriers. From the fundamental research standpoint, we disclosed novel tissue-specific *CDH1*-associated regulatory mechanisms.

Sumário

A comunicação enhancer-promotor que ocorrem devido à arquitetura da cromatina em 3D são cruciais para a expressão genética em determinado momento e tecido. Caso a maquinaria regulatória seja alterada, a expressão genética pode também alterar-se e conduzir ao desenvolvimento de doença, para as quais o cancro gástrico difuso hereditário (HDGC) não é exceção. Defeitos genéticos nos genes *CDH1* e *CTNNA1*, que codificam as moléculas de adesão epitelial caderina E e α catenina, respetivamente, predispões para o desenvolvimento de HDGC. No entanto, mais de metade das famílias permanecem geneticamente não resolvidas, apesar de ter expressão reduzida de apenas um alelo de mRNA de *CDH1* na linha germinativa. Ao todo, esses dados sugerem o *CDH1* como o principal gene desencadeador de HDGC, apontando defeitos em elementos reguladores que controlam a expressão de *CDH1* como causadores em famílias HDGC sem diagnóstico genético.

Ao longo da tese, exploramos a rede regulatória do gene *CDH1*, o seu impacto na expressão de *CDH1*/caderina E e ao nível do genoma, usando o HDGC e o cancro gástrico esporádico como modelos de doença. Para isso, 1) dissecámos os genes (não) bonafide, cujos defeitos genéticos podem causar cancro gástrico hereditário, 2) explorámos a rede regulatória do gene *CDH1* para resolver a herdabilidade em famílias HDGC, 3) desvendámos o impacto de defeitos germinativos em elementos regulatórios do gene *CDH1* coocorrendo com defeitos no gene *CDH1*, em famílias com penetrância e agressividade, 4) avaliámos potenciais enhancers da expressão de *CDH1* em linhas celulares e pacientes com cancro gástrico, 5) revelámos o papel de enhancers intrónicos do gene *CDH1* na expressão de caderina E e no desenvolvimento de metástases em cancro gástrico e 6) revelámos o impacto de elementos regulatórios do gene *CDH1* na transição de caderina E para caderina P.

Dado que a maioria das famílias HDGC e todas as famílias de cancro gástrico intestinal familiar (FIGC) permanecem geneticamente inexplicáveis, vários genes associados a síndromes de cancro gastrointestinal foram sugeridos como causadores da doença. Assim, no **artigo de revisão 1** realizámos uma pesquisa bibliográfica por indivíduos com história familiar compatível com síndromes associadas ao cancro gástrico e portadores de variantes patogénicas ou provavelmente patogénicas nos genes acima mencionados. Os genes BRCA2 e PALB2 foram os candidatos mais frequentemente mutados em indivíduos com história familiar compatível com FIGC e HDGC, respetivamente.

Os **artigos originais 1** e **2** focaram em resolver a herdabilidade em famílias HDGC, recorrendo a whole genome/exome sequencing de famílias HDGC, respetivamente. No **artigo original 1** encontrámos uma deleção codificante do gene *MLH1* numa família com viés para o desenvolvimento de cancro gástrico,

englobando um potencial elemento regulatório em estômago normal, mas ausente em cólon normal. Adicionalmente, encontrámos duas deleções, incluindo dois ou um elemento regulatório não codificante, respetivamente, que regula negativamente a expressão da caderina E. O perfil germinativo de indivíduos jovens afetados em famílias HDGC exibe regulação negativa de vias associadas ao sistema imunológico, como MHCI e imunidade mediada por células natural killers, que podem aumentar a evasão imunológica, e alteração de elementos regulatórios em estômago em todo o genoma. Por outro lado, o **artigo original 2** foca na odisseia diagnóstica de uma família HDGC não resolvida há mais de 20 anos e apresentando cancro gástrico difuso muito precoce. A alta penetrância, agressividade e desenvolvimento precoce da doença podem ser influenciados por uma deleção que engloba o gene *CDH1*, a região não codificante impactada no artigo original 1, e o gene TANGO6. A deleção *CDH1*-TANGO6 revelou uma regulação negativa da expressão de mRNA de *CDH1* quase duas vezes maior, assim como, maior impacto nas vias associadas à adesão e sistema imune, tal como a resposta a vírus e a interferon-α/-β.

O papel dos elementos regulatórios do gene *CDH1* e cancro gástrico foi abordado nos **artigos originais 3, 4** e **5**. No **artigo original 3**, dissecámos o papel das variantes genéticas e da metilação do ADN no promotor *CDH1* e seus enhancers em linhas celulares de cancro gástrico, cancros gástricos esporádicos, doentes com HDGC e controlos saudáveis. Descobrimos que as variantes genéticas e os padrões de metilação do ADN no promotor e enhancers do gene *CDH1* correlacionam-se com a expressão do mRNA de *CDH1*, identificando 7 potenciais enhancers do gene *CDH1*. O **artigo original 4** foca no impacto das deleções em elementos regulatórios do gene *CDH1* na expressão da caderina E e ao nível do genoma completo, em oposição a uma deleção codificante. No *locus* do gene *CDH1*, elementos regulatórios hipomórficos coocorrentes produzem regulação negativa de caderina E semelhante a uma deleção codificante e 50% de perda de expressão de *CDH1* se deletados independentemente. Ao nível do genoma completo, a deleção exónica afeta as vias associadas ao interferão. Contrariamente, a edição de ambos os elementos regulatórios tem um impacto maior nas vias associadas à adesão, que se traduz numa redução da formação de metástases em murganho e num aumento da sobrevivência em pacientes com cancro de mama. Sumariamente, demonstrámos dois elementos regulatórios hipomórficos que têm um impacto semelhante a uma deleção codificante em cis, no entanto, em trans atuam como protetores contra a formação de metástases, em oposição à deleção codificante. Depleção de *CDH1* desencadeou a perda completa de *CDH1*/E-caderina e aumento da expressão de *CDH3*/P-caderina na membrana plasmática (**artigo original 5**). Esta transição, provavelmente resgatando o complexo de adesão, aumentou a migração e proliferação celular, comumente observada em tumores agressivos. A transição de E para P-caderina acompanhou o aumento das interações do promotor do gene *CDH1* com um *CDH3*-eQTL, ausente em

estômago normal e nas células parentais. A deleção de *CDH3*-eQTL promove redução da expressão de *CDH3* e *CDH1*.

Do ponto de vista clínico, encontrámos novas causas germinativas de HDGC para melhorar o diagnóstico e a orientação clínica dos portadores. Do ponto de vista da investigação fundamental, revelámos novos mecanismos regulatórios associados ao gene *CDH1* específicos de cada tecido.

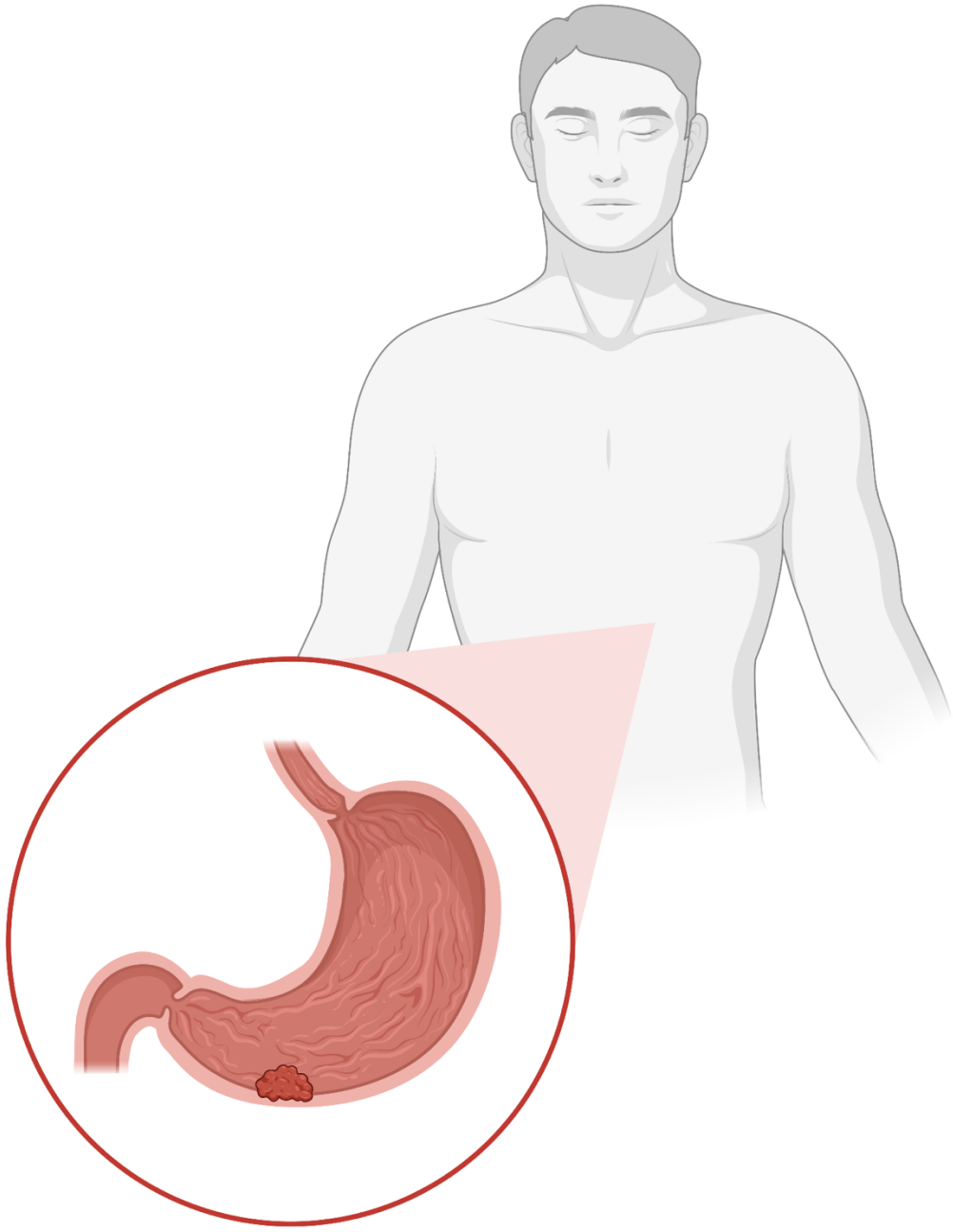
Acronyms/abbreviations

ACMG	american college of medical genetics and genomics
AMP	association for molecular pathology
BC	breast cancer
CIN	chromosomal instability
CNV	copy number variant
CRE	cis-regulatory element
CRC	colorectal cancer
DGC	diffuse gastric cancer
DSB	double strand break
FGP	fundic gland polyps
FIGC	familial intestinal gastric cancer
GAPPS	gastric adenocarcinoma and proximal polyposis of the stomach
GC	gastric cancer
GS	genomic stable
HDGC	hereditary diffuse gastric cancer
HLBC	hereditary lobular breast cancer
IGCLC	international gastric cancer linkage consortium
LAD	lamina-associating domains
LBC	lobular breast cancer
LP	likely pathogenic
HDR	homology-directed repair
lncRNA	long noncoding RNA
LOH	loss of heterozygousity
NGS	next-generation sequencing
NHEJ	non-homologous end joining repair
MSI	microsatellite instability
OC	ovarian cancer

PAM	protospacer adjacent motif
P	pathogenic
sgRNA	single-guide RNA
SNV	single nucleotide variant
SV	structural variant
TAD	topological associating domain
TCGA	the cancer genome atlas
TF	transcription factor
TFBS	transcription factor binding site
TRS	tumour risk syndromes
TSS	transcription start site
WES	whole exome sequencing
WGS	whole genom sequencing
WHO	world health orginazation
WT	wild-type

Chapter 1

General introduction



1. Genomic variation

The human genome compromises roughly three billion base pairs. Indeed, great efforts have been made to make sense of this big set of information both in health and in disease. As a result, there has been an unprecedented characterization of genomic alterations, encompassing single nucleotide variants (SNV), small insertions/deletions (indels), and structural variants (SVs). SVs may compose balanced rearrangements, including inversions, and unbalanced rearrangements, or copy number variants (CNVs), encompassing deletions and duplications.

SNVs have been firstly proposed in a knowledge-driven approach to disclose disease-causing variants for several hereditary syndromes. Detection of microsatellite instability in colorectal cancer of early adulthood patients led to the prioritization of mismatch repair genes as disease causing (Ionov et al. 1993; Thibodeau et al. 1993). Identification of SNVs in *MSH2* as the main germline genetic cause of lynch syndrome dates back to 1993 (Fishel et al. 1993; Leach et al. 1993). In the subsequent years, other mismatch repair genes were proposed also in a knowledge-driven approach, encompassing *MLH1, PMS2, PMS1, MSH6* (Bronner et al. 1994; Nicolaides et al. 1994; Miyaki et al. 1997). As more data became available, CNVs in syndrome-causing genes were acknowledged as putative germline genetic causes of several hereditary syndromes (Wijnen et al. 1998). In an attempt to disclose the missing heritability in syndrome-causing genes, gene-panel sequencing, whole exome sequencing (WES) and whole genome sequencing (WGS) were prioritized.

Next-Generation Sequencing (NGS) technologies exponential growth led to the identification of numerous genomic alterations. Notwithstanding, functional and medical interpretations of these genomic alterations have been proven challenging, mainly in the non-coding genome. Thus, most families remained genetically unsolved.

While coding genomic variants have a straightforward approach to gene dosage, non-coding variants, either intronic or intergenic, lack information for correct pathogenicity assessment. Indeed, the non-coding genome accounts for 98% of the entire genome (Istrail et al. 2004), from which most non-coding regions may be involved in gene regulation (Dunham et al. 2012). Proving non-coding variants as disease causative remains challenging, yet the 3D chromatin organization has been proposed as a major player.

2. 3D genome organization

Within the nucleus, chromatin spatial distribution is tightly organized in chromosomal territories that rarely merge within active A or repressed B compartments in time and space (**Figure 1**) (Cremer and Cremer 2001; Pombo and Dillon 2015). Compartment A associates with actively transcribed euchromatin, organized in topological-associating domains (TAD) (Cremer and Cremer 2001; B and G 2016), while the B compartment is transcriptionally inactive, with higher heterochromatin compaction and commonly localizes near the nuclear lamina (lamina-associating domains – LAD) (Cremer and Cremer 2001).

2.1. Chromatin functional unit

Higher chromatin contacts occur within TADs, which are mainly stable across cell types and species, suggesting TADs as a fundamental structural and functional unit (Bonev and Cavalli 2016a). TADs commonly display coordinated gene expression across different cell types (Dixon et al. 2012; Krefting et al. 2018). Additional organization layers encompass sub-TAD structures that display dynamic cell-type-specific enhancer- or silencer-promoter interactions and tissue-specific gene expression (**Figure 1**) (Bonev and Cavalli 2016a). Enhancer-promoter communication has been described in close proximity(Cruz-Molina et al. 2017), but also up to 850kb distant in the DNA sequence (**Figure 1**) (Lettice et al. 2003).

Architectural proteins CCCTC-binding factor (CTCF) and cohesin insulate neighbouring TADs, enabling Mb-scale communication (Phillips-Cremins et al. 2013) and generating stable TAD boundaries (Dixon et al. 2012). Co-binding of convergent CTCF motifs and cohesin were found in most loops (Rao et al. 2014). According to the loop extrusion model, cohesin holds the DNA within its ring structure to form a loop and dislocates through the DNA until a convergent CTCF site is found (Fudenberg et al. 2016). CTCF knockdown drives TADs loss in mouse embryonic stem cells, hinting on the importance of this transcription factor (TF) in TAD formation and enhancer-promoter communication (Nora et al. 2017). However, compartments A and B remained unchanged, with a reduced number of misregulated genes, pinpointing the role of other elements for compartments and gene regulation (Nora et al. 2017; Rao et al. 2017; Schwarzer et al. 2017). Loops of CTCF-independent contacts can be mediated by zinc finger Yin Yang 1 (YY1) in specific cell types, that compose homotypic attraction of chromatin structures through liquid-liquid phase separation(Hnisz et al. 2017; Beagan et al. 2017). Thus, compartmentalization may be achieved through compatible chromatin states, such as chromatin modifications and transcriptional activity comprised of phase-separated condensates (Rao et al. 2014; Andrey and Mundlos 2017a). Likely, a complex interplay of loop extrusion and homotypic attraction co-exist in shaping regulatory structures. While loop extrusion may

regulate enhancer-promoter communication, homotypic attraction may control organization into A and B compartments.

Altogether, TADs provide an insulated structural unit that defines gene regulation units, constraining enhancer activity, and shaped by CTCF and cohesin. Notwithstanding, the functional relevance of the 3D chromatin architecture is still poorly understood.

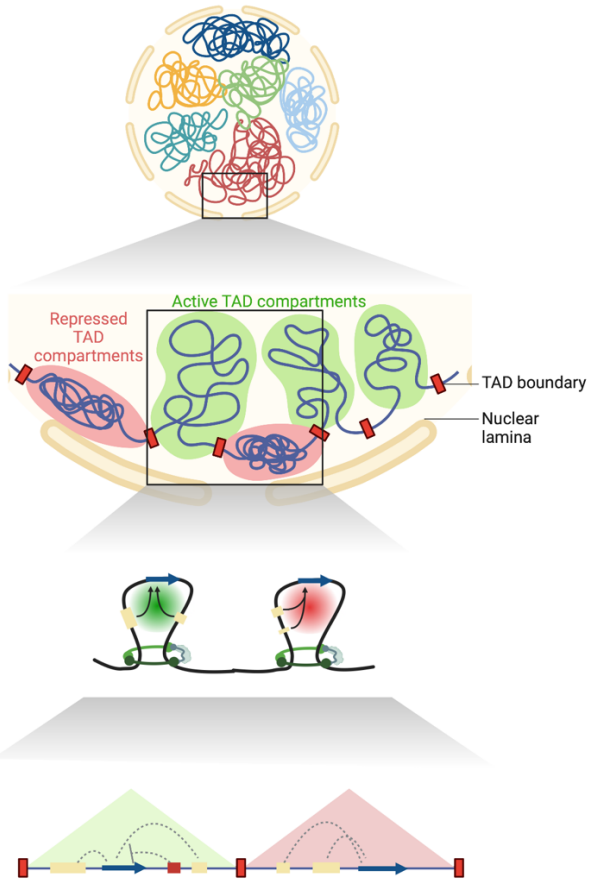


Figure 1. *Chromatin organization in the nucleus. Adapted from "Chromosome Organization in Nucleus: TADs", by BioRender.com (2022). Retrieved from https://app.biorender.com/biorender-templates*

2.2. Regulatory elements

Gene regulation in time and space is instrumental to the developmental process. A biological way to achieve this is through regulatory elements, controlling when and which specific cells express each gene (Spitz and Furlong 2012a). Located in close 3D proximity with the target genes, *cis*-regulatory elements (CREs) play an important role, in synergy with distal-regulatory elements.

Upon the 3D-chromatin folding, CREs can interact and regulate their target genes, by promoting or inhibiting them. CREs encompass promoters, acting in <1kb (Zabidi and Stark 2016), and enhancers, silencers and insulators that may act more distally (Gaszner and Felsenfeld 2006). To achieve a tight

spatiotemporal gene regulation, CREs are redundant, thus several CREs may regulate the same target gene, and a given CRE may regulate several target genes (Lettice et al. 2003). At cell-type-specific regulation, enhancer-promoter loops occur within sub-TADs shaped by architectural proteins CTCF and cohesin and open chromatin marks. However, the mechanisms behind are still poorly understood.

Gene transcription requires initiation, elongation and termination, for which the former is performed at promoters. Core promoters enable the recruitment and binding of the transcription machinery, encompassing the transcription start site (TSS) (Haberle and Stark 2018). Located upstream, proximal promoters drive transcription activation, recruiting TFs. While silencers inhibit gene expression, insulators block the effect of any enhancer or silencer with their target gene (Gaszner and Felsenfeld 2006).

Enhancers in synergy with promoters drive gene transcription, independently of enhancer orientation and distance to TSS (Bulger and Groudine 2011; Long et al. 2020). Despite position- and orientation-independent, 3D physical proximity of enhancers is essential for transcription initiation (Bulger and Groudine 2011). Enhancer-promoter communication occurs via TFs homotypic interactions in both elements that bridge their interaction and drive transcription initiation (Boija et al. 2018). Promoters are generally regulated by several cooperating enhancers with differing or overlapping spatial or temporal activities towards one or several promoters (Anderson et al. 2014). Depending on redundancy and dosage, disruption of a redundant enhancer might resist gene misexpression, while impairment of single-activity enhancers likely leads to disrupted expression. Thus, complex regulatory landscapes might combine cooperating or single-acting enhancers spatiotemporally for tight regulation of gene expression.

Enhancers harbour transcription factors binding sites (TFBS), that enable tissue-specific gene expression, contributing to the gene expression diverse pool. For several years, conservation across species was a major hint of enhancers location, yet non-conserved enhancers have been found to regulate their target genes as well (Ahituv et al. 2007). With the advances in NGS technologies and the ENCODE project, biochemical annotations may pinpoint enhancers. Chromatin accessibility methods, such as ATAC-seq (Buenrostro et al. 2013, 2015) and DNase-seq (Crawford et al. 2006), indicate regions of available chromatin for TFs binding (**Figure 2A**). For specific TFs and chromatin marks, chromatin immunoprecipitation (ChIP-seq) (Visel et al. 2009) may also provide crucial data for enhancer identification (**Figure 2A**). While H3K27ac and H3K4me1 are associated with active enhancers or promotes, H3K4me3 is suggestive of repressed regions (Ernst et al. 2011; Pradeepa et al. 2016). 3C-based technologies provide a toolbox for enhancer-promoter interrogation or silencer-promoter interactions (reviewed in section 2.2.2.) (**Figure 2A**). Notwithstanding, these methods should be used complementarily to prioritize regions for enhancer testing.

The regulatory landscape of genes within a TAD extends beyond several hundred kb, their tissue specificity and extent can be measured resourcing to regulatory sensors (**Figure 2D**) (Kothary et al. 1989; Ruf et al. 2011). Integration of a minimal promoter and reporter gene reveals the enhancer activity at any given time and tissue of all enhancers within the TAD (**Figure 2B**). Enhancer-specific activity may be tested resourcing to episomal reporter vectors (Kothary et al. 1989; Chiocchetti et al. 1997) or massive parallel reporter assays (Arnold et al. 2013; Muerdter et al. 2018) *in vitro* and *in vivo* (**Figure 2B**). The former contains a candidate enhancer, a minimal promoter and a reporter gene, while the latter accounts for a barcode to test thousands of putative enhancers in a single experiment.

CREs edition *in vitro* or *in vivo* is the ultimate assay to attribute their functional relevance, resourcing to CRISPR-Cas9 deletions or enhancer reporter assays (Jinek et al. 2012, 2013; Cong et al. 2013; Mali et al. 2013).

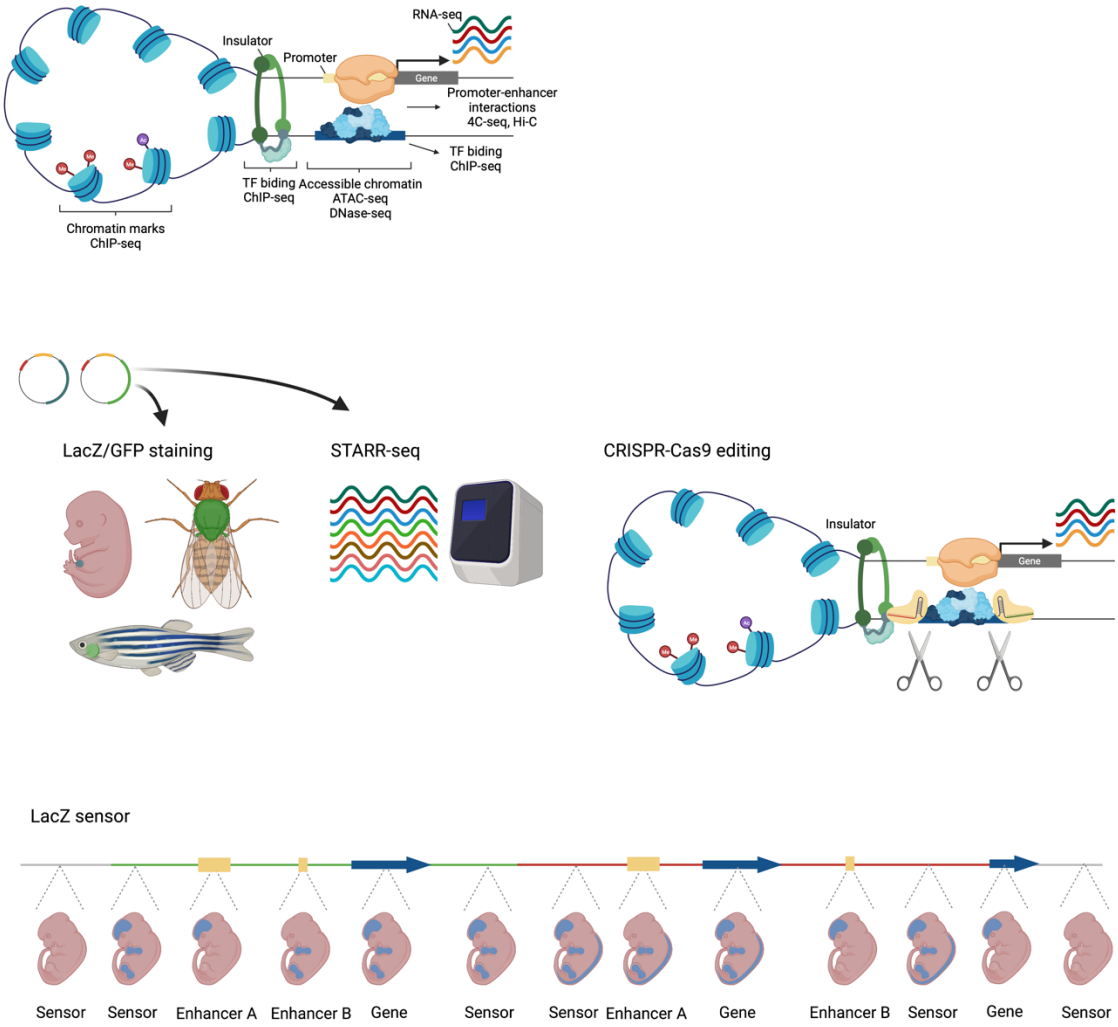


Figure 2. Methods for enhancer discovery and testing. *A) Current methods for prioritization of enhancers, B) LacZ sensors, C) In vivo functional enhancer reporter in episomal and multiplex assays, D) CRISPR-Cas9 functional edition of enhancers. Original image from Celina São José, created with BioRender.com.*

2.2.1. Deciphering functional impact with CRISPR-Cas9 editing

Clustered regularly interspaced short palindromic repeats (CRISPR) is a procaryotic defence mechanism to remove viral infecting DNA found in *Escherichia coli* (Ishino et al. 1987). A nuclease (Cas9) recognizes a protospacer adjacent motif (PAM) adjacent to a targeted sequence (crRNA or sgRNA) and performs a double strand break (DSB) in the invading DNA, editing it (Mojica et al. 2005; Jinek et al. 2012; Gasiunas et al. 2012; Shah et al. 2013). Genome engineering revolution dates back to 2013 when CRISPR-Cas9 was successfully implemented in human cancer cells and human pluripotent stem cells, for which a Nobel prize in chemistry was received in 2020 (Jinek et al. 2013; Cong et al. 2013; Mali et al. 2013). Upon a Cas9 DSB, the cell repairs the DNA damage by error-prone Non-Homologous End Joining repair (NHEJ) or Homology-Directed Repair (HDR). While NHEJ may add random base pairs to repair the DSB, HDR uses a DNA template to precisely repair the DNA damage. Resourcing to one sgRNA, indels or SNVs could be mimicked with NHEJ or HDR, respectively; while with two sgRNA, CNVs or SVs could be achieved either with NHEJ or HDR. Despite a revolutionized technique in the genetics field, CRISPR-Cas9 has a major disadvantage, the off-target effects (Fu et al. 2013). Depending on mismatches number and location, sgRNAs may bind to other genomic regions located next to a PAM, that recognized by Cas9 could be edited (Pattanayak et al. 2013). Using different Cas-nucleases or administrating a ribonucleocomplex Cas9-sgRNA have been shown to reduce the off-target effects (Kim et al. 2014; Lin et al. 2014), still progress could be made to increase CRISPR-edition specificity.

CRISPR-Cas9 edition is a powerful tool to unveil the functional impact of specific genomic alterations *in vitro* and *in vivo* (**Figure 2C**). Indeed, CRISPR-Cas9 *in vivo* recapitulation of homozygous 27-63kb deletions located in a long non-coding RNA (lncRNA) 300kb upstream of EN1 leads to complex limb malformations (Allou et al. 2021).

2.2.2. Conformational capture techniques (3C)

3D chromatin folding enables enhancers at the Mb genomic distance to communicate promoters their regulatory outputs by close spatial proximity. Despite distant in the DNA sequence, enhancer-promoter or silencer-promoter interactions regulate gene expression.

3C-based technologies capture a snapshot of interacting DNA sequences in close spatial proximity. To achieve this, these technologies capture the chromatin state through paraformaldehyde fixation (**Figure 3**) (Dekker et al. 2002). Subsequently, interacting fragments are generated by restriction enzyme digest and ligation. For a one-to-one 3C approach, with previous knowledge of both interacting partners, a PCR or RT-PCR is performed for their identification (Dekker et al. 2002). Combination of 3C with high-throughput sequencing originated several other 3C-based technologies. The circular chromosome

conformation capture followed by NGS (4C-seq) resources to an inverse PCR with primers located in the known sequence, viewpoint, through a one-to-all approach (Zhao et al. 2006). Thus, 4C-seq enables the identification of all potential interactors of a given region, for example, a promoter (Simonis et al. 2006). With high sensitivity and resolution, 4C-seq is ideal to unveil enhancer-promoter interactions in a specific locus of interest. Capture-C is a many-to-many approach that resources to biotinylated oligos and streptavidin beads to enrich and capture interactors of promoters or enhancers of interest (Hughes et al. 2014). An all-to-all technology to generate contact maps at a Mb scale is Hi-C (Lieberman-Aiden et al. 2009). This high-resolution approach unveils the 3D chromatin organization at a genome-wide scale, without any previous knowledge of a viewpoint or interacting partners (Rao et al. 2014). To enrich a Hi-C library in specific loci, one can resource to biotinylated oligos and streptavidin beads (capture-Hi-C) (Andrey and Mundlos 2017a).

3C-based technologies have been refined in recent years to increase resolution and increase knowledge in regulatory elements, notably in association with other functional elements identification, described in the previous section.

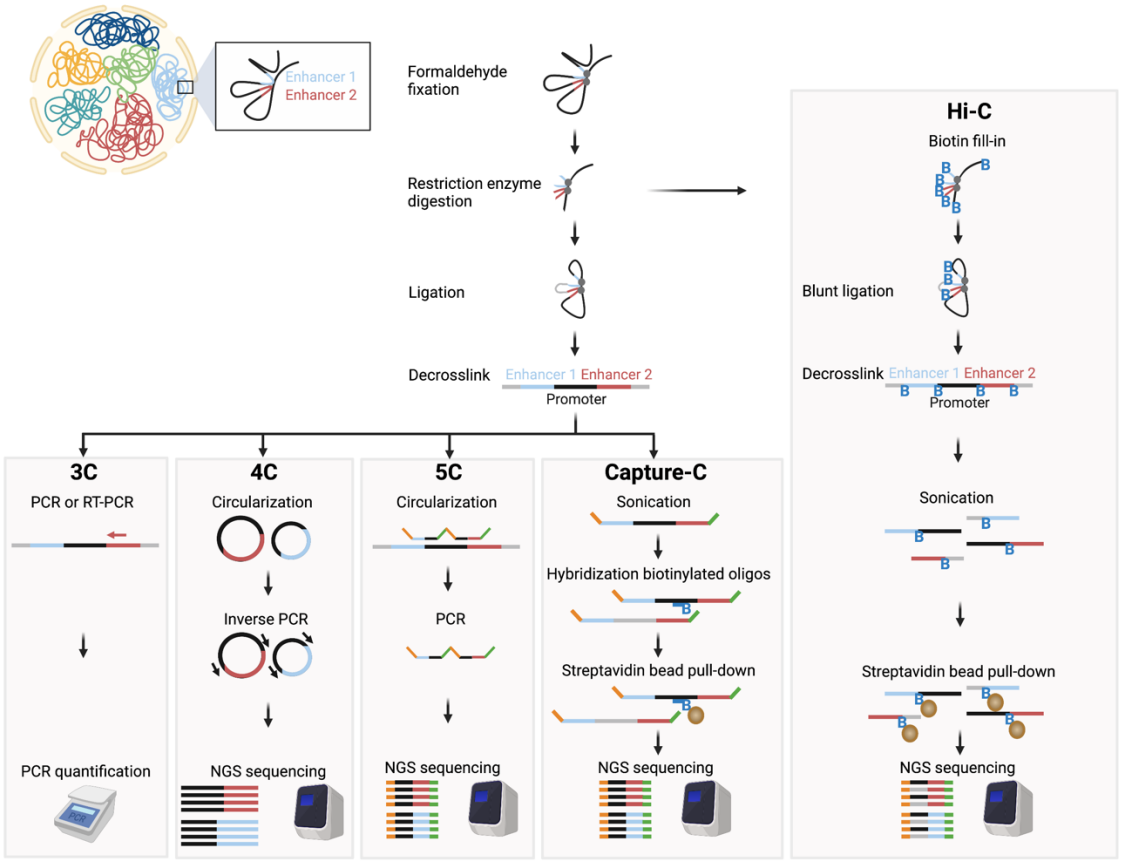


Figure 3. *3C-based methods for identification of enhancer-promoter communication (Krijger, P. H. L., 2020). Original image from Celina São José, created with BioRender.com.*

2.2.3. Impact of regulatory elements in disease

TADs regulate enhancer-promoter interactions and their boundaries place limits to genes regulatory networks (**Figure 4A**). Impairment of TADs boundaries or their structure may lead to enhancer-promoter ectopic interactions and, therefore, gene misexpression. Indeed, genomic variants may switch CREs position and dosage, reshaping enhancer landscapes and TADs, leading to gain or loss of function.

Effects on enhancer dosage generally occur in intra-TAD SVs, disrupting several cooperating enhancers. Given the high enhancer redundancy, deletion of individual elements will be tolerated in most cases, namely in a heterozygous state. Yet, deletion of several elements may lead to gene loss of function. This is the case for *Ihh*, which displays 9 tissue-specific enhancers working in addictive manner (Will et al. 2017). Deletion of these composition and dosage enhancers led to different degrees of *Ihh* knockdown with distinct tissue expression patterns (Will et al. 2017). Alternatively, duplication of such elements may result in gain of gene function or overexpression, as observed for the *Ihh* gene (Will et al. 2017). According to current knowledge, intra-TAD non-coding inversions may have a mild/absent effect on gene expression, as regulatory elements are position- and orientation-independent.

Inter-TAD SVs may re-structure TADs in new configurations, repositioning TAD boundaries. Inversions encompassing TADs boundaries may lead to TADs re-shuffling (**Figure 4B**). Relocation of *EPHA4* enhancers in close proximity to the *WNT6* gene leads to *WNT6* overexpression, according to *EPHA4* tissue-specificity, and *EPHA4* downregulation from lack of promoter-enhancer communication (Lupiáñez et al. 2015).

Deletions that include the TAD boundary may lead to TADs fusion, such as a large deletion of the *EPHA4* locus (**Figure 4C**) (Bianco et al. 2018). TADs fusion may result in gain or loss of function, depending on the increase or loss of enhancer-promoter communications. In glioblastoma, SVs have been shown to relocate *GFI1* and *GFI1B* genes to hijack super-enhancers (Northcott et al. 2014).

Duplications that cross TAD boundaries may originate new insulated chromatin domains, denominated neo-TADs (**Figure 4D**). Indeed, duplications at the *Sox9* locus boundary element and the *Kcnj2* gene hijacks *Sox9* enhancers, leading to *Kcnj2* overexpression and cooks syndrome predisposition (Kurth et al. 2009).

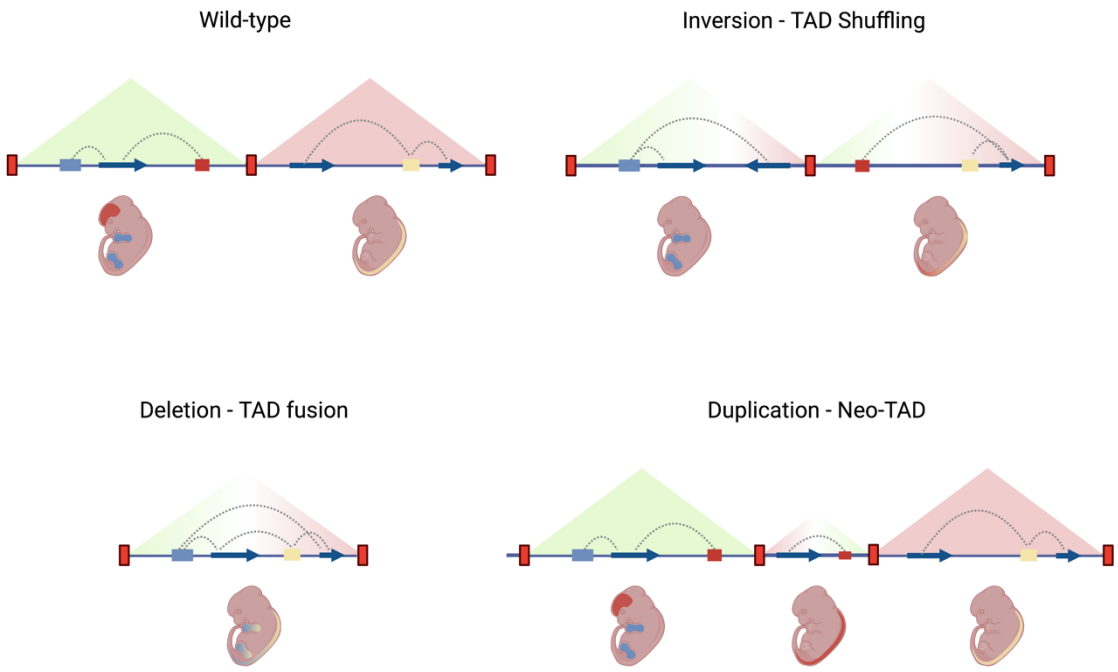


Figure 4. Possible impacts of SVs in TADs configuration. *A) Wildtype, B) Inversion - TAD shuffling, C) Deletion - TAD fusion, D) Duplication - neo-TAD formation. Adapted from Spielmann M., 2018, created with BioRender.com.*

Recent sequencing developments have called SVs for which the functional impact in still unknown. Only the integration of these layers, biochemical annotations, 3C-based technologies, enhancer assays and CRISPR-Cas9 edition, increases the knowledge of the functional impact of non-coding SVs and predicts their phenotypic consequences.

The technological advances triggered the disclosure of genetic determinants, particularly in the non-coding genome, and gastric cancer is no exception. The next section is dedicated at exploring gastric cancer.

3. Gastric cancer

3.1. Epidemiology and risk factors

Gastric cancer (GC) comprises a spectrum of diseases that develop upon a malignant transformation of cells in the stomach. According to the latest statistics, over 1 million individuals have been diagnosed with GC, ranking fifth most incident cancer worldwide (Ferlay et al. 2021). Almost 60% of the cases are late diagnosed in stage III or IV, which may explain a 5-year survival of nearly 20% (Amin et al. 2017; Ajani et al. 2017). Thus, in 2020, 800 thousand individuals died, resulting in the fourth leading cause of cancer-related mortality (Ferlay et al. 2021). Despite a steady tendency to decrease, related with increased hygiene standards, consumption of fresh vegetables and fruits and *H. pylori* eradication (Moayyedi et al. 2002) the world health organization estimates that 1,77 million individuals will be diagnosed with GC in 2040 (Sonnenberg 2011; Ferlay et al. 2021). Most GC cases occur in developing countries located in Eastern Asia, Eastern Europe and South America (Ferlay et al. 2021).

GC development may be a result of a synergistic effect of environmental risk factors, such as *Helicobacter pylori* or EBV infection and diet, and genetic risk factors. *H. pylori* is thought as the major risk factor for GC development, mainly through inflammation of the gastric mucosa, entitled gastritis (Morgan et al. 2022), covered in section 3.2.

3.2. Histological and molecular classification

Most GCs are adenocarcinomas, originated from epithelial cells, yet less frequent tumours may occur from other cell types, including lymphoma, sarcoma, and others (Ajani et al. 2017).

Diagnosis and treatment of GC rely on histological classification of tumours, thus several systems have been proposed, Laurén (LAUREN 1965), World Health Organization (WHO) (Nagtegaal et al. 2020) and Carneiro (Carneiro et al. 1995), from which Laurén and WHO classifications are the most widely used. Histology features are the foundation of Laurén classification to categorize tumours in intestinal-type, diffuse-type and mixed-type (LAUREN 1965). WHO basis the classification on histological features, in which GC can be categorized into adenocarcinoma (papillary, tubular and mucinous), signet ring cell carcinoma or other poorly cohesive carcinomas, mixed and undifferentiated (Nagtegaal et al. 2020).

Table 1. *Histological and molecular gastric cancer classifications.*

Laurén 1965	Carneiro 1995	WHO 2010	TCGA
Intestinal	Glandular	Papillary Tubular Mucinous	↑↑CIN ↓GS ↓↓EBV^{+}/MSI
Diffuse	Isolated cell	Poorly cohesive SRCC	↑↑GS ↓CIN ↓↓EBV^{+}/MSI
Mixed	Mixed	Mixed	↑↑ EBV^{+}
Undetermined	Solid Rare variants	Undifferentiated Rare variants	↑MSI ↓↓CIN/GS

Intestinal and diffuse histotypes display distinct histological and molecular pathogenesis. According to the Correa's model, intestinal-type gastric cancer (IGC) originates in a stepwise manner, mainly through the cooperation between host genetic factors and environmental factors, encompassing *H. pylori* infection, diet, smoke, among others (Correa et al. 1975). Briefly, environmental factors predispose to inflammation of the gastric mucosa, that evolves from atrophic/chronic gastritis, intestinal metaplasia, dysplasia to malignant transformation (Figueiredo et al. 2013) (**Figure 5A**).

The diffuse histotype evolves from gastric mucosa, in the absence of a clear pathomechanism and reduced environmental influence. Cells display poorly cohesive histopathology and/or signet ring phenotype, that infiltrates the stomach wall, with poor endoscopy diagnosis yield and reduced survival (**Figure 5B**) (Assumpção et al. 2020).

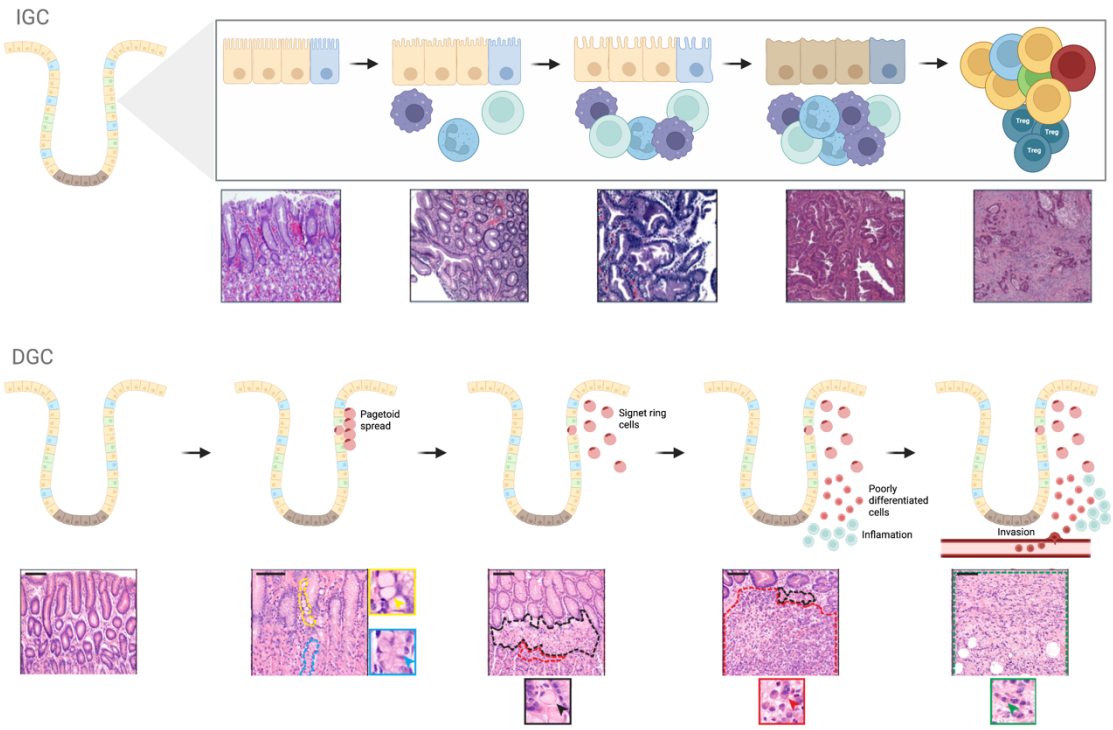


Figure 5. Gastric cancer pathogenesis. *A) Correa's cascade for intestinal gastric cancer (IGC) development, B) diffuse gastric cancer (DGC) development. Adapted from Tan, P., 2015 and Monster, J. L., 2022, created with BioRender.com.*

Despite a very well described cascade for IGCs, environmental factors fail to explain most GCs, paving the way for the influence of genetic factors. At the somatic level, SNVs, CNVs, epigenetic modifications, among others may represent GC molecular signatures and directed therapy targets (Gravalos and Jimeno 2008; Garcia-Pelaez et al. 2021a). Tumours with *ERBB1* amplifications may benefit from targeted therapy encompassing HER2 inhibitor.

Thus, the cancer genome atlas (TCGA) defined a GC molecular classification according to EBV positive infection, microsatellite instability (MSI), genomic stable (GS) and chromosomal instability (CIN) (**Figure 6**) (Bass et al. 2014). Most GC cases display *TP53* or *CDH1* mutations, other frequently mutated genes include *SMAD4*, *PIK3CA*, *RHOA*, *ARID1A* and *KRAS* (Bass et al. 2014). GS cancers mostly display diffuse histotype, *CDH1* mutations and, therefore, cell adhesion pathways impaired (Bass et al. 2014). Contrarily, CIN GCs are mainly of intestinal-type, displaying TP53 mutations and RTK-RAS activation (Bass et al. 2014).

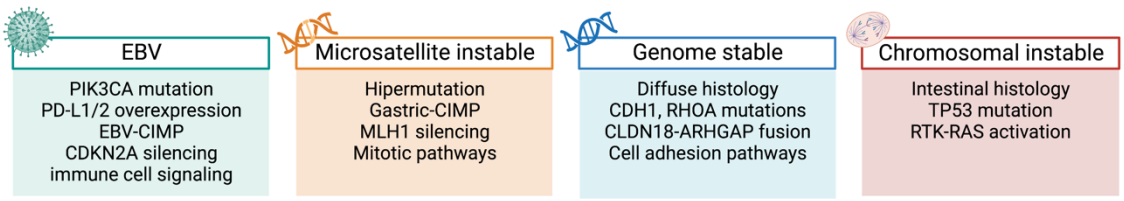


Figure 6. *TCGA gastric cancer molecular classification.*

3.3. Diagnosis, treatment, and survival

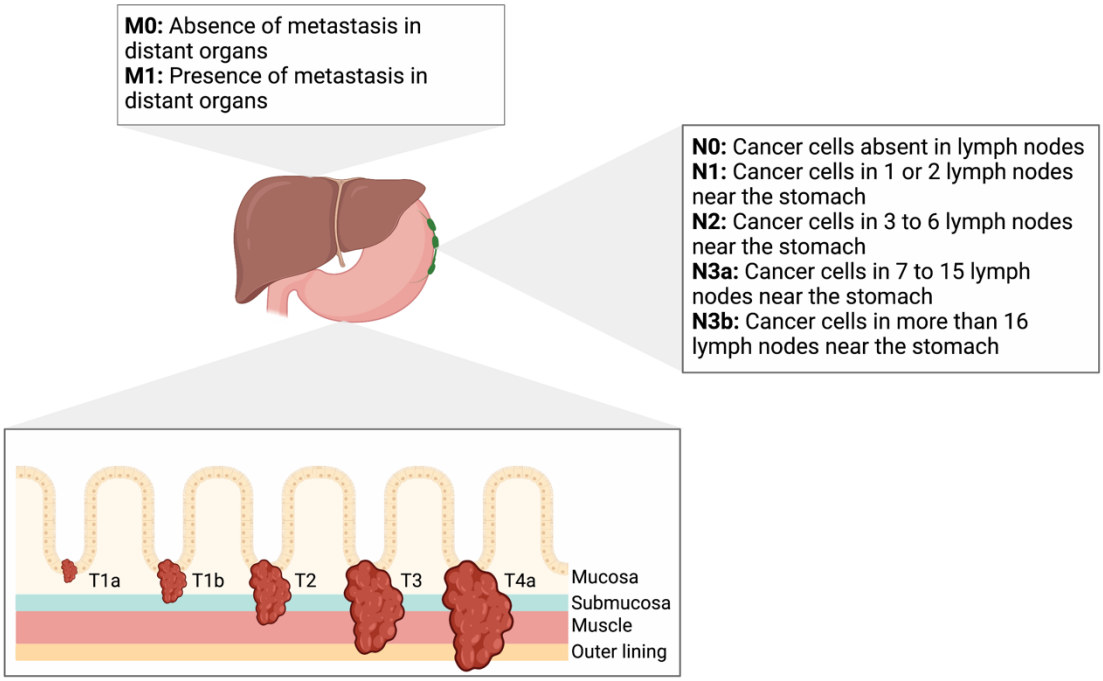
At an early-stage GC is asymptomatic, or displays unspecific symptoms, encompassing anorexia, early satiety, indigestion, abdominal pain and weight loss (Smyth et al. 2016). Thus, diagnosis is hampered with a significant impact in treatment options and, therefore, survival.

Overall, GC prevention plays a major role in reducing GC mortality. Recommendations include better dietary habits, with increased fresh vegetables and fruits, reduced alcohol consumption and smoking, and *H. pylori* eradication (Tsugane and Sasazuki 2007). GC early detection is key to reduce GC-associated mortality, as performed in Japan. Massive screening programs increased the detection of early-stage GC by 45%, with a shift in 5-year survival of 2.24-fold (Kim et al. 2018). Alternative methods used in Asian countries include cutting-edge surgical procedures and chemotherapy (Japanese gastric cancer treatment guidelines, 2014).

Upper endoscopy is the gold standard for GC screening, enabling both visualization and biopsies collection for histological classification (Smyth et al. 2020). The increased cost and invasiveness reimburse with increased diagnosis sensitivity. Combination with other imaging methods, such as ultrasound, provides information on the presence of superficial cancer (Mocellin and Pasquali 2015). Double-contrast barium radiographs with photofluorography is frequently used to identify several gastric lesions, but with reduced sensitive. Metastasis are screened resourcing to abdominal and chest computed tomography scan or magnetic resonance to identify liver, lungs or peritoneum metastasis (Ma et al. 2022). Since peritoneal relapse is common in resected DGCs, explorative laparoscopies are recommended in tumours planned for resection (Smyth et al. 2020).

During diagnosis, a patient undergoes a series of imaging tests to clinically stage the tumour and guide therapy. According to the cTNM staging, a tumour can be categorized in relation to the depth of gastric wall invasion (T), infiltration of regional lymph node (N) and presence of distant metastasis (M) (**Figure 7A**). Each of these points add up to a score that assigns a clinical staging (**Figure 7B**) (Amin et al. 2017).

A



B

			M0					M1
			N0	N1	N2	N3a	N3b	Any N
			0	1-2	3-6	7-15	>15	Metastasis
M0	T1	Mucosa, submucosa	1A	1B	IIA	IIB	IIIB	IV
	T2	Muscularis propria	1B	IIA	IIB	IIIA	IIIB	
	T3	Subserosa	IIA	IIB	IIIA	IIIB	IIIC	
	T4a	Serosa	IIB	IIIA	IIIA	IIIB	IIIC	
	T4b	Adjacent structure	IIIA	IIIB	IIIB	IIIC	IIIC	
M1	Any T	Metastasis	IV					

Figure 7. Gastric cancer clinical stagging. *A) Categories of tumour stagging, B) Score sheet for clinical stagging. Original image from Celina São José, created with BioRender.com.*

Treatment with curative intent requires endoscopic or surgical resection. Endoscopic resection is recommended for early GC stages (T1a) lacking lymph node metastasis risk factors, such as poor differentiation (Hatta et al. 2020). While gastrectomy and lymphadenectomy is performed in T1 tumours with cancer cells in lymph nodes (N+) and T2 to T4a with absent distant metastasis (M0) (Smyth et al. 2020).

Perioperative chemotherapy (before and after surgery) has been shown to increase 5-year survival in 12% by reduction of tumour size and microscopic propagation (Cunningham et al. 2006). Interestingly, postoperative radiotherapy does not increase survival in combination with perioperative or adjuvant (after surgery) chemotherapy when gastrectomy with lymphadenectomy is performed (Cats et al. 2018).

Current chemotherapy regimens include trastuzumab for HER2-positive GCs (with *ERBB-2* oncogene amplification), which increases overall survival in inoperable tumours (Bang et al. 2010), and VEGFR2-positive (Fuchs et al. 2014). Immunotherapy has been growing in the treatment of several cancers and GC is no exception. Anti-PD-1 checkpoint blockage antibodies, nivolumab and pembrolizumab, that survival in advanced lines of GC therapy (Kang et al. 2017). However, heterogeneity remains a major obstacle for targeted therapy, since tumour clones may escape chemotherapy, thus, more research on novel targets is essential.

For advanced unresectable or metastatic (M1) GCs, chemotherapy improves both survival and overall quality of life if patients are fit, as opposed to supportive care (Vanbutsele et al. 2020). Palliative care should also include nutritional advice to avoid weight loss and radiotherapy for painful metastases (Rosania et al. 2016).

Five-year survival rates differ according to the tumour stage, raking between 57% for stage I and 5% for stage IV (**Table 2**) (Amin et al. 2017; Ajani et al. 2017).

Table 2. *Five-year survival rates in gastric cancer.*

Clinical stage	5-year survival (%)
Stage I	57%
Stage II	37%
Stage III	26%
Stage IV	5%

These results highlight the importance of early-stage diagnosis, through biomarkers or screening programs. In western Europe most GC cases are diagnosed in advanced stages, yielding poor survival rates. In east Asia, massive screening programs and cutting-edge surgical techniques result in earlier-stage diagnosis and improved overall survival (Japanese gastric cancer treatment guidelines, 2014; Kim et al. 2018).

A small fraction of GCs displays a hereditary nature. The next section is dedicated at exploring the GC hereditary component.

4. Genetic predisposition to gastric cancer

We recently reviewed the current knowledge on the germline predisposition and clinical features from patients and families affected by GC (**Appendix A**).

Evidence for familial predisposition to GC dates back historically to the 18th Century, was observed in the Bonaparte Family and described by Sokoloff in 1938 (Sokoloff 1938). It took 60 years for a direct association between a germline heritable genetic defect and increased risk to DGC development to be discovered (Guilford et al. 1998). In 1999, the International Gastric Cancer Linkage Consortium (IGCLC) established for the first-time diagnostic criteria and guidelines for the clinical management of GC families (Caldas et al. 1999).

It is fairly established that 1-3% of all GCs are hereditary, and generically classified into three main tumour risk syndromes (TRS): (1) hereditary diffuse gastric cancer (HDGC); (2) gastric adenocarcinoma and proximal polyposis of the stomach (GAPPS), and; (3) familial intestinal gastric cancer (FIGC) (Caldas et al. 1999; Worthley et al. 2012).

Both HDGC and GAPPS have well-recognized heritable mechanisms that confer high risk for developing mainly GC, while FIGC remains widely understudied, presenting a large disease spectrum and generally occurring at older age (Carvalho et al. 2021). Despite their low incidence, these heritable diseases deserve special attention due to their high morbidity and mortality rates (Vos et al. 2019). In fact, according to the most recent global cancer statistics, GC is the fourth most deadly cancer worldwide(Sung et al. 2021). Therefore, it is of uttermost importance to identify patients with GC predisposition from birth, to improve their genetic testing, surveillance, and/or risk-reduction, ultimately to prevent disease onset. This can be achieved by increasing the knowledge on the inherited factors associated to GC (Pocurull et al. 2021).

It is essential to have access to a detailed family history, patient's age of onset, tumours' histological classification, and reports of secondary cancers from all affected individuals, for proper allocation of isolated or familial cases of GC, clustering into a particular GC-TRS (Choi and Kim 2016). Laurén and WHO classifications (LAUREN 1965; Nagtegaal et al. 2020) are considered the foundation for histopathological characterization of gastric tumours according to their subtypes and should be used when collecting data from patients with GC predisposing syndromes.

4.1. Hereditary diffuse gastric cancer (HDGC)

HDGC is an autosomal dominant TRS, predisposing to DGC and lobular breast cancer (LBC). It was the first recognized heritable GC-TRS and is historically known for its association with germline variants in the *CDH1* gene, but recently *CTNNA1* variants were also found segregating with early-onset DGC (**Figure 8**) (Guilford et al. 1998; Majewski et al. 2013). The definition of HDGC evolved along the years, as well as the clinical criteria driving genetic testing. Clinical practice

guidelines were updated three times so far, first to better identify at-risk individuals for DGC, and later considering the development of LBC in female carriers of *CDH1* germline variants (Caldas et al. 1999; Fitzgerald et al. 2010; van der Post et al. 2015; Blair et al. 2020). In addition, the discovery of personal or family history of cleft-lip and/or cleft-palate in DGC patients or families carrying *CDH1* loss of function variants, led to the contemplation of these phenotypes, when co-occurring with DGC, as criteria for genetic testing (Blair et al. 2020). A detailed description on the evolution and the current clinical criteria according to IGCLC is depicted in **Figure 9**. To acknowledge the fact that *CDH1* variants may predispose preferentially to LBC in families lacking DGC, the 2020 clinical guidelines created a novel entity – hereditary lobular breast cancer (HLBC). HLBC families will remain under this designation, while no DGCs are detected in carriers, and their management will be discussed on a case-by-case basis (**Figure 2**; (Blair et al. 2020)). Notwithstanding, HLBC entity remains controversial, in which several HLBC families who initially lacked DGC history exhibit high rates of occult signet ring cell gastric carcinoma (Gamble et al. 2022). A clinical favour of DGC or LBC predisposition may rather originate from different geographical origins, environmental factors, modifier genetic variants, among others. Given the high incidence of LBC in HDGC, García-Pelaez et al suggest widening the 2020 clinical criteria to maximize diagnosis of families lacking DGC phenotype (Garcia-Pelaez et al. 2023). The LBC-centred suggested criteria include (1) at least 2 BC cases, 1 case of confirmed LBC, (2) GC or BC history, 1 confirmed LBC regardless of age, and (3) isolated LBC in individuals aged <55 years (Garcia-Pelaez et al. 2023).

According to the American College of Medical Genetics and Genomics and the Association for Molecular Pathology (ACMG/AMP) criteria, germline *CDH1* pathogenic (P) or likely pathogenic (LP) SNVs and CNVs, (Lee et al. 2018), or *CTNNA1* SNVs are considered causative of HDGC. Importantly, families fulfilling clinical criteria, but lacking a causal variant in *CDH1* or *CTNNA1* are still considered by IGCLC as HDGC-like families, understood as genetically unsolved families, and in need for disease surveillance (Caldas et al. 1999; Fitzgerald et al. 2010; van der Post et al. 2015; Blair et al. 2020).

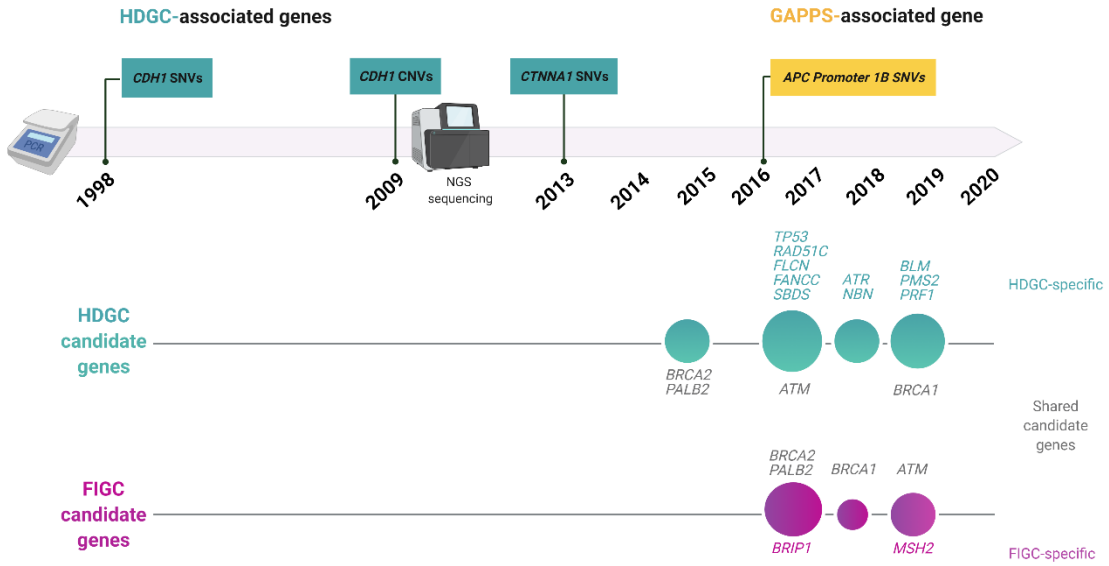


Figure 8. *Genetic timeline of bona fide and candidate genes for hereditary diffuse gastric cancer (HDGC), gastric adenocarcinoma and proximal polyposis of the stomach (GAPPS) and familial intestinal gastric cancer (FIGC). SNVs, single-nucleotide variants; CNVs, copy number variants.*

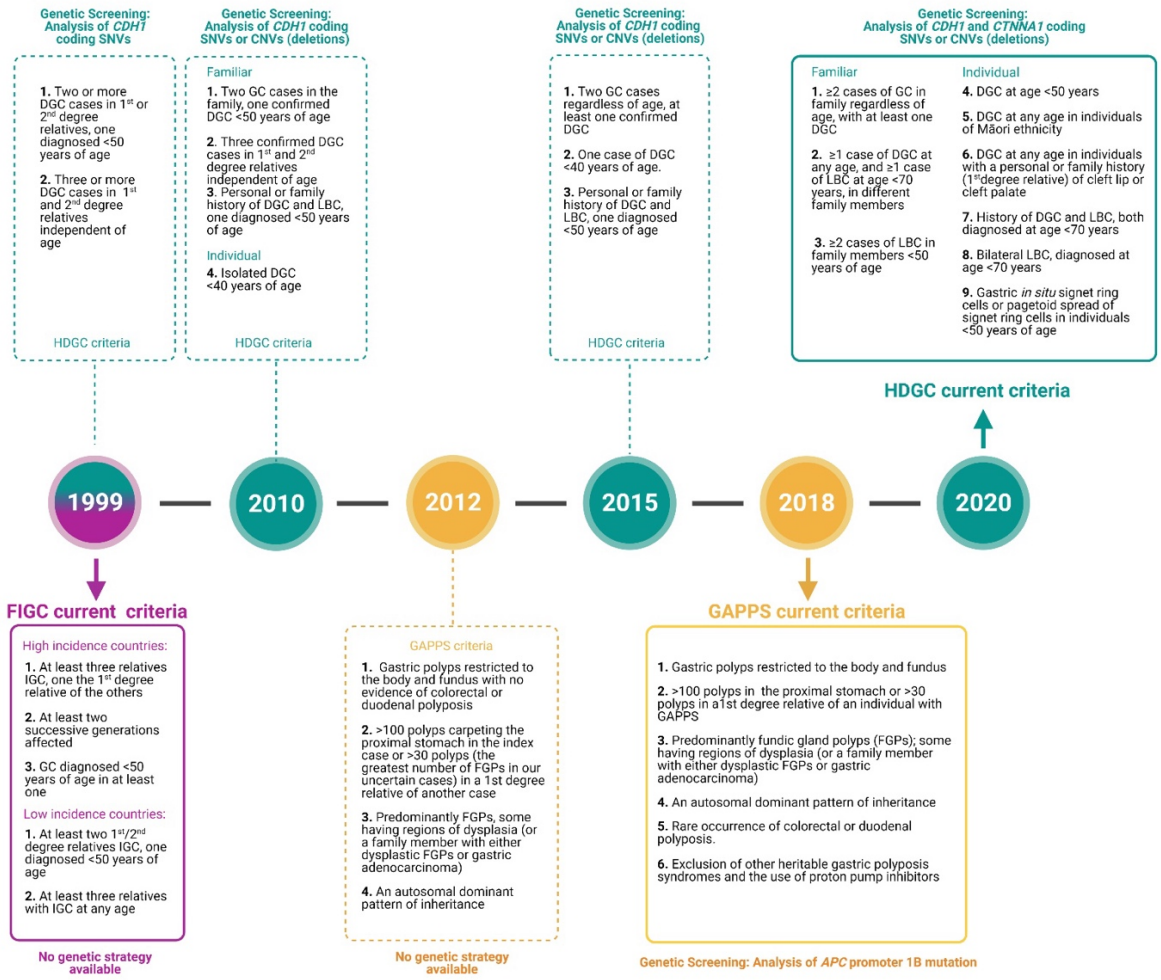


Figure 9. *Evolution of clinical criteria and genetic screening guidelines for the main GC-TRS: HDGC, GAPPS and FIGC. GC, gastric cancer; DGC, diffuse gastric cancer; LBC, lobular breast cancer; FGP(s), Fundic Gland Polyp(s); IGC, intestinal gastric cancer.*

4.1.1. Bonafide HDGC genes: CDH1 and CTNNA1

CDH1 / E-cadherin

The *CDH1* gene encodes for E-cadherin, a type I transmembrane glycoprotein, that mediates intercellular adhesion in epithelial mammalian tissues in a Ca^{2+} dependent manner (Harris and Tepass 2010)This protein is composed of a cytoplasmatic, a transmembrane and an extracellular domain (**Figure 10**). The extracellular domain establishes homophilic interactions with E-cadherin proteins in neighbouring cells, while the cytoplasmatic domain interacts with p120 and β-catenin proteins, promoting its stabilization (**Figure 10**). *CDH1*/E-cadherin loss of function in epithelial cells often promotes cell-to-cell adhesion impairment, increased migratory and invasive potential, increased survival, and altered signalling mainly related to the interaction with stroma and the tissue microenvironment (Shapiro et al. 1995; Aberle et al. 1996; Okamoto et al. 2005). Thus, E-cadherin loss of function has a close association with the infiltrative and invasive phenotypes observed in HDGC hallmark tumours (DGC and LBC) and the congenital malformations cleft lip/cleft palate (**Table 3**; (Aberle et al. 1996; Frebourg et al. 2006; Kaurah et al. 2007; Humar et al. 2009; Tsugeno et al. 2020; Green et al. 2022). Indeed, location of *CDH1* variants differs between HDGC and congenital malformations cleft lip/cleft palate, in which variants associated with the latter cluster in the linker between the extracellular domains (Cerrato-Izaguirre et al. 2022). The relationship of germline *CDH1* loss of function variants with DGC and LBC is also reminiscent from the sporadic context, where somatic *CDH1* loss of function is a common and determinant event for the initiation and development of DGC and LBC sporadic counterparts (Berx et al. 1995; Selvanathan et al. 2020).

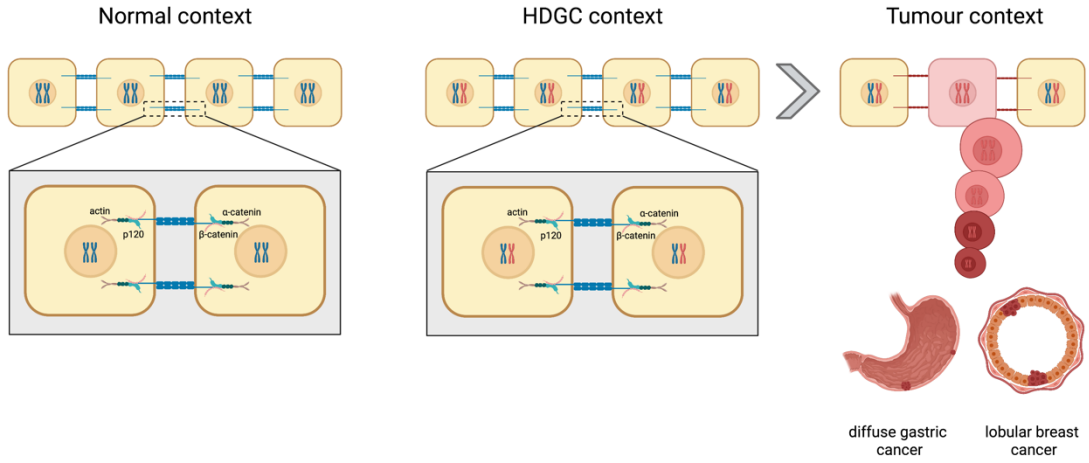


Figure 10. *Molecular characterization of normal stomach tissue, HDGC normal stomach tissue and HDGC gastric tumour.*

Heterozygous germline *CDH1* alterations, either SNVs or CNVs, occur in 10-40% of HDGC families, depending on the geographical background (Guilford et al. 1998; Oliveira et al. 2009a). Truncating SNVs in the nonsense-mediated decays (NMD) competent region of the gene (Karam et al. 2008), enclosing frameshift, nonsense and many splice-site variants, account for the majority of HDGC-

associated *CDH1* variants (**Table 3**; (Oliveira et al. 2015)). Large deletions, encompassing both 5'- and 3'-regulatory or coding regions of the *CDH1* gene were identified in HDGC families in 2009, adding a novel disease mechanism predisposing for HDGC (**Table 3**, **Figure 8**; (Oliveira et al. 2009a)). Truncating SNVs (upstream of the NMD-boundary), start loss variants and large deletions have a clear clinical significance, and are thus generally considered actionable in HDGC families (Garcia-Pelaez et al. 2023). These occur throughout the gene without any hotspot so far identified and follow the classical tumour suppressor gene model (**Table 3**). Despite reports of *CDH1* missense variants in HDGC, evidence of their associations is still scarce (Corso et al. 2022; Garcia-Pelaez et al. 2023). However, variants depicted in clinvar initially as missense have later been proved to impact splicing, generating premature stop codon (Barbosa-Matos et al. 2021)

Carriers of germline heterozygous *CDH1* causative variants remain healthy until a second event affects the remaining wild-type (WT) allele. Once the WT is targeted, through a second hit mechanism, and *CDH1* somatic expression is completely lost, in either the stomach or breast, a tumour may arise often at early age (Oliveira et al. 2015). Somatic *CDH1* promoter methylation, large deletions causing loss of heterozygosity (LOH) or a second SNV have been identified as second hit mechanisms in HDGC tumours (Grady et al. 2000; Barber et al. 2008; Oliveira et al. 2009b). The most frequent second inactivation hit in HDGC primary tumours is *CDH1* promoter hypermethylation, while LOH prevails in lymph node metastases (**Table 3**; (Oliveira et al. 2009b)).

Although academically established as an HDGC cause, *CDH1* variant type diversity remain a controversial territory in clinical practice. Therefore, to standardize *CDH1* variants' classification and to attribute a value to its actionability, Lee K and colleagues developed specific variant curation guidelines for the analysis of germline *CDH1* sequence variants, according to the ACMG/AMP criteria (**Table 3**) (Richards et al. 2015; Lee et al. 2018). This classification system allows allocating *CDH1* variants into five categories (benign - B, likely benign - LB, variant of unknown significance - VUS, likely pathogenic – LP, and pathogenic – P), and consider P or LP *CDH1* variants as actionable in the HDGC context (Lee et al. 2018; Blair et al. 2020). The definition of actionability of P/LP variants triggers a cascade that involves carrier genetic testing in extended families, surveillance, and/or prophylactic measures (Blair et al. 2020). Nevertheless, predicting VUS actionability, particularly of missense variants, is still incredibly challenging, due to the scarce clinical data available for this type of variants. Given the high frequency of missense variants found in families with HDGC criteria, it is important to understand which of them may induce cryptic splicing with severe consequences, such as premature truncation. This knowledge would highly likely contribute for better classification into more benign of pathogenic classes (Figueiredo et al. 2021; Barbosa-Matos et al. 2021).

Table 3. Molecular and clinical and characteristics of HDGC and GAPPS and their causal genes.

Syndrome	Gene	Spectrum	Main Molecular Variation	Accionable Variants	Hotspot	2nd hit	Penetrance			MANE Transcript (Refseq)	ACMG/AMP Criteria (Yes/No) [PMID Ref]
							GC in males	GC in females	LBC		
HDGC	*CDH1*	DGC, LBC, CLP	Truncating SNVs; large deletions	P/LP	No	Promoter methylation; LOH; 2nd SNV	42%-70% by 80 years of age	33%-55% by 80 years of age	39%-55% by 80 years of age	NM_004360.5	Yes
	CTNNA1	DGC	Truncating SNVs	P/LP	No	NA	NA		NA	NM_001903.5	No
GAPPS	*APC*	GC with Polyposis	Promoter SNVs	P/LP	Promoter 1b	NA	13%		NA	NM_000038.6	No
References (PMID):						19269290 10973239	26182300 31246251 21813476	26182300 31246251	26182300 11729114		30311375

Footnote: GC, Gastric Cancer; DGC, Diffuse Gastric Cancer; LBC, Lobular Breast Cancer; CLP, Clefts of the Lip and/or Palate; FGP(s), Fundic Gland Polyp(s); SNVs, single nucleotide variants; ACMG/AMP, American College of Medical Genetics and Genomics and the Association for Molecular Pathology; P, Pathogenic; LP, Likely Pathogenic; LOH, Loss of heterozygosity; NA, Not Available.

CDH1-associated disease risk

According to Hansford S. and colleagues (Hansford et al. 2015), *CDH1* variant carriers that met 2010 clinical criteria (Fitzgerald et al. 2010) had a GC cumulative incidence of 70% in males and 56% in females by 80 years of age (**Table 1**). Recently, Roberts and colleagues estimated the cumulative risk to develop GC by 80 years of age to be 42% in male carriers and 33% in female carriers (Roberts et al. 2019), of which 33% met the 2015 less stringent clinical criteria (**Table 3**) (Van Der Post et al. 2015). Estimations of cumulative risk to develop LBC range between 39% and 55% in females carrying *CDH1* variants (**Table 3**) (Pharoah et al. 2001; Hansford et al. 2015; Roberts et al. 2019; Xicola et al. 2019). The cumulative risk differs from study to study, and largely depends on whether families included in the risk calculation were ascertained or non-ascertained, if these arose in countries with high or low incidence of GC and/or breast cancer (BC), were exposed to different modifier genetic and environmental factors, and importantly on the set of *CDH1* variants under study. This is quite evident from the Roberts et al study, which has shown that the disease associated risk decreased when non-ascertained cohorts of *CDH1* pathogenic variant carriers were used (Roberts et al. 2019).

In summary, carriers of *CDH1* P/LP variants display high, but variable, DGC and/or LBC associated risk, and the factors that underlie differential risks to develop DGC, LBC or both remain elusive.

CTNNA1 / α-E-catenin

The *CTNNA1* gene encodes for the αE-catenin protein, which plays a role in adherens junctions (**Figure 10**). E-cadherin and αE-catenin act together with p120 and β-catenin, which are αE-catenin binding partners, to mediate the cytoplasmic linkage with the actin cytoskeleton (**Figure 10**) (van Roy 2014; Takeichi 2018).

Lobo et al. recently reviewed the literature describing the occurrence of germline *CTNNA1* variants in HDGC families (Lobo et al. 2021). This study concluded that 13 HDGC families have been described to carry *CTNNA1* germline variants (Majewski et al. 2013; Hansford et al. 2015; van der Post et al. 2018; Weren et al. 2018; Benusiglio et al. 2019; Clark et al. 2020), 10 of which classified as P or LP, according to the ACMG/AMP guidelines for *CDH1* variant classification in HDGC (Lee et al. 2018). This literature review also suggested that P, but not LP variants in *CTNNA1* seem to associate with DGC, accounting as a disease cause for <2% of families/individuals fulfilling HDGC criteria (Lobo et al. 2021). Considering that HDGC is very rare in the population, germline pathogenic variants in *CTNNA1* are even rarer, but quite well supported by clinical and molecular evidence.

Given the above evidence, *CTNNA1* variant analysis has been introduced in the 2020 HDGC clinical testing guidelines, and the gene considered as a bona fide HDGC predisposing gene (**Figure 9**; (Blair et al. 2020)). Supporting this decision is the fact that asymptomatic *CTNNA1* variant carriers were found to display DGC foci in prophylactic total gastrectomy specimens, raising the hypothesis that *CDH1* and *CTNNA1* pathogenic variants may have a similar behaviour for DGC development (Benusiglio et al. 2019). Thus, families that fulfil HDGC clinical criteria and lack a *CDH1* P/LP SNV or CNV are now advised to undergo *CTNNA1* testing (**Figure 9**; (Blair et al. 2020)). *CTNNA1*-positive families described to date display mainly DGC, and a clear association with LBC and cleft lip, with or without cleft palate, has not been reported (**Table 3**).

Classification using ACMG criteria is not available for *CTNNA1* variants, thus, they have been mainly classified using the general ACMG criteria (**Table 3**) (Richards et al. 2015; Lobo et al. 2021). The lack of specific ACMG guidelines for classification of these variants occurring in HDGC patients and families complicates interpretation and prediction of actionability. Most of the germline variants carried by HDGC families, reported to date, were truncating, including frameshift and nonsense variants (**Table 3**) (Majewski et al. 2013; Hansford et al. 2015; Weren et al. 2018; Benusiglio et al. 2019; Clark et al. 2020). Data on *CTNNA1*-disease associated risk are still scarce, yet first estimations suggest a risk of developing DGC by 80 years of age ranging between 49% and 57% (Coudert et al. 2022). To the best of our knowledge LBC estimations have not been performed due to the lack of LBC cases in *CTNNA1* carriers.

4.1.2. Management of HDGC families carrying CDH1 and CTNNA1 causative variants

Individuals at risk of HDGC development are clinically managed according to the 2020 HDGC clinical guidelines (Blair et al. 2020). Upon identification of a *CDH1* causal variant, carriers are offered risk-reduction measures available to prevent cancer. These include prophylactic removal of the stomach, regardless of endoscopic results, in carriers older than 18 years old, for DGC prevention in both

genders; and bilateral risk-reducing mastectomy for LBC prevention in females (Blair et al. 2020). Considering that both DGC and LBC may arise at early age in *CDH1* variant carriers, prophylactic surgeries may be considered between age 20 and 30 (van der Post et al. 2015). However, risk-reduction gastrectomy entails initial weight loss, bile reflux gastritis controlled with sucralfate, dumping syndrome and life-long calcium supplementation for improved bone mineral density (Forrester et al. 2022; Gamble et al. 2023).

CDH1 variant carriers may also undergo DGC and LBC surveillance by endoscopic/multiple biopsies and/or breast magnetic resonance imaging (MRI), while waiting for prophylactic surgeries, if unfit or upon refusal of prophylactic interventions. Systematic DGC endoscopic surveillance with random biopsies yields a 50-67% diagnostic yield (Lee et al. 2023; Asif et al. 2023), while endoscopic surveillance targeted biopsies 25-26% detection rate (Mi et al. 2018; Lee et al. 2023). Yet, endoscopy screening yields a different detection rate according to expert consensus and single institution experience (Asif et al. 2023). The clinical recommendations for surveillance by breast MRI begin at age 30 (Blair et al. 2020).

For *CTNNA1*-positive families, given the current lack of knowledge, surveillance is favoured over prophylactic surgery for DGC in variant carriers. As LBC predisposition due to *CTNNA1* germline variants has not been documented, BC surveillance in asymptomatic carriers should be considered on a case-by-case basis (Blair et al. 2020).

CDH1-positive HLBC families are advised to undergo annual breast surveillance with MRI or bilateral risk-reducing mastectomy, and may be recategorized into HDGC, if a DGC develops or a DGC precursor lesion is identified during surveillance in a carrier family member (Blair et al. 2020).

4.1.3. Unknowns: CDH1-related HDGC

Although the knowledge on the implications of bearing a germline *CDH1* variant has grown considerably, since the first reports of their causality (Guilford et al. 1998), many uncertainties remain about penetrance, variant actionability, and the spectrum of the disease in families carrying different *CDH1* germline variants. Most of the studies performed to date to understand penetrance in *CDH1* carrier families (Pharoah et al. 2001; Hansford et al. 2015; Roberts et al. 2019; Xicola et al. 2019), are biased towards the previous IGCLC criteria that mainly focused on risk for DGC. The exception is the study performed by Roberts et al. 2019, which predicted penetrance using an unascertained cohort, and demonstrated that disease risk for DGC may be lower than initially predicted using ascertained cohorts (van der Post et al. 2015; Roberts et al. 2019). However, several aspects related with penetrance in *CDH1* carrier families remain elusive, namely penetrance of confirmed DGC and LBC, penetrance modifiers and variant-specific penetrance.

The *CDH1*-associated disease-spectrum, beyond DGC and LBC also remains unclear, as some apparently HDGC-unrelated early onset cancers may also develop in carriers. While intestinal-type GC has been completely excluded from the *CDH1* spectrum (Caldas et al. 1999; Van Der Post et al. 2015), an association with colorectal cancer (CRC) and ovarian cancer (OC) has yet to be acknowledged (Richards et al. 1999; Oliveira et al. 2002; Stuebs et al. 2018; Roberts et al. 2019; Figueiredo et al. 2019).

A few genotype-phenotype association studies have been performed, considering the molecular classification of *CDH1* variants (Lo et al. 2019; Selvanathan et al. 2020). Nonetheless, a clear correlation of cancer and non-cancer phenotypes with variant molecular types remain to be performed.

CDH1 missense variants are often identified in HDGC families, but their functional consequence is hard to predict, and therefore their potential causal effect and actionability in HDGC remain a matter of debate. Missense variants do not affect the size of the E-cadherin protein, and many times not even the protein localization, however, E-cadherin-mediated signalling seems to be modified in many instances (Melo et al. 2017; Lee et al. 2018) Whether the latter consequence is sufficient to predispose to HDGC remains unknown, and only genotype-phenotype directed to families and individuals carrying these variants, will shed light into their potential causality. To date, it has been reported that missense variants seem to be enriched in HLBC cases (Corso et al. 2018).

Recently, genetic modifiers, epigenetic alterations, or environmental factors have been suggested to play a role in hereditary cancer predisposition (Taeubner et al. 2018). However, how specific genetic and environmental factors, or their combination, contribute to increase HDGC predisposition, severity or disease spectrum remains unexplained.

Evidence of drug targetable approaches in E-cadherin deficient cells may pave the way for targeted-therapy or chemo-prevention in HDGC patients (Beetham et al. 2019; Brew et al. 2021). Druggable vulnerabilities in E-cadherin deficient cells include allosteric AKT inhibitors (Bougen-Zhukov et al. 2019, 2022) and HDAC inhibitors (Decourtye-Espiard et al. 2021).

H. pylori and *Epstein-Barr* virus infections have been reported to be shared among individuals in the same environment, and sometimes from the same family (Ellison-Loschmann et al. 2017), however, their contribution to HDGC remains understudied.

4.1.4. Unknowns: CTNNA1-related HDGC

Knowledge on *CTNNA1*-associated HDGC is still very scarce, therefore most of what has been realized for *CDH1*-associated HDGC is yet to be discovered in *CTNNA1*-HDGC. *CTNNA1* carriers are predisposed to DGC, but whether LBC and cleft lip with or without cleft palate are also part of the disease spectrum remains unknown. Disease-risk estimations, variant actionability, genotype-

phenotype correlations, and general approaches for surveillance and prophylactic surgery are all currently missing items in knowledge about *CTNNA1*-associated HDGC.

4.2. Gastric adenocarcinoma and proximal polyposis of the stomach (GAPPS)

GAPPS is a GC-predisposing syndrome with an autosomal dominant inheritance pattern with tropism to the proximal stomach (Worthley et al. 2012).

Predisposition to develop gastric polyps distinguishes GAPPS from other GC syndromes. The spectrum of disease includes fundic gland polyps (FGP) – with hyperplastic, adenomatous and mixed-type features with hyperproliferative aberrant pits –, foveolar-type dysplasia of variable grades, intestinal-type and foveolar-type adenomas and intestinal-type and mixed-type gastric adenocarcinoma (Worthley et al. 2012; De Boer et al. 2018; Gullo et al. 2021)

Fundic gland polyposis with multifocal dysplasia have been detected as early as 10 years old, however, GC age of onset ranges from 23 to 75 years (Worthley et al. 2012; Repak et al. 2016). GAPPS families display a heterogeneous phenotype, mainly in the number of polyps, that ranges between 30 to hundreds (Worthley et al. 2012; Rudloff 2018). Intestinal and mixed type GC have been detected in 13% of patients from families with a GAPPS diagnosis (**Table 1**).

Current diagnostic criteria for GAPPS are depicted in **Figure 2** (Rudloff 2018). The most recent updates in clinical criteria exclude other heritable gastric polyposis syndromes, and individuals that use proton pump inhibitors, which induce FGP per se. GAPPS has been considered a variant of familial adenomatous polyposis (FAP) which is restricted to the stomach, although recently colorectal polyps have also been described in patients from GAPPS families (Provenzale et al. 2016). Therefore, a diagnosis of FAP or attenuated FAP (AFAP) should be ruled out with the absence of polyposis in the remaining gastrointestinal tract (Rudloff 2018).

GAPPS has been shown to be caused by SNVs at the *APC* promoter 1B, and for this reason *APC* promoter 1B is currently tested upon clinical suspicion of GAPPS (**Figure 1**). To date, germline variants at the *APC* promoter 1B hotspot are considered the main, and only known, cause of GAPPS.

4.2.1. Bonafide GAPPS genes: APC

The *APC* gene encodes the Adenomatous Polyposis Coli protein, a negative regulator of β-catenin and, thus, of the WNT signalling pathway. *APC* downregulation leads to an increase in cytoplasmatic β-catenin, that is translocated to the nucleus, promoting increased proliferation (Ilyas and Tomlinson, 1997).

Germline SNVs in *APC*, with a hotspot in promoter 1B (YY1 Transcription Factor Binding motif), have been identified as the main trigger to develop GAPPS (**Table 1**; (Li et al. 2016)). In stomach, promoter 1B has ~15-fold higher transcription than promoter 1A, due to methylation of the latter in cells from the gastric mucosa (Hosoya et al. 2009; Rohlin et al. 2011). Therefore, lack of YY1 transcription factor binding may lead to reduced promoter 1B activity and reduced *APC* tumour suppressive phenotype in the stomach. In the colon, there is normal promoter 1A-driven transcription, that compensates for the lack of promoter 1B-driven transcription. This provides the rational for a GAPPS-restricted, or at least preferential, gastric phenotype (Rudloff 2018). However, identical promoter 1B variants have been identified in <1% of classical FAP families with both gastric and duodenal involvement (Snow et al. 2015).

In 2016, Li J. and colleagues found three germline SNVs and one small deletion in *APC* promoter 1B in an Australian family (c.-195A>C co-segregated with c.-125delA), an American family (c.-191T>C) and a Canadian family (c.-192A>G) (Li et al. 2016). These variants were considered P/LP and accepted as disease causing (Landrum et al. 2014). A combination of variants at the *APC* promoter 1B (e.g., c.-195A>C and c.-125delA) was found associated with severe phenotypes, but their individual contribution remains unknown (Li et al. 2016). The variant c.-191T>C was found in 12 additional families from different ancestries, indicating a potential hotspot (Repak et al. 2016; Mitsui et al. 2018; Foretova et al. 2019; Kanemitsu et al. 2021). GAPPS-associated phenotypes and disease expressivity may be heterogeneous, even among families carrying the same causal variant. While some patients may develop very early onset GC (Repak et al. 2016), old carriers remain disease free (Foretova et al. 2019) . Supporting incomplete penetrance is the fact that patients from 2nd and 3rd generations present with severe phenotypes, while 1st generation obligate carriers remain disease free (Worthley et al. 2012). Altogether, these observations suggest that genetic modifiers or environmental factors may play a role in GAPPS cancer predisposition and penetrance.

Currently, ACMG clinical criteria for *APC* variants classification is only available for FAP syndrome (**Table 1**; (Hegde et al. 2014)).

4.2.2. Management: GAPPS

Given the low number of GAPPS cases reported and patients heterogeneity, management of GAPPS families is mostly carried out on a case-by-case basis (Tacheci et al. 2021). Endoscopic surveillance with biopsies is recommended and prophylactic gastrectomy is offered to GAPPS families without a clear timing guideline (Worthley et al. 2012; Repak et al. 2016; Li et al. 2016). Nevertheless, prolonged surveillance in FGP patients is not recommended, due to a potential rapid progression. Further, first-degree relatives of families diagnosed with GAPPS, carrying the disease-associated genotype, are advised to perform upper

gastrointestinal endoscopy and a colonoscopy (Worthley et al. 2012; Rudloff 2018).

4.2.3. Unknowns: APC-related GAPPS

Twenty-five families, encompassing 110 patients that fulfilled GAPPS criteria, have been reported across the globe, namely from Australia, North America, Europe, and Japan (Worthley et al. 2012; Yanaru-Fujisawa et al. 2012; Declich et al. 2013; Repak et al. 2016; Li et al. 2016; Beer et al. 2017; Rudloff 2018; Mitsui et al. 2018; Foretova et al. 2019; Kunovsky et al. 2019; Kanemitsu et al. 2021; Tacheci et al. 2021). The number of polyposis-affected patients in GAPPS families has likely been underestimated due to a bias in surveillance and diagnosis towards the stomach (Tacheci et al. 2021), and because colonic polyposis cases were not considered as part of the GAPPS disease spectrum (Li et al. 2016). For this reason, the prevalence and penetrance of GAPPS-related phenotypes remains fairly unknown (Yen et al. 1993; Li et al. 2016; Tacheci et al. 2021).

Although the somatic landscape of GAPPS has started to be disclosed, more studies are needed to distinguish between sporadic FGPs and GAPPS-associated FGPs (Lagarde et al. 2010; Kadiyska et al. 2014; Pavicic et al. 2014).

Even though H. pylori infection is thought to be absent in GAPPS patients, sparse data is available highlighting the lack of knowledge regarding the role of this and other environmental factors in tumour development in the context of this disease (Worthley et al. 2012; Rudloff 2018; Tacheci et al. 2021).

4.3. Familial intestinal gastric cancer (FIGC)

FIGC has been generally understood as an autosomal dominant disease with one or several family members affected by intestinal type GC (IGC) (Caldas et al. 1999). However, no global consensus has been reached regarding clinical criteria for this disease. It is generally accepted that criteria analogous to the Amsterdam criteria for hereditary non-polyposis colorectal cancer, are used in countries where GC incidence is high (e.g., Japan, China, Korea, Portugal). These criteria become more restrictive in countries with low GC incidence (e.g., USA, Canada, UK) (Caldas et al. 1999), as depicted in **Figure 9**.

Whist germline variants at *CDH1*, *CTNNA1* and *APC* genes are well established as the main triggers to develop HDGC and GAPPS, FIGC remains almost completely genetically unexplained, with unknown tumour spectrum and age of onset. Moreover, given the lack of FIGC-causal genes, clinical criteria have not been validated nor revisited since 1999 (Caldas et al. 1999).

To date, rare patients or families suspected of FIGC or displaying isolated IGC at early age have been shown to display germline potentially causative variants in

BRCA2, *BRCA1*, *BRIP1* and *MSH2* (Lee et al. 2000; Sahasrabudhe et al. 2017; Slavin et al. 2017; Weren et al. 2018; Ichikawa et al. 2018; Tedaldi et al. 2019; Carvalho et al. 2021). However, both the diversity of genes apparently involved, and the fact that these genes predispose to different diseases, has hampered the classification of these candidate genes as FIGC-associated.

A recent study pinpointed FIGC as an independent clinical and molecular entity, differing at the molecular level, from of both sporadic IGC and HDGC (Carvalho et al. 2021). The average age of onset of IGC in FIGC families was shown to be considerably higher than that of HDGC, but at least 10 years lower than that of sporadic IGC. In addition, the germline, and somatic mutational landscapes of FIGC patients and tumours was shown to be distinctive from those of sporadic IGC. Specifically, FIGC probands carried far more *TP53* germline common variants than sporadic IGC cases; and far more somatic common variants in *BRCA2*, *ATM*, *FOXF1*, *FHIT*, *SDHB*, *MSH6*, *CTNNA1* and *PXN*. Also, the MSI frequency was significantly higher in FIGC tumours, as compared to the sporadic counterpart (Carvalho et al. 2021). The tumour spectrum of FIGC families was shown to be broad, predisposing primarily to the stomach, but very often to CRC and BC (Carvalho et al. 2021). As high penetrance gene defects were rare in FIGC, Carvalho et al suggested FIGC as a polygenic disease (Carvalho et al. 2021), a thesis that needs further research to be proven.

Altogether, these findings claim for a proposal to redefine FIGC clinical criteria. Carvalho and colleagues suggested that a minimum of two cases of GC diagnosed at any age, with one GC being histologically confirmed IGC, would be needed for clinical confirmation of FIGC (Carvalho et al. 2021).

4.3.1. Management

Due to the scarce information about genetic causes and disease spectrum, management of FIGC patients is not clear. Notwithstanding, Sereno and colleagues recommended periodic endoscopy in first degree relatives, starting 10 years earlier than the age of onset of the youngest family member diagnosed with IGC (Sereno et al. 2011). According to Corso et al 2013, yearly gastroduodenoscopy starting at 40 years of age or 5 years earlier than the age of onset of the youngest family member diagnosis, could also be accepted. Given the high rate of H. pylori infection, eradication is recommended in these families (Corso et al. 2013).

4.3.2. Unknowns: FIGC

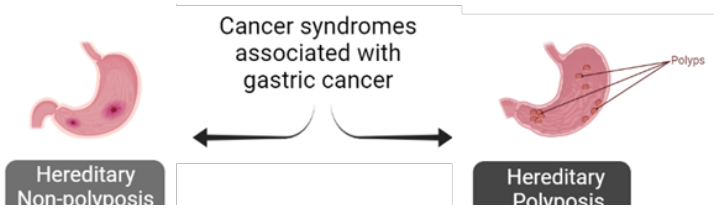
FIGC remains vastly unknown, and genetically unresolved to date. Consequently, no objective and targeted clinical guidelines or criteria for testing can be designed. Prevalence and penetrance are unknown, and management is

suboptimal. The tumour spectrum also remains unclear, and no consensus exists on the most probable age of onset of IGC in the context of FIGC, so that cancer screening and surveillance can be optimized.

4.4. Inherited gastric cancer in other cancer-associated syndromes

GC and further gastric lesions can still be observed in other inherited cancer disorders as part of their tumour spectrum such as hereditary breast and ovarian cancer (HBOC), Li–Fraumeni syndrome (LFS), Lynch syndrome (LS), familial adenomatous polyposis (FAP), juvenile polyposis syndrome (JPS), MUTYH-associated polyposis (MAP), Peutz-Jeghers syndrome (PJS) and PTEN hamartoma tumour syndrome (PHTS) (**Table 4**). Despite not being the primary cancer, a noticeable clustering of GC cases has been observed in subsets of families with these TRS. The lifetime risk for developing GC in these heritable syndromes is also highly variable. Based on the syndromes presented in **Table 4**, the risk for GC can be elevated for JPS, PJS and LS, ranging from 10 to nearly 30%. On the other hand, current literature suggests that GC is likely to occur rather infrequently in FAP, LFS, MAP, PHTS or HBOC, and be largely dependent on environmental risk factors, linked or not with geographical background (Setia et al. 2015). This has recently been described for LFS, where a comparison of cancer spectrum in TP53 germline carrier individuals from Japan and France revealed stomach cancer being significantly more frequent in Japan than in France (7.8% vs 1.3%) (Funato et al. 2021). Besides distinct disease spectra and mutated genes, the geographic origin of the population where these TRS occur, also affects the GC risk (Capelle et al. 2010; Ikenoue et al. 2019). GC surveillance measures may be stimulated in cases where GC occurs more frequently than expected in families diagnosed with one of the above described TRS, or if patients and their relatives (affected by a TRS above) live in a country with high GC incidence (Vasen et al. 2013).

Table 4. *TRS without primary predisposition to GC but presenting GC as part of their disease spectrum.*



	Syndrome	Inheritance	Cancer Spectrum	Major Associated Genes	References (PMID)
Hereditary Non-polyposis syndromes	Hereditary breast and ovarian cancer syndrome (HBOC)	AD	Breast, ovarian, prostate, pancreatic and gastric cancer	*BRCA1* *BRCA2*	32259785 16369438 12373604
	Li-Fraumeni syndrome (LFS)	AD	Soft tissue sarcoma, female breast cancer, osteosarcoma, CNS tumor, gastric, lung, and adrenocortical cancers; schoroid plexus carcinoma, leukemia and rhabdomyosarcoma of embryonal anaplastic subtype	*TP53*	32457520 32259785 3409256 26014290 21552135
	Lynch syndrome (LS)	AD	Colorectal, endometrial, ovarian, gastric, pancreatic, urothelial, pancreatic, brain(usually glioblastoma), biliary tract and small-bowel cancers; sebaceous adenomas/carcinomas and keratoacanthomas	*MLH1* *PMS2* *MSH2* *EPCAM* *MSH6*	32930401 31487681 28754778
Hereditary Polyposis syndromes	Familial adenomatous polyposis (FAP)	AD	Colorectal, duodenal, gastric, and thyroid cancer; desmoid tumors, osteomas, odontomas, hepatoblastomas and medulloblastomas	*APC*	28185118 1316858 2896968 *NCCN 2020
	Juvenile polyposis syndrome (JPS)	AD	Colorectal, gastric, small intestine and pancreatic cancer	*SMAD4* *BMPR1A*	31108454 25645574 *NCCN 2020
	MUTYH-associated adenomatous polyposis (MAP)	AR	Colorectal and duodenal cancer; gastric polyps and gastric cancer	*MUTYH*	19732775 27194394 *NCCN 2020
	Peutz-Jeghers syndrome (PJS)	AD	Colorectal, gastric, breast, small intestine, pancreatic, ovarian, uterus, cervical, and lung cancer	*STK11*	24143323 20051941 20581245 *NCCN 2020
	PTEN hamartoma tumor syndrome (PHTS)	AD	Breast, endometrial and follicular thyroid cancer; gastrointestinal hamartomas or ganglioneuromas, colon cancer, lipomas, renal cell carcinoma and melanoma	*PTEN*	32533092 32259785

Footnote: AD, Autosomal Dominant; AR, Autosomal Recessive; CNS, Central Nervous System; NCCN 2020 (National Comprehensive Cancer Network, 2021).

5. References

Aberle H, Schwartz H, Kemler R (1996) Cadherin-catenin complex: protein interactions and their implications for cadherin function. J Cell Biochem 61:514–23. https://doi.org/10.1002/(SICI)1097-4644(19960616)61:4%3C514::AID-JCB4%3E3.0.CO;2-R

Ahituv N, Zhu Y, Visel A, et al (2007) Deletion of ultraconserved elements yields viable mice. PLoS Biol 5:1906–1911. https://doi.org/10.1371/JOURNAL.PBIO.0050234

Ajani JA, Lee J, Sano T, et al (2017) Gastric adenocarcinoma. Nat Rev Dis Primers 3:. https://doi.org/10.1038/NRDP.2017.36

Allou L, Balzano S, Magg A, et al (2021) Non-coding deletions identify Maenli lncRNA as a limb-specific En1 regulator. Nature 592:93–98. https://doi.org/10.1038/S41586-021-03208-9

Amin MB, Greene FL, Edge SB, et al (2017) The Eighth Edition AJCC Cancer Staging Manual: Continuing to build a bridge from a population-based to a more "personalized" approach to cancer staging. CA Cancer J Clin 67:93–99. https://doi.org/10.3322/CAAC.21388

Anderson E, Devenney PS, Hill RE, Lettice LA (2014) Mapping the Shh long-range regulatory domain. Development 141:3934–3943. https://doi.org/10.1242/DEV.108480

Andrey G, Mundlos S (2017) The three-dimensional genome: regulating gene expression during pluripotency and development. Development 144:3646–3658. https://doi.org/10.1242/DEV.148304

Arnold CD, Gerlach D, Stelzer C, et al (2013) Genome-wide quantitative enhancer activity maps identified by STARR-seq. Science 339:1074–1077. https://doi.org/10.1126/SCIENCE.1232542

Asif B, Sarvestani AL, Gamble LA, et al (2023) Cancer surveillance as an alternative to prophylactic total gastrectomy in hereditary diffuse gastric cancer: a prospective cohort study. Lancet Oncol 24:383–391. https://doi.org/10.1016/S1470-2045(23)00057-8

Assumpção PP, Barra WF, Ishak G, et al (2020) The diffuse-type gastric cancer epidemiology enigma. BMC Gastroenterol 20:. https://doi.org/10.1186/S12876-020-01354-4

Bang YJ, Van Cutsem E, Feyereislova A, et al (2010) Trastuzumab in combination with chemotherapy versus chemotherapy alone for treatment of HER2-positive advanced gastric or gastro-oesophageal junction cancer (ToGA): a phase 3, open-label, randomised controlled trial. Lancet 376:687–697. https://doi.org/10.1016/S0140-6736(10)61121-X

Barber M, Murrell A, Ito Y, et al (2008) Mechanisms and sequelae of E-cadherin silencing in hereditary diffuse gastric cancer. J Pathol 216:295–306. https://doi.org/10.1002/PATH.2426

Barbosa-Matos R, Silva RL, Garrido L, et al (2021) The *CDH1* c.1901C>T Variant: A Founder Variant in the Portuguese Population with Severe Impact in mRNA Splicing. Cancers (Basel) 13:. https://doi.org/10.3390/CANCERS13174464

Bass AJ, Thorsson V, Shmulevich I, et al (2014) Comprehensive molecular characterization of gastric adenocarcinoma. Nature 513:202–209. https://doi.org/10.1038/NATURE13480

Beagan JA, Duong MT, Titus KR, et al (2017) YY1 and CTCF orchestrate a 3D chromatin looping switch during early neural lineage commitment. Genome Res 27:1139–1152. https://doi.org/10.1101/GR.215160.116/-/DC1

Beer A, Streubel B, Asari R, et al (2017) Gastric adenocarcinoma and proximal polyposis of the stomach (GAPPS) - a rare recently described gastric polyposis syndrome - report of a case. Z Gastroenterol 55:1131–1134. https://doi.org/10.1055/s-0043-117182

Beetham H, Chen A, Telford BJ, et al (2019) A high-throughput screen to identify novel synthetic lethal compounds for the treatment of E-cadherin-deficient cells. Sci Rep 9:. https://doi.org/10.1038/S41598-019-48929-0

Benusiglio PR, Colas C, Guillerm E, et al (2019) Clinical implications of *CTNNA1* germline mutations in asymptomatic carriers. Gastric Cancer 22:899–903. https://doi.org/10.1007/S10120-018-00907-7

Berx G, Cleton-Jansen AM, Nollet F, et al (1995) E-cadherin is a tumour/invasion suppressor gene mutated in human lobular breast cancers. EMBO J 14:6107–6115. https://doi.org/10.1002/J.1460-2075.1995.TB00301.X

Bianco S, Lupiáñez DG, Chiariello AM, et al (2018) Polymer physics predicts the effects of structural variants on chromatin architecture. Nat Genet 50:662–667. https://doi.org/10.1038/S41588-018-0098-8

Blair VR, McLeod M, Carneiro F, et al (2020) Hereditary diffuse gastric cancer: updated clinical practice guidelines. Lancet Oncol 21:e386–e397. https://doi.org/10.1016/S1470-2045(20)30219-9

Boija A, Klein IA, Sabari BR, et al (2018) Transcription Factors Activate Genes through the Phase-Separation Capacity of Their Activation Domains. Cell 175:1842-1855.e16. https://doi.org/10.1016/J.CELL.2018.10.042

Bougen-Zhukov N, Decourtye-Espiard L, Mitchell W, et al (2022) E-Cadherin-Deficient Cells Are Sensitive to the Multikinase Inhibitor Dasatinib. Cancers (Basel) 14:. https://doi.org/10.3390/CANCERS14071609

Bougen-Zhukov N, Nouri Y, Godwin T, et al (2019) Allosteric AKT Inhibitors Target Synthetic Lethal Vulnerabilities in E-Cadherin-Deficient Cells. Cancers (Basel) 11:. https://doi.org/10.3390/CANCERS11091359

Brew T, Bougen-Zhukov N, Mitchell W, et al (2021) Loss of E-Cadherin Leads to Druggable Vulnerabilities in Sphingolipid Metabolism and Vesicle Trafficking. Cancers (Basel) 14:. https://doi.org/10.3390/CANCERS14010102

Buenrostro JD, Giresi PG, Zaba LC, et al (2013) Transposition of native chromatin for fast and sensitive epigenomic profiling of open chromatin, DNA-binding proteins and nucleosome position. Nat Methods 10:1213–1218. https://doi.org/10.1038/NMETH.2688

Buenrostro JD, Wu B, Chang HY, Greenleaf WJ (2015) ATAC-seq: A Method for Assaying Chromatin Accessibility Genome-Wide. Curr Protoc Mol Biol 109:21.29.1-21.29.9. https://doi.org/10.1002/0471142727.MB2129S109

Bulger M, Groudine M (2011) Functional and mechanistic diversity of distal transcription enhancers. Cell 144:327–339. https://doi.org/10.1016/J.CELL.2011.01.024

Caldas C, Carneiro F, Lynch HT, et al (1999) Familial gastric cancer: overview and guidelines for management. J Med Genet 36:873–880

Capelle LG, Van Grieken NCT, Lingsma HF, et al (2010) Risk and epidemiological time trends of gastric cancer in Lynch syndrome carriers in the Netherlands. Gastroenterology 138:487–92. https://doi.org/10.1053/j.gastro.2009.10.051

Carneiro F, Seixas M, Sobrinho-Simões M (1995) New elements for an updated classification of the carcinomas of the stomach. Pathol Res Pract 191:571–584. https://doi.org/10.1016/S0344-0338(11)80878-2

Carvalho J, Oliveira P, Senz J, et al (2021) Redefinition of familial intestinal gastric cancer: clinical and genetic perspectives. J Med Genet 58:1–11. https://doi.org/10.1136/JMEDGENET-2019-106346

Cats A, Jansen EPM, van Grieken NCT, et al (2018) Chemotherapy versus chemoradiotherapy after surgery and preoperative chemotherapy for resectable gastric cancer (CRITICS): an international, open-label, randomised phase 3 trial. Lancet Oncol 19:616–628. https://doi.org/10.1016/S1470-2045(18)30132-3

Cerrato-Izaguirre D, Chirino YI, Prada D, et al (2022) Somatic Mutational Landscape in Mexican Patients: *CDH1* Mutations and chr20q13.33 Amplifications Are Associated with Diffuse-Type Gastric Adenocarcinoma. Int J Mol Sci 23:. https://doi.org/10.3390/IJMS231911116

Chiocchetti A, Tolosano E, Hirsch E, et al (1997) Green fluorescent protein as a reporter of gene expression in transgenic mice. Biochim Biophys Acta 1352:193–202. https://doi.org/10.1016/S0167-4781(97)00010-9

Choi YJ, Kim N (2016) Gastric cancer and family history. Korean J Intern Med 31:1042–1053. https://doi.org/10.3904/KJIM.2016.147

Clark DF, Michalski ST, Tondon R, et al (2020) Loss-of-function variants in *CTNNA1* detected on multigene panel testing in individuals with gastric or breast cancer. Genet Med 22:840–846. https://doi.org/10.1038/S41436-020-0753-1

Cong L, Ran FA, Cox D, et al (2013) Multiplex genome engineering using CRISPR/Cas systems. Science 339:819–823. https://doi.org/10.1126/SCIENCE.1231143

Correa P, Haenszel W, Cuello C, et al (1975) A model for gastric cancer epidemiology. Lancet 2:58–60. https://doi.org/10.1016/S0140-6736(75)90498-5

Corso G, Figueiredo J, La Vecchia C, et al (2018) Hereditary lobular breast cancer with an emphasis on E-cadherin genetic defect. J Med Genet 55:431–441. https://doi.org/10.1136/JMEDGENET-2018-105337

Corso G, Magnoni F, Massari G, et al (2022) *CDH1* germline mutations in healthy individuals from families with the hereditary diffuse gastric cancer syndrome. J Med Genet 59:313–317. https://doi.org/10.1136/JMEDGENET-2021-108226

Corso G, Roncalli F, Marrelli D, et al (2013) History, pathogenesis, and management of familial gastric cancer:

original study of John XXIII's family. Biomed Res Int 2013:385132. https://doi.org/10.1155/2013/385132
Coudert M, Drouet Y, Delhomelle H, et al (2022) First estimates of diffuse gastric cancer risks for carriers of *CTNNA1* germline pathogenic variants. J Med Genet 59:1189–1195. https://doi.org/10.1136/JMG-2022-108740
Crawford GE, Holt IE, Whittle J, et al (2006) Genome-wide mapping of DNase hypersensitive sites using massively parallel signature sequencing (MPSS). Genome Res 16:123–131. https://doi.org/10.1101/GR.4074106
Cunningham D, Allum WH, Stenning SP, et al (2006) Perioperative chemotherapy versus surgery alone for resectable gastroesophageal cancer. N Engl J Med 355:11–20. https://doi.org/10.1056/NEJMOA055531
De Boer WB, Ee H, Kumarasinghe MP (2018) Neoplastic Lesions of Gastric Adenocarcinoma and Proximal Polyposis Syndrome (GAPPS) Are Gastric Phenotype. Am J Surg Pathol 42:1–8. https://doi.org/10.1097/PAS.0000000000000924
Declich P, Arrigoni GS, Omazzi B, et al (2013) Sporadic fundic gland polyps and proximal polyposis associated with gastric adenocarcinoma share a common antral G cell hyperplasia. Gut 62:1088–9. https://doi.org/10.1136/gutjnl-2012-304100
Decourtye-Espiard L, Bougen-Zhukov N, Godwin T, et al (2021) E-Cadherin-Deficient Epithelial Cells Are Sensitive to HDAC Inhibitors. Cancers (Basel) 14:. https://doi.org/10.3390/CANCERS14010175
Dekker J, Rippe K, Dekker M, Kleckner N (2002) Capturing chromosome conformation. Science 295:1306–1311. https://doi.org/10.1126/SCIENCE.1067799
Ellison-Loschmann L, Sporle A, Corbin M, et al (2017) Risk of stomach cancer in Aotearoa/New Zealand: A Māori population based case-control study. PLoS One 12:. https://doi.org/10.1371/JOURNAL.PONE.0181581
Ernst J, Kheradpour P, Mikkelsen TS, et al (2011) Mapping and analysis of chromatin state dynamics in nine human cell types. Nature 473:43–49. https://doi.org/10.1038/NATURE09906
Ferlay J, Colombet M, Soerjomataram I, et al (2021) Cancer statistics for the year 2020: An overview. Int J Cancer 149:778–789. https://doi.org/10.1002/IJC.33588
Figueiredo C, Garcia-Gonzalez MA, Machado JC (2013) Molecular pathogenesis of gastric cancer. Helicobacter 18 Suppl 1:28–33. https://doi.org/10.1111/HEL.12083
Figueiredo J, Melo S, Carneiro P, et al (2019) Clinical spectrum and pleiotropic nature of *CDH1* germline mutations. J Med Genet 56:199–208. https://doi.org/10.1136/JMEDGENET-2018-105807
Figueiredo J, Mercadillo F, Melo S, et al (2021) Germline *CDH1* G212E Missense Variant: Combining Clinical, In Vitro and In Vivo Strategies to Unravel Disease Burden. Cancers (Basel) 13:. https://doi.org/10.3390/cancers13174359
Fitzgerald RC, Hardwick R, Huntsman D, et al (2010) Hereditary diffuse gastric cancer: updated consensus guidelines for clinical management and directions for future research. J Med Genet 47:436–444. https://doi.org/10.1136/jmg.2009.074237
Foretova L, Navratilova M, Svoboda M, et al (2019) GAPPS - Gastric Adenocarcinoma and Proximal Polyposis of the Stomach Syndrome in 8 Families Tested at Masaryk Memorial Cancer Institute - Prevention and Prophylactic Gastrectomies. Klin Onkol 32:2S109-2S117. https://doi.org/10.14735/AMKO2019S109
Forrester JD, Foster D, Ford JM, et al (2022) Surgery for Hereditary Diffuse Gastric Cancer: Long-Term Outcomes. Cancers (Basel) 14:. https://doi.org/10.3390/CANCERS14030728
Frebourg T, Oliveira C, Hochain P, et al (2006) Cleft lip/palate and *CDH1*/E-cadherin mutations in families with hereditary diffuse gastric cancer. J Med Genet 43:138–142. https://doi.org/10.1136/jmg.2005.031385
Fu Y, Foden JA, Khayter C, et al (2013) High-frequency off-target mutagenesis induced by CRISPR-Cas nucleases in human cells. Nat Biotechnol 31:822–826. https://doi.org/10.1038/NBT.2623
Fuchs CS, Tomasek J, Yong CJ, et al (2014) Ramucirumab monotherapy for previously treated advanced gastric or gastro-oesophageal junction adenocarcinoma (REGARD): an international, randomised, multicentre, placebo-controlled, phase 3 trial. Lancet 383:31–39. https://doi.org/10.1016/S0140-6736(13)61719-5
Funato M, Tsunematsu Y, Yamazaki F, et al (2021) Characteristics of Li-Fraumeni Syndrome in Japan; A Review Study by the Special Committee of JSHT. Cancer Sci 112:2821–34. https://doi.org/10.1111/cas.14919
Gamble LA, Lopez R, Rajasimhan S, et al (2023) Micronutrient Supplementation and Bone Health After Prophylactic Total Gastrectomy in Patients With *CDH1* Variants. J Clin Endocrinol Metab. https://doi.org/10.1210/CLINEM/DGAD137
Gamble LA, Rossi A, Fasaye GA, et al (2022) Association Between Hereditary Lobular Breast Cancer Due to *CDH1* Variants and Gastric Cancer Risk. JAMA Surg 157:18–22. https://doi.org/10.1001/JAMASURG.2021.5118
Garcia-Pelaez J, Barbosa-Matos R, Gullo I, et al (2021) Histological and mutational profile of diffuse gastric cancer: current knowledge and future challenges. Mol Oncol 15:2841–2867. https://doi.org/10.1002/1878-0261.12948
Garcia-Pelaez J, Barbosa-Matos R, Lobo S, et al (2023) Genotype-first approach to identify associations between *CDH1* germline variants and cancer phenotypes: a multicentre study by the European Reference Network on Genetic Tumour Risk Syndromes. Lancet Oncol 24:91–106. https://doi.org/10.1016/S1470-2045(22)00643-X
Gasiunas G, Barrangou R, Horvath P, Siksnys V (2012) Cas9-crRNA ribonucleoprotein complex mediates specific DNA cleavage for adaptive immunity in bacteria. Proc Natl Acad Sci U S A 109:. https://doi.org/10.1073/PNAS.1208507109/-/DCSUPPLEMENTAL/PNAS.201208507SI.PDF
Gaszner M, Felsenfeld G (2006) Insulators: exploiting transcriptional and epigenetic mechanisms. Nat Rev Genet 7:703–713. https://doi.org/10.1038/NRG1925
Grady WM, Willis J, Guilford PJ, et al (2000) Methylation of the *CDH1* promoter as the second genetic hit in hereditary diffuse gastric cancer. Nat Genet 26:16–17. https://doi.org/10.1038/79120
Gravalos C, Jimeno A (2008) HER2 in gastric cancer: a new prognostic factor and a novel therapeutic target. Ann Oncol 19:1523–1529. https://doi.org/10.1093/ANNONC/MDN169
Green BL, Fasaye GA, Samaranayake SG, et al (2022) Frequent cleft lip and palate in families with pathogenic germline *CDH1* variants. Front Genet 13:. https://doi.org/10.3389/FGENE.2022.1012025
Guilford P, Hopkins J, Harraway J, et al (1998) E-cadherin germline mutations in familial gastric cancer. Nature 392:402–405. https://doi.org/10.1038/32918
Gullo I, van der Post RS, Carneiro F (2021) Recent advances in the pathology of heritable gastric cancer syndromes. Histopathology 78:125–147. https://doi.org/10.1111/HIS.14228
Haberle V, Stark A (2018) Eukaryotic core promoters and the functional basis of transcription initiation. Nat Rev Mol Cell Biol 19:621–637. https://doi.org/10.1038/S41580-018-0028-8
Hansford S, Kaurah P, Li-Chang H, et al (2015) Hereditary Diffuse Gastric Cancer Syndrome: *CDH1* Mutations and Beyond. JAMA Oncol 1:23–32. https://doi.org/10.1001/jamaoncol.2014.168
Harris TJ, Tepass U (2010) Adherens junctions: from molecules to morphogenesis. Nat Rev Mol Cell Biol 11:502–514. https://doi.org/10.1038/nrm2927
Hatta W, Gotoda T, Koike T, Masamune A (2020) History and future perspectives in Japanese guidelines for endoscopic resection of early gastric cancer. Dig Endosc 32:180–190. https://doi.org/10.1111/DEN.13531
Hegde M, Ferber M, Mao R, et al (2014) ACMG technical standards and guidelines for genetic testing for inherited colorectal cancer (Lynch syndrome, familial adenomatous polyposis, and MYH-associated polyposis). Genet Med 16:101–116. https://doi.org/10.1038/GIM.2013.166
Hnisz D, Shrinivas K, Young RA, et al (2017) A Phase Separation Model for Transcriptional Control. Cell 169:13–23. https://doi.org/10.1016/J.CELL.2017.02.007
Hosoya K, Yamashita S, Ando T, et al (2009) Adenomatous polyposis coli 1A is likely to be methylated as a passenger in human gastric carcinogenesis. Cancer Lett 285:182–189. https://doi.org/10.1016/J.CANLET.2009.05.016
Hughes JR, Roberts N, Mcgowan S, et al (2014) Analysis of hundreds of cis-regulatory landscapes at high resolution in a single, high-throughput experiment. Nat Genet 46:205–212. https://doi.org/10.1038/NG.2871
Humar B, Blair V, Charlton A, et al (2009) E-cadherin deficiency initiates gastric signet-ring cell carcinoma in mice and man. Cancer Res 69:2050–2056. https://doi.org/10.1158/0008-5472.CAN-08-2457
Ichikawa H, Wakai T, Nagahashi M, et al (2018) Pathogenic germline BRCA1/2 mutations and familial predisposition to gastric cancer. JCO Precis Oncol 2:1–8. https://doi.org/10.1200/PO.18.00097
Ikenoue T, Arai M, Ishioka C, et al (2019) Importance of gastric cancer for the diagnosis and surveillance of Japanese Lynch syndrome patients. J Hum Genet 64:1187–1194. https://doi.org/10.1038/s10038-019-0674-5
Ilyas M, Tomlinson IP (1997) The interactions of APC, E-cadherin and beta-catenin in tumour development and progression. J Pathol 182:128–37. https://doi.org/10.1002/(SICI)1096-9896(199706)182:2<128::AID-PATH839>3.0.CO;2-Q

Ionov Y, Peinado MA, Malkhosyan S, et al (1993) Ubiquitous somatic mutations in simple repeated sequences reveal a new mechanism for colonic carcinogenesis. Nature 363:558–561. https://doi.org/10.1038/363558A0
Ishino Y, Shinagawa H, Makino K, et al (1987) Nucleotide sequence of the iap gene, responsible for alkaline phosphatase isozyme conversion in Escherichia coli, and identification of the gene product. J Bacteriol 169:5429–5433. https://doi.org/10.1128/JB.169.12.5429-5433.1987
Jinek M, Chylinski K, Fonfara I, et al (2012) A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity. Science 337:816–821. https://doi.org/10.1126/SCIENCE.1225829
Jinek M, East A, Cheng A, et al (2013) RNA-programmed genome editing in human cells. Elife 2:. https://doi.org/10.7554/ELIFE.00471
Kadiyska TK, Todorov TP, Bichev SN, et al (2014) APC promoter 1B deletion in familial polyposis--implications for mutation-negative families. Clin Genet 85:452–7. https://doi.org/10.1111/cge.12210
Kanemitsu K, Iwatsuki M, Yamashita K, et al (2021) Two Asian families with gastric adenocarcinoma and proximal polyposis of the stomach successfully treated via laparoscopic total gastrectomy. Clin J Gastroenterol 14:92–97. https://doi.org/10.1007/S12328-020-01290-6
Kang YK, Boku N, Satoh T, et al (2017) Nivolumab in patients with advanced gastric or gastro-oesophageal junction cancer refractory to, or intolerant of, at least two previous chemotherapy regimens (ONO-4538-12, ATTRACTION-2): a randomised, double-blind, placebo-controlled, phase 3 trial. Lancet 390:2461–2471. https://doi.org/10.1016/S0140-6736(17)31827-5
Karam R, Carvalho J, Bruno I, et al (2008) The NMD mRNA surveillance pathway downregulates aberrant E-cadherin transcripts in gastric cancer cells and in *CDH1* mutation carriers. Oncogene 27:4255–4260. https://doi.org/10.1038/ONC.2008.62
Kaurah P, MacMillan A, Boyd N, et al (2007) Founder and recurrent *CDH1* mutations in families with hereditary diffuse gastric cancer. JAMA 297:2360–2372. https://doi.org/10.1001/JAMA.297.21.2360
Kim H, Hwang Y, Sung H, et al (2018) Effectiveness of Gastric Cancer Screening on Gastric Cancer Incidence and Mortality in a Community-Based Prospective Cohort. Cancer Res Treat 50:582–589. https://doi.org/10.4143/CRT.2017.048
Kim S, Kim D, Cho SW, et al (2014) Highly efficient RNA-guided genome editing in human cells via delivery of purified Cas9 ribonucleoproteins. Genome Res 24:1012–1019. https://doi.org/10.1101/GR.171322.113
Kothary R, Clapoff S, Darling S, et al (1989) Inducible expression of an hsp68-lacZ hybrid gene in transgenic mice. Development 105:707–714. https://doi.org/10.1242/DEV.105.4.707
Kunovsky L, Kala Z, Potrusil M, et al (2019) A Central European family with gastric adenocarcinoma and proximal polyposis of the stomach. Gastrointest Endosc 90:523–525. https://doi.org/10.1016/j.gie.2019.05.001
Kurth I, Klopocki E, Stricker S, et al (2009) Duplications of noncoding elements 5' of SOX9 are associated with brachydactyly-anonychia. Nat Genet 41:862–863. https://doi.org/10.1038/NG0809-862
Lagarde A, Rouleau E, Ferrari A, et al (2010) Germline APC mutation spectrum derived from 863 genomic variations identified through a 15-year medical genetics service to French patients with FAP. J Med Genet 47:721–2. https://doi.org/10.1136/jmg.2010.078964
Landrum MJ, Lee JM, Riley GR, et al (2014) ClinVar: public archive of relationships among sequence variation and human phenotype. Nucleic Acids Res 42:. https://doi.org/10.1093/NAR/GKT1113
LAUREN P (1965) THE TWO HISTOLOGICAL MAIN TYPES OF GASTRIC CARCINOMA: DIFFUSE AND SO-CALLED INTESTINAL-TYPE CARCINOMA. AN ATTEMPT AT A HISTO-CLINICAL CLASSIFICATION. Acta Pathol Microbiol Scand 64:31–49. https://doi.org/10.1111/APM.1965.64.1.31
Lee CYC, Olivier A, Honing J, et al (2023) Endoscopic surveillance with systematic random biopsy for the early diagnosis of hereditary diffuse gastric cancer: a prospective 16-year longitudinal cohort study. Lancet Oncol 24:107–116. https://doi.org/10.1016/S1470-2045(22)00700-8
Lee JH, Han SU, Cho H, et al (2000) A novel germ line juxtamembrane Met mutation in human gastric cancer. Oncogene 19:4947–4953. https://doi.org/10.1038/SJ.ONC.1203874
Lee K, Krempely K, Roberts ME, et al (2018) Specifications of the ACMG/AMP variant curation guidelines for the analysis of germline *CDH1* sequence variants. Hum Mutat 39:1553–1568. https://doi.org/10.1002/humu.23650
Lettice LA, Heaney SJH, Purdie LA, et al (2003) A long-range Shh enhancer regulates expression in the developing limb and fin and is associated with preaxial polydactyly. Hum Mol Genet 12:1725–1735. https://doi.org/10.1093/HMG/DDG180
Li J, Woods SL, Healey S, et al (2016) Point Mutations in Exon 1B of APC Reveal Gastric Adenocarcinoma and Proximal Polyposis of the Stomach as a Familial Adenomatous Polyposis Variant. Am J Hum Genet 98:830–842. https://doi.org/10.1016/j.ajhg.2016.03.001

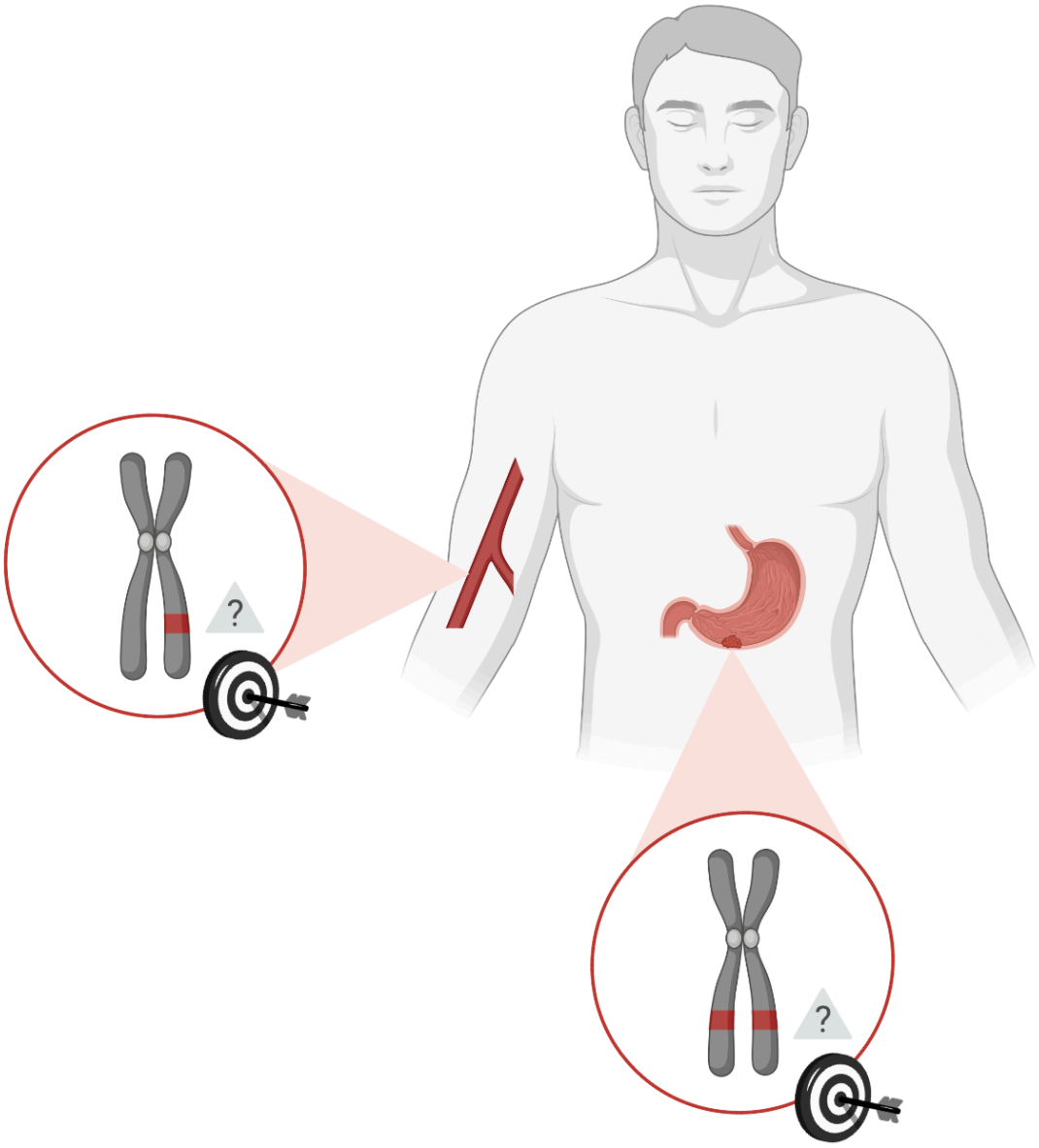
Lieberman-Aiden E, Van Berkum NL, Williams L, et al (2009) Comprehensive mapping of long-range interactions reveals folding principles of the human genome. Science 326:289–293. https://doi.org/10.1126/SCIENCE.1181369
Lin S, Staahl BT, Alla RK, Doudna JA (2014) Enhanced homology-directed human genome engineering by controlled timing of CRISPR/Cas9 delivery. Elife 3:e04766. https://doi.org/10.7554/ELIFE.04766
Lo W, Zhu B, Sabesan A, et al (2019) Associations of *CDH1* germline variant location and cancer phenotype in families with hereditary diffuse gastric cancer (HDGC). J Med Genet 56:. https://doi.org/10.1136/JMEDGENET-2018-105361
Lobo S, Benusiglio PR, Coulet F, et al (2021) Cancer predisposition and germline *CTNNA1* variants. Eur J Med Genet 64:104316. https://doi.org/10.1016/j.ejmg.2021.104316
Long HK, Osterwalder M, Welsh IC, et al (2020) Loss of Extreme Long-Range Enhancers in Human Neural Crest Drives a Craniofacial Disorder. Cell Stem Cell 27:765-783.e14. https://doi.org/10.1016/J.STEM.2020.09.001
Lupiáñez DG, Kraft K, Heinrich V, et al (2015) Disruptions of topological chromatin domains cause pathogenic rewiring of gene-enhancer interactions. Cell 161:1012–1025. https://doi.org/10.1016/J.CELL.2015.04.004
Ma D, Zhang Y, Shao X, et al (2022) PET/CT for Predicting Occult Lymph Node Metastasis in Gastric Cancer. Curr Oncol 29:6523–6539. https://doi.org/10.3390/CURRONCOL29090513
Majewski IJ, Kluijt I, Cats A, et al (2013) An α-E-catenin (*CTNNA1*) mutation in hereditary diffuse gastric cancer. J Pathol 229:621–629. https://doi.org/10.1002/path.4152
Mali P, Yang L, Esvelt KM, et al (2013) RNA-guided human genome engineering via Cas9. Science 339:823–826. https://doi.org/10.1126/SCIENCE.1232033
Melo S, Figueiredo J, Fernandes MS, et al (2017) Predicting the Functional Impact of *CDH1* Missense Mutations in Hereditary Diffuse Gastric Cancer. Int J Mol Sci 18:. https://doi.org/10.3390/IJMS18122687
Mi EZ, Mi EZ, di Pietro M, et al (2018) Comparative study of endoscopic surveillance in hereditary diffuse gastric cancer according to *CDH1* mutation status. Gastrointest Endosc 87:408–418. https://doi.org/10.1016/J.GIE.2017.06.028
Mitsui Y, Yokoyama R, Fujimoto S, et al (2018) First report of an Asian family with gastric adenocarcinoma and proximal polyposis of the stomach (GAPPS) revealed with the germline mutation of the APC exon 1B promoter region. Gastric Cancer 21:1058–1063. https://doi.org/10.1007/S10120-018-0855-5
Moayyedi P, Axon ATR, Feltbower R, et al (2002) Relation of adult lifestyle and socioeconomic factors to the prevalence of Helicobacter pylori infection. Int J Epidemiol 31:624–631. https://doi.org/10.1093/IJE/31.3.624
Mocellin S, Pasquali S (2015) Diagnostic accuracy of endoscopic ultrasonography (EUS) for the preoperative locoregional staging of primary gastric cancer. Cochrane Database Syst Rev 2015:. https://doi.org/10.1002/14651858.CD009944.PUB2
Mojica FJM, Díez-Villaseñor C, García-Martínez J, Soria E (2005) Intervening sequences of regularly spaced prokaryotic repeats derive from foreign genetic elements. J Mol Evol 60:174–182. https://doi.org/10.1007/S00239-004-0046-3
Morgan E, Arnold M, Camargo MC, et al (2022) The current and future incidence and mortality of gastric cancer in 185 countries, 2020-40: A population-based modelling study. EClinicalMedicine 47:. https://doi.org/10.1016/J.ECLINM.2022.101404
Muerdter F, Boryn ŁM, Woodfin AR, et al (2018) Resolving systematic errors in widely used enhancer activity assays in human cells. Nat Methods 15:141–149. https://doi.org/10.1038/NMETH.4534
Nagtegaal ID, Odze RD, Klimstra D, et al (2020) The 2019 WHO classification of tumours of the digestive system. Histopathology 76:182–188. https://doi.org/10.1111/HIS.13975
Northcott PA, Lee C, Zichner T, et al (2014) Enhancer hijacking activates GFI1 family oncogenes in medulloblastoma. Nature 511:428–434. https://doi.org/10.1038/NATURE13379

Okamoto R, Irie K, Yamada A, et al (2005) Recruitment of E-cadherin associated with alpha- and beta-catenins and p120ctn to the nectin-based cell-cell adhesion sites by the action of 12-O-tetradecanoylphorbol-13-acetate in MDCK cells. Genes Cells 10:435–445. https://doi.org/10.1111/J.1365-2443.2005.00846.X
Oliveira C, Bordin MC, Grehan N, et al (2002) Screening E-cadherin in gastric cancer families reveals germline mutations only in hereditary diffuse gastric cancer kindred. Hum Mutat 19:510–517. https://doi.org/10.1002/HUMU.10068
Oliveira C, Pinheiro H, Figueiredo J, et al (2015) Familial gastric cancer: genetic susceptibility, pathology, and implications for management. Lancet Oncol 16:e60-70. https://doi.org/10.1016/s1470-2045(14)71016-2
Oliveira C, Senz J, Kaurah P, et al (2009a) Germline *CDH1* deletions in hereditary diffuse gastric cancer families. Hum Mol Genet 18:1545–1555. https://doi.org/10.1093/hmg/ddp046
Oliveira C, Sousa S, Pinheiro H, et al (2009b) Quantification of epigenetic and genetic 2nd hits in *CDH1* during hereditary diffuse gastric cancer syndrome progression. Gastroenterology 136:2137–2148. https://doi.org/10.1053/j.gastro.2009.02.065
Pattanayak V, Lin S, Guilinger JP, et al (2013) High-throughput profiling of off-target DNA cleavage reveals RNA-programmed Cas9 nuclease specificity. Nat Biotechnol 31:839–843. https://doi.org/10.1038/NBT.2673
Pavicic W, Nieminen TT, Gylling A, et al (2014) Promoter-specific alterations of APC are a rare cause for mutation-negative familial adenomatous polyposis. Genes Chromosomes Cancer 53:857–64. https://doi.org/10.1002/gcc.22197
Pharoah PD, Guilford P, Caldas C (2001) Incidence of gastric cancer and breast cancer in *CDH1* (E-cadherin) mutation carriers from hereditary diffuse gastric cancer families. Gastroenterology 121:1348–1353. https://doi.org/10.1053/gast.2001.29611
Pocurull A, Herrera-Pariente C, Carballal S, et al (2021) Clinical, Molecular and Genetic Characteristics of Early Onset Gastric Cancer: Analysis of a Large Multicenter Study. Cancers (Basel) 13:. https://doi.org/10.3390/CANCERS13133132
Pradeepa MM, Grimes GR, Kumar Y, et al (2016) Histone H3 globular domain acetylation identifies a new class of enhancers. Nat Genet 48:681–686. https://doi.org/10.1038/NG.3550
Provenzale D, Gupta S, Ahnen DJ, et al (2016) Genetic/Familial High-Risk Assessment: Colorectal Version 1.2016, NCCN Clinical Practice Guidelines in Oncology. J Natl Compr Canc Netw 14:1010–1030. https://doi.org/10.6004/JNCCN.2016.0108
Rao SSP, Huntley MH, Durand NC, et al (2014) A 3D map of the human genome at kilobase resolution reveals principles of chromatin looping. Cell 159:1665–1680. https://doi.org/10.1016/J.CELL.2014.11.021
Repak R, Kohoutova D, Podhola M, et al (2016) The first European family with gastric adenocarcinoma and proximal polyposis of the stomach: case report and review of the literature. Gastrointest Endosc 84:718–725. https://doi.org/10.1016/J.GIE.2016.06.023
Richards FM, McKee SA, Rajpar MH, et al (1999) Germline E-cadherin gene (*CDH1*) mutations predispose to familial gastric cancer and colorectal cancer. Hum Mol Genet 8:607–610. https://doi.org/10.1093/HMG/8.4.607
Richards S, Aziz N, Bale S, et al (2015) Standards and guidelines for the interpretation of sequence variants: a joint consensus recommendation of the American College of Medical Genetics and Genomics and the Association for Molecular Pathology. Genet Med 17:405–424. https://doi.org/10.1038/GIM.2015.30
Roberts ME, Ranola JMO, Marshall ML, et al (2019) Comparison of *CDH1* Penetrance Estimates in Clinically Ascertained Families vs Families Ascertained for Multiple Gastric Cancers. JAMA Oncol 5:1325–1331. https://doi.org/10.1001/jamaoncol.2019.1208
Rohlin A, Engwall Y, Fritzell K, et al (2011) Inactivation of promoter 1B of APC causes partial gene silencing: evidence for a significant role of the promoter in regulation and causative of familial adenomatous polyposis. Oncogene 30:4977–4989. https://doi.org/10.1038/ONC.2011.201
Rosania R, Chiapponi C, Malfertheiner P, Venerito M (2016) Nutrition in Patients with Gastric Cancer: An Update. Gastrointest Tumors 2:178–187. https://doi.org/10.1159/000445188
Rudloff U (2018) Gastric adenocarcinoma and proximal polyposis of the stomach: diagnosis and clinical perspectives. Clin Exp Gastroenterol 11:447–459. https://doi.org/10.2147/CEG.S163227
Ruf S, Symmons O, Uslu VV, et al (2011) Large-scale analysis of the regulatory architecture of the mouse genome with a transposon-associated sensor. Nat Genet 43:379–387. https://doi.org/10.1038/NG.790
Sahasrabudhe R, Lott P, Bohorquez M, et al (2017) Germline Mutations in PALB2, BRCA1, and RAD51C, Which Regulate DNA Recombination Repair, in Patients With Gastric Cancer. Gastroenterology 152:983-986.e6. https://doi.org/10.1053/j.gastro.2016.12.010
Selvanathan A, Nixon CY, Zhu Y, et al (2020) *CDH1* Mutation Distribution and Type Suggests Genetic Differences between the Etiology of Orofacial Clefting and Gastric Cancer. Genes (Basel) 11:. https://doi.org/10.3390/GENES11040391
Sereno M, Aguayo C, Guillen Ponce C, et al (2011) Gastric tumours in hereditary cancer syndromes: clinical features, molecular biology and strategies for prevention. Clin Transl Oncol 13:599–610
Setia N, Clark JW, Duda DG, et al (2015) Familial gastric cancers. Oncologist 20:1365–1377. https://doi.org/10.1634/theoncologist.2015-0205
Shah SA, Erdmann S, Mojica FJM, Garrett RA (2013) Protospacer recognition motifs: mixed identities and functional diversity. RNA Biol 10:891–899. https://doi.org/10.4161/RNA.23764
Shapiro L, Fannon AM, Kwong PD, et al (1995) Structural basis of cell-cell adhesion by cadherins. Nature 374:327–337. https://doi.org/10.1038/374327A0
Simonis M, Klous P, Splinter E, et al (2006) Nuclear organization of active and inactive chromatin domains uncovered by chromosome conformation capture-on-chip (4C). Nat Genet 38:1348–1354. https://doi.org/10.1038/NG1896
Slavin T, Neuhausen SL, Rybak C, et al (2017) Genetic Gastric Cancer Susceptibility in the International Clinical Cancer Genomics Community Research Network. Cancer Genet 216–217:111–119. https://doi.org/10.1016/J.CANCERGEN.2017.08.001
Smyth EC, Nilsson M, Grabsch HI, et al (2020) Gastric cancer. Lancet 396:635–648. https://doi.org/10.1016/S0140-6736(20)31288-5
Smyth EC, Verheij M, Allum W, et al (2016) Gastric cancer: ESMO Clinical Practice Guidelines for diagnosis, treatment and follow-up. Ann Oncol 27:v38–v49. https://doi.org/10.1093/ANNONC/MDW350
Snow AK, Tuohy TMF, Sargent NR, et al (2015) APC promoter 1B deletion in seven American families with familial adenomatous polyposis. Clin Genet 88:360–365. https://doi.org/10.1111/CGE.12503
Sokoloff B (1938) Predisposition to cancer in the Bonaparte family. The American Journal of Surgery 40:673–678. https://doi.org/10.1016/S0002-9610(38)90653-1
Sonnenberg A (2011) Time trends of mortality from gastric cancer in Europe. Dig Dis Sci 56:1112–1118. https://doi.org/10.1007/S10620-010-1553-2
Spitz F, Furlong EEM (2012) Transcription factors: from enhancer binding to developmental control. Nat Rev Genet 13:613–626. https://doi.org/10.1038/NRG3207
Stuebs F, Heidemann S, Caliebe A, et al (2018) *CDH1* mutation screen in a BRCA1/2-negative familial breast-/ovarian cancer cohort. Arch Gynecol Obstet 297:147–152. https://doi.org/10.1007/S00404-017-4551-1
Sung H, Ferlay J, Siegel RL, et al (2021) Global Cancer Statistics 2020: GLOBOCAN Estimates of Incidence and Mortality Worldwide for 36 Cancers in 185 Countries. CA Cancer J Clin 71:209–249. https://doi.org/10.3322/CAAC.21660
Tacheci I, Repak R, Podhola M, et al (2021) Gastric adenocarcinoma and proximal polyposis of the stomach (GAPPS) - A Helicobacter-opposite point. Best Pract Res Clin Gastroenterol 50–51:. https://doi.org/10.1016/J.BPG.2021.101728
Taeubner J, Wieczorek D, Yasin L, et al (2018) Penetrance and Expressivity in Inherited Cancer Predisposing Syndromes. Trends Cancer 4:718–728. https://doi.org/10.1016/J.TRECAN.2018.09.002
Takeichi M (2018) Multiple functions of α-catenin beyond cell adhesion regulation. Curr Opin Cell Biol 54:24–29. https://doi.org/10.1016/j.ceb.2018.02.014
Tedaldi G, Pirini F, Tebaldi M, et al (2019) Multigene Panel Testing Increases the Number of Loci Associated with Gastric Cancer Predisposition. Cancers (Basel) 11:. https://doi.org/10.3390/CANCERS11091340
Thibodeau SN, Bren G, Schaid D (1993) Microsatellite instability in cancer of the proximal colon. Science 260:816–819. https://doi.org/10.1126/SCIENCE.8484122
Tsugane S, Sasazuki S (2007) Diet and the risk of gastric cancer: review of epidemiological evidence. Gastric Cancer 10:75–83. https://doi.org/10.1007/S10120-007-0420-0

Tsugeno Y, Nakano K, Nakajima T, et al (2020) Histopathologic Analysis of Signet-ring Cell Carcinoma In Situ in Patients With Hereditary Diffuse Gastric Cancer. Am J Surg Pathol 44:1204–1212. https://doi.org/10.1097/PAS.0000000000001511
van der Post RS, van Dieren J, Grelack A, et al (2018) Outcomes of screening gastroscopy in first-degree relatives of patients fulfilling hereditary diffuse gastric cancer criteria. Gastrointest Endosc 87:397-404.e2. https://doi.org/10.1016/J.GIE.2017.04.016
van der Post RS, Vogelaar IP, Carneiro F, et al (2015) Hereditary diffuse gastric cancer: Updated clinical guidelines with an emphasis on germline *CDH1* mutation carriers. J Med Genet 52:361–374. https://doi.org/10.1136/jmedgenet-2015-103094
Van Der Post RS, Vogelaar IP, Manders P, et al (2015) Accuracy of Hereditary Diffuse Gastric Cancer Testing Criteria and Outcomes in Patients With a Germline Mutation in *CDH1*. Gastroenterology 149:897–906. https://doi.org/10.1053/j.gastro.2015.06.003
van Roy F (2014) Beyond E-cadherin: roles of other cadherin superfamily members in cancer. Nat Rev Cancer 14:121–134. https://doi.org/10.1038/nrc3647
Vanbutsele G, Van Belle S, Surmont V, et al (2020) The effect of early and systematic integration of palliative care in oncology on quality of life and health care use near the end of life: A randomised controlled trial. Eur J Cancer 124:186–193. https://doi.org/10.1016/J.EJCA.2019.11.009
Vasen HFA, Blanco I, Aktan-Collan K, et al (2013) Revised guidelines for the clinical management of Lynch syndrome (HNPCC): recommendations by a group of European experts. Gut 62:812–23. https://doi.org/10.1136/gutjnl-2012-304356
Visel A, Blow MJ, Li Z, et al (2009) ChIP-seq accurately predicts tissue-specific activity of enhancers. Nature 457:854–858. https://doi.org/10.1038/NATURE07730
Vos JR, Giepmans L, Röhl C, et al (2019) Boosting care and knowledge about hereditary cancer: European Reference Network on Genetic Tumour Risk Syndromes. Fam Cancer 18:281–284. https://doi.org/10.1007/S10689-018-0110-6
Weren RDA, Van Der Post RS, Vogelaar IP, et al (2018) Role of germline aberrations affecting *CTNNA1*, MAP3K6 and MYD88 in gastric cancer susceptibility. J Med Genet 55:669–674. https://doi.org/10.1136/JMEDGENET-2017-104962
Will AJ, Cova G, Osterwalder M, et al (2017) Composition and dosage of a multipartite enhancer cluster control developmental expression of Ihh (Indian hedgehog). Nat Genet 49:1539–1545. https://doi.org/10.1038/NG.3939
Worthley DL, Phillips KD, Wayte N, et al (2012) Gastric adenocarcinoma and proximal polyposis of the stomach (GAPPS): a new autosomal dominant syndrome. Gut 61:774–779. https://doi.org/10.1136/gutjnl-2011-300348
Xicola RM, Li S, Rodriguez N, et al (2019) Clinical features and cancer risk in families with pathogenic *CDH1* variants irrespective of clinical criteria. J Med Genet 56:838–843. https://doi.org/10.1136/jmedgenet-2019-105991
Yanaru-Fujisawa R, Nakamura S, Moriyama T, et al (2012) Familial fundic gland polyposis with gastric cancer. Gut 61:1103–4. https://doi.org/10.1136/gutjnl-2011-301384
Yen T, Stanich PP, Axell L, Patel SG (1993) APC-Associated Polyposis Conditions
Zabidi MA, Stark A (2016) Regulatory Enhancer-Core-Promoter Communication via Transcription Factors and Cofactors. Trends Genet 32:801–814. https://doi.org/10.1016/J.TIG.2016.10.003
Zhao Z, Tavoosidana G, Sjölinder M, et al (2006) Circular chromosome conformation capture (4C) uncovers extensive networks of epigenetically regulated intra- and interchromosomal interactions. Nat Genet 38:1341–1347. https://doi.org/10.1038/NG1891
(2017) Japanese gastric cancer treatment guidelines 2014 (ver. 4). Gastric Cancer 20:1–19. https://doi.org/10.1007/S10120-016-0622-4

Chapter 2

Rational and aims



Gene expression in time and space is tightly regulated through physical enhancer-promoter communication within the nucleus. The 3D chromatin folding enables physical contacts between promoters and distal enhancers, across large genomic distances, being essential for cell-type-specific gene expression. When disrupted, regulatory landscapes may lead to gene misexpression and, consequently, disease.

This thesis was focused on the most frequent gastric cancer syndrome, hereditary diffuse gastric cancer (HDGC).

Single nucleotide variants (SNVs) in genes encoding epithelial adhesion molecules E-cadherin (*CDH1*) and α-catenin (*CTNNA1*) have been described as the main germline genetic defects triggering HDGC, in 10-40% (*CDH1* SNVs and CNVs) and 2% (*CTNNA1* SNVs) of the cases. Therefore, more than half of HDGC families remain genetically unexplained. Interestingly, genetically unsolved patients display *CDH1* mRNA monoallelic downregulation in the germline, a common feature observed in families that bear causing SNVs and CNVs. These results pinpoint *CDH1* as the major gene triggering HDGC, and we hypothesise that unsolved HDGC cases are caused by defects in regulatory elements controlling *CDH1* expression.

Thus, this thesis is dedicated at exploring the *CDH1* regulatory network and its impact in *CDH1*/E-cadherin expression, using HDGC and sporadic gastric cancer as disease models.

The search for genetic mechanisms causing hereditary cancer syndromes reached a bottleneck after the recent efforts using gene-panels sequencing, whole exome sequencing and whole genome sequencing, as only a small fraction of cases have been solved. This led us to hypothesise that, the missing heritability might be caused by disruption of regulatory elements controlling expression of syndromes-associated genes, for which HDGC is no exception.

From the clinical standpoint this is relevant to address, as HDGC *CDH1* carriers are predisposed to diffuse gastric cancer and lobular breast cancer development in early adulthood. Current management criteria recommend prophylactic gastrectomy and mastectomy of at-risk individuals to prevent disease onset. In genetically unsolved families, prophylactic surgical measures are not recommended, and clinical management is based on family history alone, leading to cancer development or surgery in individuals, that may not be at risk.

Our major expectation is to identify novel HDGC germline genetic causes to improve diagnosis and clinical management in carriers. From the fundamental research standpoint, we expect to disclose novel tissue-specific *CDH1*-associated regulatory mechanisms.

During this thesis, we addressed the following specific aims:

A- Disclose the missing heritability in hereditary gastric cancer

Review paper 1, chapter 3

In this paper we search the literature for gastric cancer predisposing families to disclose other genes beyond *CDH1* that may predispose to diffuse or intestinal gastric cancer (DGC/IGC) development

B- Unveil the germline landscape of HDGC patients, pinpointing *CDH1* potential regulatory elements

Original paper 1, chapter 3

In this paper we sequenced a cohort composed of 21 HDGC-like families and explored the *CDH1* regulatory network to unravel whether *CDH1* regulatory elements may predispose to HDGC or create a genetic background favourable to DGC development

C- Disclose the missing heritability in HDGC patients, beyond *CDH1* coding variants

Original paper 2, chapter 3

In this paper we describe a HDGC family with a large *CDH1* deletion, encompassing the neighbouring gene. Herein we mimicked the CNV to test if *CDH1* regulatory elements beyond the *CDH1* coding sequence may aggravate the HDGC phenotype

D- Investigate the role of genetic variants and DNA methylation in *CDH1* promoter and enhancers for E-cadherin expression

Original paper 3, chapter 4

Herein we tested the methylation status with the *CDH1 locus* enhancers of cell lines and gastric specimens for *CDH1* loss-associated enhancers identification

E- Explore the impact of exonic and hypomorphic *CDH1* intronic enhancers in E-cadherin expression

Original paper 4, chapter 4

In this paper we CRISPR-Cas9 edited exonic and intronic regions to unveil the impact of non-coding rather than coding sequences in *CDH1* expression and at the genome-wide level in metastasis formation

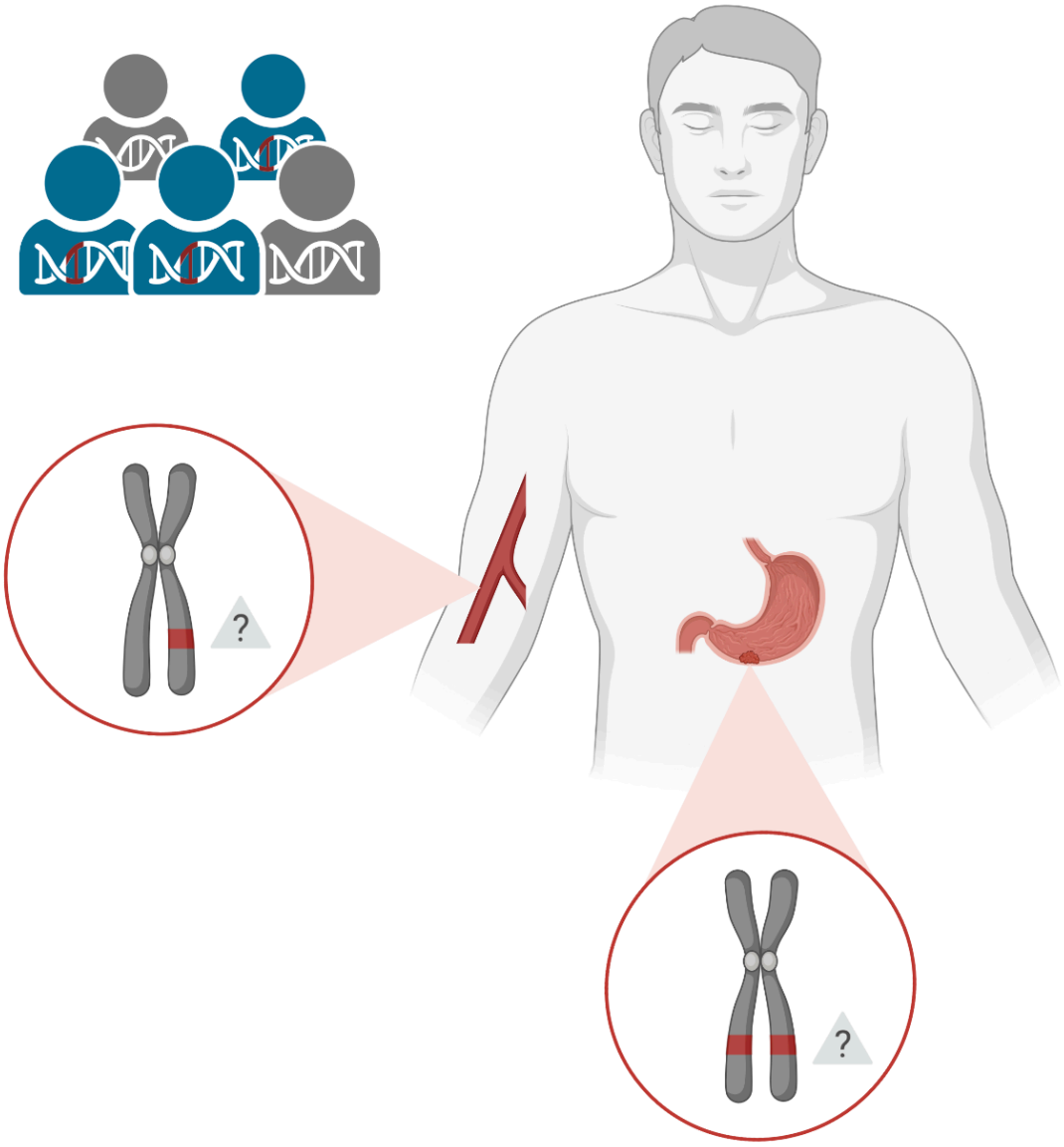
F- Unveil the molecular mechanisms behind E- to P-cadherin switch in gastric cancer

Original paper 5, chapter 4

Herein, we developed a E- to P-cadherin switch model in a gastric cancer cell line to dissect the regulatory elements behind *CDH1/CDH3 locus* expression

Chapter 3

Germline regulatory elements



Review paper 1

Gastric cancer genetic predisposition and clinical presentations: established heritable causes and potential candidate genes

José Garcia-Pelaez[1,2,3,8]†, Rita Barbosa-Matos[1,2,4,8]†, **Celina São José**[1,2,3,8]†, Sónia Sousa[1,2,8], Irene Gullo[1,2,5,6,8], Nicoline Hoogerbrugge[7,8], Fátima Carneiro[1,2,5,6,8] and Carla Oliveira[1,2,5,8*]

† These authors contributed equally to this work

[1] Instituto de Investigação e Inovação em Saúde, Porto, Portugal

[2] Instituto de Patologia e Imunologia Molecular da Universidade do Porto, Porto, Portugal

[3] Doctoral Programme in Biomedicine, Faculty of medicine, University of Porto, Porto, Portugal

[4] International Doctoral Programme in Molecular and Cellular Biotechnology Applied to Health Sciences from Institute of Biomedical Sciences Abel Salazar, University of Porto, Porto, Portugal

[5] Faculty of Medicine of the University of Porto, Porto, Portugal

[6] Centro Hospitalar e Universitário S. João, Porto, Portugal

[7] Department of Human Genetics, Radboud university medical center, Radboud Institute for Molecular Life Sciences, Nijmegen, The Netherlands

[8] Full Member of the European Reference Network on Genetic Tumor Risk Syndromes (ERN GENTURIS) – Project ID No 739547

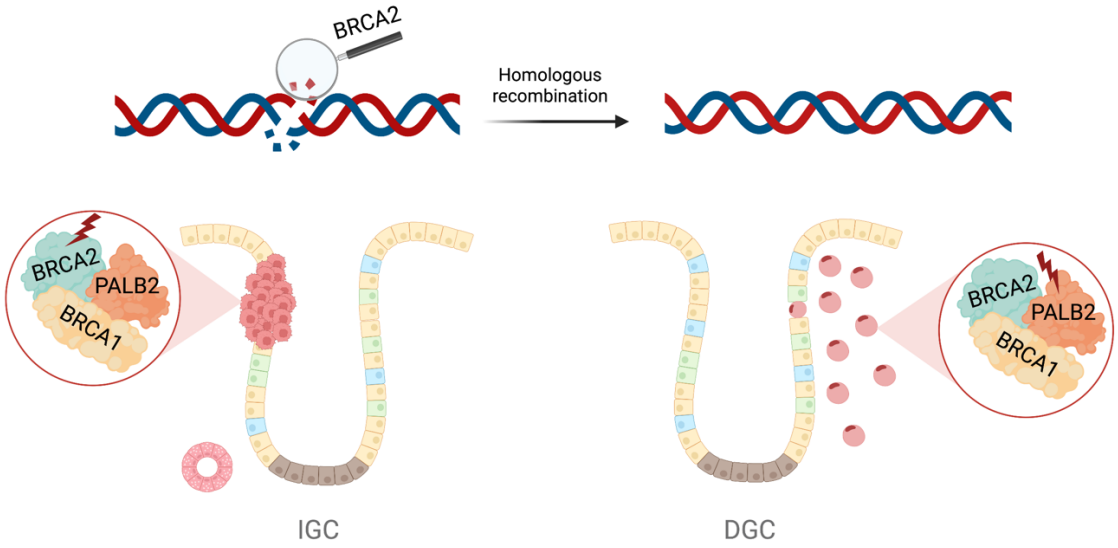
European Journal of Medical Genetics, 65(1), 104401 (appendix A).

Author statement

Celina São José: Investigation, methodology, conceptualization, formal analysis, writing – original draft, writing – review and editing.

Disclose the missing heritability in hereditary gastric cancer, beyond the *CDH1* gene

Herein, we search for individuals/families with clinical presentation and/or family history compatible with GC-associated syndromes and carrying pathogenic or likely pathogenic variants, independent of the genes. We found 27 families with family history compatible with HDGC or FIGC, harbouring 28 P/LP variants in 16 TRS-associated genes, mainly associated with DNA repair. *BRCA2* and *PALB2* were the most frequently mutated candidate genes in individuals with family history compatible with FIGC and HDGC, respectively. Consolidation of *PALB2* and *BRCA2* as FIGC- or HDGC-associated genes, respectively, holds promise and worth additional research.



Abstract

Tumour risk syndromes (TRS) are characterized by increased risk of early-onset cancers in a familial context. High cancer risk is mostly driven by loss-of-function variants in a single cancer-associated gene. Presently, predisposition to diffuse gastric cancer (DGC) is explained by *CDH1* and *CTNNA1* pathogenic and likely pathogenic variants (P/LP), causing Hereditary Diffuse Gastric Cancer (HDGC); while *APC* promoter 1B single nucleotide variants predispose to Gastric Adenocarcinoma and Proximal Polyposis of the Stomach (GAPPS). Familial Intestinal Gastric Cancer (FIGC), recognized as a GC-predisposing disease, remains understudied and genetically unsolved. GC can also occur in the spectrum of other TRS. Identification of heritable causes, allows defining diagnostic testing criteria, helps to clinically classify GC families into the appropriate TRS, and allows performing pre-symptomatic testing identifying at-risk individuals for downstream surveillance, risk reduction and/or treatment. However, most of HDGC, some GAPPS, and most FIGC patients/families remain unsolved, expecting a heritable factor to be discovered. The missing heritability in GC-associated tumour risk syndromes (GC-TRS) is likely explained not by a single major gene, but by a diversity of genes, some, predisposing to other TRS. This would gain support, if GC-enriched small families or apparently isolated early-onset GC cases were hiding a family history compatible with another TRS. Herein, we revisited current knowledge on GC-TRS, and searched in the literature for individuals/families bearing P/LP variants predisposing for other TRS, but whose probands display a clinical presentation and/or family history also fitting a GC-TRS criteria. We found 27 families with family history compatible with HDGC or FIGC, harbouring 28 P/LP variants in 16 TRS-associated genes, mainly associated with DNA repair. *PALB2* or *BRCA2* were the most frequently mutated candidate genes in individuals with family history compatible with HDGC and FIGC, respectively. Consolidation of *PALB2* and *BRCA2* as HDGC- or FIGC-associated genes, respectively, holds promise and worth additional research. This analysis further highlighted the influence, that proband's choice and small or unreported family history have, for a correct TRS diagnosis, genetic screening, and disease management. In this review, we provide a rational for identification of particularly relevant candidate genes in GC-TRS.

Cancer predisposition in HDGC and FIGC: candidate genes

GC, genetically determined from birth, is a rare occurrence, and the well-established predisposing genes described above are still insufficient to explain disease in many of these rare patients and families. Indeed, 60-90% of HDGC, some GAPPS, and almost all FIGC patients/families remain unsolved, and expect their cancer heritability to be discovered. So, one of the most important questions in the field is how additional genes (candidate genes), with the potential to confer high predisposition to GC, could be found in molecularly unexplained families. One possibility is that the monogenic nature of disease in unsolved cases, is explained not by the major gene in that syndrome (e.g., *CDH1* in HDGC), but by an array of different predisposing genes, some of them known to predispose to other TRS (e.g., different MMR genes in LS). On the other hand, GC may occur as part of the cancer spectrum in other syndromes (as described in **Table 2**). Therefore, GC isolated cases, of either intestinal or diffuse histology, detected at early age with unknown or unreported family history, prior to genetic testing, may hide family histories compatible with classification into other TRS.

To address these possibilities, we screened the literature for reports on next generation sequencing (NGS) studies, of either gene-panels, exomes, or whole genomes which described potentially actionable germline variants in families that could fulfil criteria of GC-associated TRS. This analysis has two main limitations. One is related to FIGC: because no clinical criteria exist to select families at risk of FIGC, a few cases have been submitted for NGS studies. The other limitation is related with GAPPS: as most GAPPS families are reported to carry promoter 1B *APC* variants, and GAPPS is very rare, few families remain unsolved.

The search strategy herein used relied on very specific requirements. Selected cases needed to be considered genetically solved and present sufficient clinical information for allocation in at least one of the following criteria: (1) families/individuals fulfilling IGCLC-HDGC criteria 2020 (Blair et al. 2020a); (2) isolated cases of IGC <70 years old or families with >=2 GC cases one confirmed IGC, herein classified as FIGC, and; (3) families/individuals fulfilling GAPPS 2018 criteria. Selected studies are depicted in **Table 3** and **Supplementary Table S1.** Data is presented by affected gene and its association with the above listed criteria. Selection of publications was performed with PUBMED.

In PUBMED, we used as QUERY “Familial gastric cancer germline mutations” to search for manuscripts from 1993 to 2021. We retrieved 595 publications that were manually inspected to identify the above listed requirements. Fourteen publications were selected: 26182300 (Hansford et al. 2015), 28024868 (Sahasrabudhe et al. 2017), 25340522 (Gaston et al. 2014), 29706558 (Fewings et al. 2018), 28875981 (Vogelaar et al. 2017), 29330337 (Weren et al. 2018), 11042681 (Lee et al. 2000), 31608315 (Ichikawa et al. 2018), 29025585 (Slavin et al. 2017), 25576241 (Donner et al. 2015), 12920089 (Kim et al. 2003), 15173255 (Keller et al. 2004), 32066632 (Carvalho et al. 2021), 33525650 (Herrera-Pariente et al. 2021).

Additionally, two manuscripts were added manually, one was *in press* ((te Paske et al. 2021), PMID 34075207)) and the other was published but failed to be retrieved with the above query ((Tedaldi et al. 2019), PMID 31514334). The authors, due to co-authorship, knew both manuscripts.

We identified 71 distinct families, fulfilling criteria for either HDGC or FIGC, and bearing 66 variants in 34 different genes. After classifying all variants, using the ACMG guidelines (Richards et al. 2015), we ended up with 27 families carrying 28 different P/LP variants across 16 candidate genes (**Figure 3, Table 3**). Thirteen out of these 16 (81%) candidate genes are involved in DNA damage response pathways, such as homologous recombination repair and mismatch repair pathways. Additionally, several candidate genes were associated with other hereditary syndromes such as BC and Fanconi anaemia (**Figure 3**, **Table 4**). In case available information on variants was insufficient for their confirmation as P or LP, these were excluded from this final compilation. Finally, the frequency of P/LP variants in HDGC/FIGC families per each gene was calculated as presented below:

Frequency of P/ LP variants in HDGC families per gene (%):

$$\frac{Nº of HDGC families bearing\ P/LP\ variants}{Nº of families screened fulfilling HDGC criteria}$$

Frequency of P/ LP variants in FIGC families per gene (%):

$$\frac{Nº of FIGC families bearing\ P/LP\ variants}{Nº of families screened fulfilling FIGC criteria}$$

This allowed us to understand if a candidate gene was more associated with HDGC or FIGC, and to compare its mutation frequency with DGC and IGC in a sporadic context (**Table 4**).

From the 27 families identified, 20 fulfilled HDGC criteria and carried 21 P/LP variants in 14 different genes. Seven families fulfilled FIGC criteria and carried eight P/LP variants in six different genes (**Figure 3, Table 3**). *ATM*, *BRCA1*, *BRCA2* and *PALB2* variants were present in both HDGC and FIGC (**Figure 3, Table 3**). The same variant was never found in more than one family, except for the *PALB2* variant c.1240C>T, described in a HDGC and a FIGC family. In addition, two families carried more than one P/LP variant in two distinct genes, and eventually representing an oligogenic context (**Table 3**).

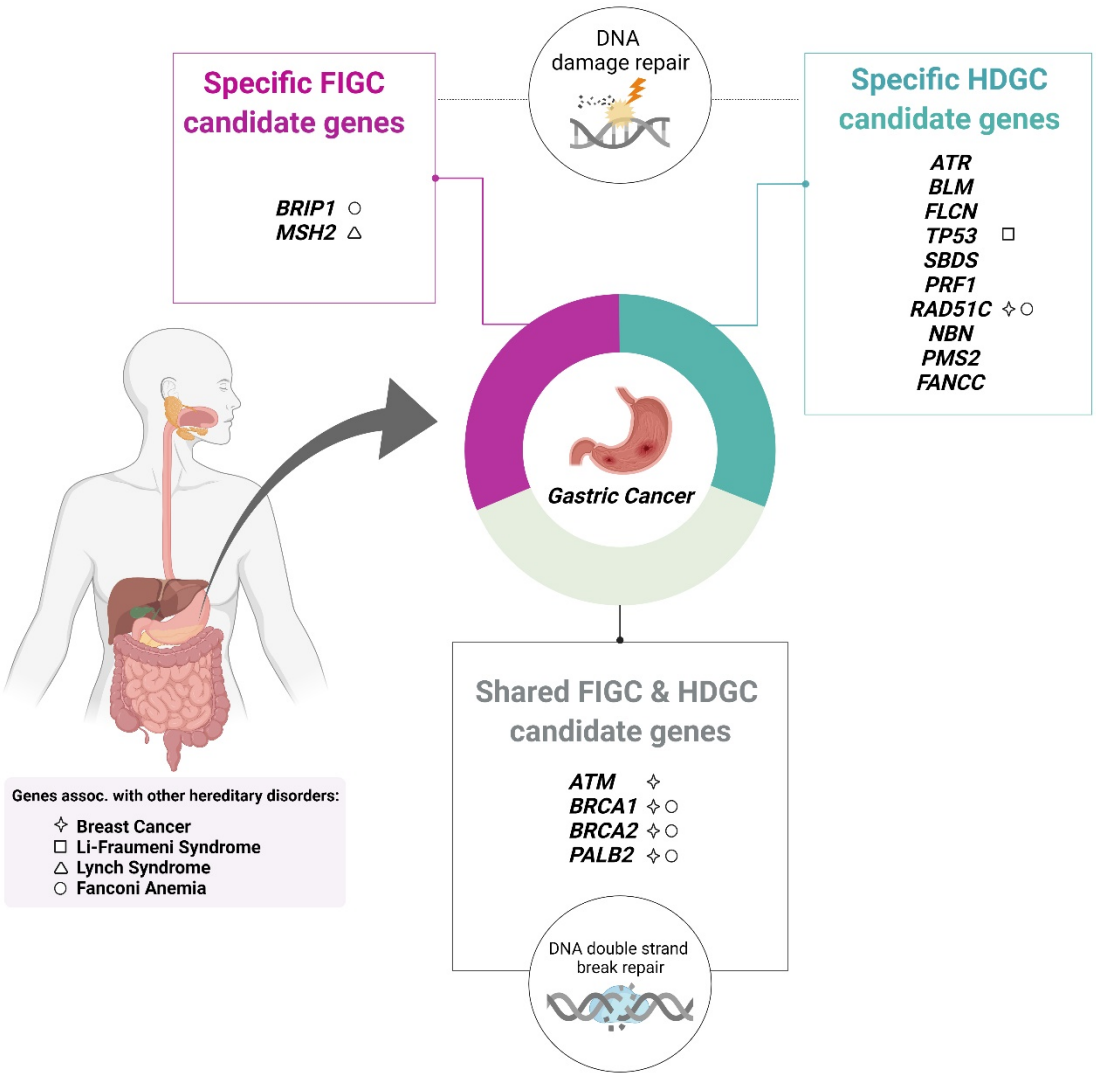


Figure 3. List of candidate genes retrieved from our literature review and its associated biological processes.

Table 3. Summary, description, and classification of P or PL variants found in individuals fulfilling HDGC criteria, and the FIGC criteria listed above. Phenotypes present in different individuals are separated by ';' and phenotypes present in the same individual are separated by ','. Probands are depicted in bold and '+' means a carrier of the variant not proband.

Gene	mRNA position HGVS	αα position HGVS	Molecular type	Primary syndrome	Alternative syndromes	Spectrum	Probands information	Relatives information	ACMG criteria	ACMG classification	References (PMID)
ATM	c.7705_7706del	p.Asp2569*	Frameshift	HDGC	NA	GC	**DGC 66y**	GC; GC; GC; GC; GC; GC; GC	PVS1, PM2, PS4	P	29025585
	c.1564_1565del	p.Glu522Ilefs*43	Frameshift	HDGC	FIGC	GC	IGC 54y	>=2 GC cases regardless of age, 1 confirmed DGC; >=2 GC cases ≤60 years in 1st-degree relatives; 1 IGC 54y	PVS1, PM2, PS4	P	31514334
	c.2192dup	p.Tyr731*	Nonsense	HDGC	NA	GC	**DGC 32y**	**No additional phenotypes reported**	PVS1, PM2	LP	31514334
ATR	c.6075T>A	Tyr2025*	Nonsense	HDGC	NA	GC/BC	**DGC 28y (#)**	SRCC 45y; DGC 60y*; SRCC; SRCC; CRC 75y; PRSC 60s; BC 40s	PVS1, PM2	LP	29706558
BLM	c.2395del	p.Cys799Valfs*16	Frameshift	HDGC	NA	GC/BC	**LBC 50y, DGC 54y**	>=2 GC cases regardless of age, 1 confirmed DGC	PM2, PVS1	LP	31514334
BRCA1	c.188T>A≈	p.Leu63*	Nonsense	FIGC	HBOC	GC/BC	**IGC 66y**	BC 37y; CXC 44y; LC; GC	PM2, PS4, PVS1; PP5	P	31608315
	c.406del	p.Arg136Aspfs*27	Frameshift	HDGC	NA	GC	**DGC 47y**	GC	PVS1, PM2. PP5	P	31514334
BRCA2	c.6037A>T	p.Lys2013*	Nonsense	FIGC	NA	GC	**IGC 60y**	GC	PM2, PS4, PVS1	P	31514334
	c.89dupA	p.Leu29fs	Frameshift	FIGC	NA	GC/BC	**IGC 73y**	GC 50y; GC 55y; BC 50y	PM2, PVS1	LP	31608315
	c.6922A>T≈	p.Lys2308*	Nonsense	FIGC	HBOC	GC/BC	**IGC 66y**	BC 37y; CXC 44y; LC; GC	PM2, PVS1	LP	31608315
	c.2808_2811del	p.Ala938Profs*21	Frameshift	FIGC	NA	GC	**IGC 58y, ESC 58y**	**No additional phenotypes reported**	PM2, PS4, PVS1, PP5	P	29025585
	c.3860_3863delATAA	p.Asn1287Ilefs*5	Frameshift	HDGC	NA	GC/BC	**DGC 64y**	GC 21y; BC 50y	PM2, PVS1, PP5	P	26182300
BRIP1	c.124_127TGTT/ c.128_131delTGTT	p.Leu43fs*11	Frameshift	FIGC	NA	GC	**IGC 19y**	**No additional phenotypes reported**	PM2, PVS1; PS4_mod	P	29025585
FANCC	c.1642C>T	p.Arg548*	Nonsense	HDGC	NA	GC	**DGC 50y (poorly differentiated)**	GC	PVS1, PS3, PM4, PM2	P	29025585
FLCN	c.59del	p.Phe20fs*35	Frameshift	HDGC	NA	GC	**DGC 17y**	**No additional phenotypes reported**	PVS1, PM2, PS4	P	29025585
MSH2	c.367-?_645+?del	NA	Frameshift	FIGC	NA	GC	**IGC 57y**	CRC	PM2, PVS1, PP5	P	31514334
NBN	c.1124+1G>C	NA	Splice-site	HDGC	NA	GC/BC	**DGC 28y (#)**	SRCC 45y; DGC 60y*; SRCC; SRCC; CRC 75y; PRSC 60s; BC 40s	PVS1, PM2	LP	29706558
PALB2	c.1240C>T	p.Arg414*	Nonsense	FIGC	NA	GC	**IGC 69y**	**No further information available**	PVS1, PM2, PS4	P	28024868
	c.1192delG	p.Val398fs	Frameshift	HDGC	HBOC	GC/BC	**DGC 45y**	GC 38y; BC 42y; BC 45y; CRC 59y; PRSC 75y	PVS1, PM2	LP	26182300
	c.1240C>T	p.Arg414*	Nonsense	HDGC	NA	Unk	**DGC 48y**	**No further information available**	PVS1, PM2, PS4	P	28024868
	c.3201+1G>T	NA	Splice-site	HDGC	LS	GC	**DGC 50y**	CRC 60y; GC 60y; GC 50y; GC; GC 50y; GC 52y; GC 57y; GC 58y; GC 58y; GC 50y; GC 52y; GC 50y	PVS1, PM2	LP	28024868
	c.755_756CT	p.Leu253fs	Frameshift	HDGC	NA	GC/BC	**DGC 55y**	DGC 53y; DGC 44y; BC; GC; LC; LRYC	PVS1; PM2, PS4	P	29706558
	c.535C>T	p.Gln179*	Nonsense	HLBC	NA	BC	**BLBC 62y, 66y**	**No additional phenotypes reported**	PVS1, PM2	LP	31514334
	c.2718G>A	p.Trp906*	Nonsense	HLBC	NA	BC	**LBC 62y**	Bilateral LBC or family history of >=2 LBC cases <50y	PVS1, PM2, PS4	P	31514334
PMS2	c.2182_2183delAC	p.Thr728Serfs*7	Frameshift	HDGC	NA	GC	**DGC 73y**	GC	PVS1, PM2	LP	31514334
PRF1	c.1122G>A	p.Trp374*	Nonsense	HDGC	NA	GC	**DGC 65y**	GC	PVS1, PM2, PS4	P	31514334
RAD51C	c.709C>T	p.Arg237*	Nonsense	HDGC	LS	GC	**DGC 73y, CRC**	GC 45y; GC 60y; CRC 38y	PM2, PS4, PVS1	P	28024868
SBDS	c.184A>T/ c.183_184delTAinsCT	p.Lys62*	Nonsense	HDGC	NA	GC	**DGC 40y**	**No additional phenotypes reported**	PVS1	LP	29025585
TP53	c.637C>T	p.Arg213*	Nonsense	HDGC	HBOC	GC/BC	**DGC 40y, BC 42y**	**No additional phenotypes reported**	PM2, PVS1, PS3_moderate, PS4	P	29025585

Footnote: (#) carrier of two LP variants in ATR and NBN. (≈) Same family carrying simultaneously a *BRCA1* and a *BRCA2* variant. NA- not available; HGVS - Human Genome Variation Society; ACMG - American College of Medical Genetics and Genomics; HDGC - Hereditary Diffuse Gastric Cancer; FIGC - Familial Intestinal Gastric Cancer; DGC - Diffuse Gastric Cancer; SRCC - Signet Ring Cell Carcinoma; GC - Gastric Cancer; BC - Breast Cancer; IGC - Intestinal Gastric Cancer; ESC - Esophageal Cancer; LBC - Lobular Breast Cancer; BLBC - Bilateral Lobular Breast Cancer; CRC - Colorectal Cancer; LS - Lynch Syndrome; SRCC - Signet Ring Cell Carcinoma; PRSC - Prostatic Cancer; CXC - Cervical Cancer; LC - Lung Cancer; LRYC - Laryngeal Cancer; P – pathogenic; LP – likely pathogenic; fs – frameshift.

Table 4. Summary of candidate genes identified in HDGC and FIGC families.

	Candidate Genes					HDGC				FIGC				Non-assertained		
Gene bearing germline variant	Main described function	Main associated syndrome (OMIM)	Total nº of families screened	GC-associated syndromes in carriers of P/LP variants per gene	Families screened per gene fulfilling HDGC clinical criteria	HDGC families bearing P/LP variants	Frequency of P/LP variants in HDGC families per gene (%)	Mutation frequency in sporadic DGC cases [PMID: 33724653] (%)	Families screened per gene fulfilling FIGC clinical criteria	FIGC families bearing P/LP variants	Frequency of P/LP variants FIGC families per gene (%)	Mutation frequency in sporadic IGC cases [PMID: 33724653] (%)	Nº of non-ascertain families	Non-ascertained families bearing P/LP variants	References (PMID)
PALB2	TS; Homologous recombination and repair of DNA, colocalized with *BRCA2*	Breast cancer, susceptibility to (AD; OMIM ID: 114480); Fanconi anemia, complementation group N (OMIM ID: 610832); Pancreatic cancer, susceptibility to, 3 (OMIM ID: 613348)	462	HDGC, HLBC, FIGC, HBOC, LS	275	6	2%	NA	11	1	9%	1% (1/75; single publication)	7	1	31514334, 26182300, 28024868, 29706558
FANCC	Delays the onset of apoptosis and promotes homologous recombination and repair of DNA	Fanconi anemia, complementation group C (AR; OMIM ID: 227645)	51	HDGC	16	1	6%	2.9% (2/72; single publication)	3	0	0%	NA	14	0	29025585
BRIP1	Double-strand break repair function	Breast cancer, early-onset, susceptibility to (AD; OMIM ID: 114480); Fanconi anemia, complementation group J (OMIM ID: 609054)	51	FIGC	16	0	0%	3% (1/34; single publication)	6	1	17%	NA	14	0	29025585
BRCA1	TS; Genome stability maintenance, specifically the homologous recombination and repair of DNA	Breast-ovarian cancer, familial, 1 (AD; OMIM ID: 604370); Fanconi anemia, complementation group S (AR; OMIM ID: 617883); Pancreatic cancer, susceptibility to, 4 (OMIM ID: 614320)	103	HDGC, FIGC, HBOC	78	1	1%	4% (4/93)	17	1	6%	8% (7/93)	7	0	31514334, 31608315
BRCA2	TS; Genome stability maintenance, specifically the homologous recombination and repair of DNA	Breast cancer, male, susceptibility to (AD; OMIM ID: 114480); Breast-ovarian cancer, familial, 2 (AD; OMIM ID: 612555); Fanconi anemia, complementation group D1 (AR; OMIM ID: 605724); Glioblastoma 3 (AR; OMIM ID: 613029); Medulloblastoma (AD; OMIM ID: 155255); Pancreatic cancer 2 (OMIM ID: 613347); Prostate cancer (AD; OMIM ID: 176807); Wilms tumor (OMIM ID:194070)	298	FIGC, HDGC, HBOC	238	1	0%	5% (15/298)	23	4	17%	9% (25/279)	21	0	31514334, 26182300, 31608315, 29025585
RAD51C	Homologous recombination and repair of DNA	Breast-ovarian cancer, familial, susceptibility to, 3 (OMIM ID: 613399); Fanconi anemia, complementation group O (AR; OMIM ID: 613390)	200	HDGC, LS	31	1	3%	NA	0	0	NA	NA	0	0	28024868
ATM	Cell cycle checkpoint kinase required for cell response to DNA damage and for genome stability.	Ataxia-telangiectasia (AR; OMIM ID: 208900); Breast cancer, susceptibility to (AD; OMIM ID: 114480)	147	HDGC, FIGC	94	3	3%	9% (24/268)	14	0	0%	8% (11/140)	21	0	31514334, 29025585
ATR	DNA damage response	Seckel syndrome 1 (AR; OMIM ID: 210600); Cutaneous telangiectasia and cancer syndrome, familial (AD; OMIM ID: 614564)	22	HDGC	22	1	5%	4% (4/106)	0	0	NA	8% (6/75; single publication)	0	0	29706558
MSH2	TS; Component of the post-replicative DNA mismatch repair system	Colorectal cancer, hereditary nonpolyposis, type 1 (AD; OMIM ID: 120435); Mismatch repair cancer syndrome 2 (AR; OMIM ID: 619096); Muir-Torre syndrome (AD; OMIM ID: 158320)	96	FIGC	78	0	0%	2% (2/106)	11	1	9%	6% (1/18; single publication)	7	0	31514334
PMS2	Component of the post-replicative DNA mismatch repair system	Mismatch repair cancer syndrome 4 (AR; OMIM ID: 619101); Colorectal cancer, hereditary nonpolyposis, type 4 (OMIM ID: 614337)	96	HDGC	78	1	1%	2% (2/122; single publication)	11	0	0%	3% (4/140)	7	0	31514334
SBDS	Essential role in ribosome biogenesis	Shwachman-Diamond syndrome (AR; OMIM ID:260400); Aplastic anemia, susceptibility to (OMIM ID: 609135)	51	HDGC	16	1	6%	3% (2/72; single publication)	3	0	0%	NA	14	0	29025585
BLM	Genome stability maintenance	Bloom syndrome (AR; OMIM ID: 210900)	96	HDGC	78	1	1%	4% (4/106)	11	0	0%	4% (4/93)	7	0	31514334
PRF1	Involved in cytolysis and in the regulation of the immune system.	Hemophagocytic lymphohistiocytosis, familial, 2 (AR; OMIM ID:603553); Aplastic anemia (OMIM ID: 609135); Lymphoma, non-Hodgkin (OMIM ID: 605027)	96	HDGC	78	1	1%	4% (4/106)	11	0	0%	6% (1/18; single publication)	7	0	31514334
FLCN	TS; Cellular response to amino acid availability, modulation of EGFR and mTOR signaling pathways and mTOR signaling pathways	Birt-Hogg-Dube syndrome (AD; OMIM ID: 135150); Pneumothorax, primary spontaneous (AD; OMIM ID: 173600)	51	HDGC	16	1	6%	1% (1/72; single publication)	3	0	0%	5% (5/93)	14	0	29025585
TP53	TS; Regulates cell cycle arrest, apoptosis, senescence, DNA repair, or changes in metabolism; controls the expression of target genes	Li-Fraumeni syndrome (AD; OMIM ID: 151623); Adrenocortical carcinoma, pediatric (AD; OMIM ID: 202300); Basal cell carcinoma 7 (AD; OMIM ID: 614740); Choroid plexus papilloma (AD; OMIM ID:260500); Colorectal cancer (AD; OMIM ID: 114500); Glioma susceptibility 1 (AD; OMIM ID: 137800); Bone marrow failure syndrome 5 (AD; OMIM ID: 618165)	51	HDGC, HBOC	16	1	6%	41% (167/407)	3	0	0%	57% (132/232)	14	0	29025585
TOTAL						20				8				1	

The gene with the highest number of variants classified as P/LP found in families fulfilling HDGC criteria was *PALB2* (16p12.2). *PALB2* gene encodes a protein with the same name that is part of the homologous recombination system, it has been extensively associated with predisposition to BC (OMIM Nº 114480; (Tischkowitz and Xia 2010; Antoniou et al. 2014), and considered as a HBOC-associated gene (Yang et al. 2016; Velázquez et al. 2019; Daly et al. 2021). *PALB2* protein is colocalized with *BRCA2*, which may explain its relationship with BC. *PALB2* was found mutated in six different families fulfilling HDGC criteria: two clearly HLBC (HDGC without known GC involvement); three clearly HDGC (without known LBC involvement); and one which fulfilled simultaneously HDGC and HBOC criteria (**Table 3**). As an independent entity, HLBC families may carry variants in genes exclusively predisposing to LBC, if a *CDH1* variant cannot not be found. *PALB2* may be one of such genes, but its clear association with the LBC phenotype has yet to be acknowledged.
One of the variants found in a HDGC case (DGC at 48 years), was also observed in an IGC patient at 69 years (c.1240C>T; **Table 3**), however, for both cases no information about cancer in relatives or family history was available. Overall, we found that 2% (6/275) of families displaying HDGC criteria carry a *PALB2* variant (**Table 4**, **Supplementary Table S2**). This frequency is higher than that found in the sporadic setting, where *PALB2* was found mutated in a single IGC case (1%; 1/75), at the somatic level (**Table 4**, **Supplementary Table S2**) (Garcia-Pelaez et al. 2021b). *PALB2* was also previously found infrequently mutated in the germline of patients with apparently sporadic GC from The Cancer Genome Atlas (TCGA) study (1.2%; 4/321), and in similar patients from another study (1.1%; 5/443) (Lu et al. 2015; Huang et al. 2018). However, the GC histology was not disclosed in these studies. Altogether, these data support *PALB2* germline P/LP variants as a potential cause of HDGC, although the mechanism remains unclear. Few families with IGC have been identified bearing *PALB2* germline variants, a phenotype divergent from HDGC. A deeper analysis of the *PALB2* LoF mode of action is needed to establish its formal role in this HDGC.

Another HDGC-candidate gene in our analysis is the *ATM* gene, located at chromosome 11q22.3, and considered the main trigger of ataxia-telangiectasia (OMIM Nº 208900) (Savitsky et al. 1995), associated with familial susceptibility to BC (OMIM Nº 114480) and early-onset GC, as a moderate-penetrant gene (Zhou et al. 2020; Setia et al. 2021). *ATM* encoded protein is a cell cycle checkpoint kinase that regulates DNA damage response through p53 and BRCA1 (Jin and Oh 2019; Zhou et al. 2020). Three germline *ATM* variants were identified and classified as P/LP (**Table 3**). From those, two variants were identified in families meeting HDGC clinical criteria, while one was found in a family that met both HDGC and FIGC selection criteria (**Table 3**). *ATM* therefore accounted for 3% (3/94) of HDGC families, and somatic variants were also found in sporadic DGC (9%; 24/268; **Table 4**, **Supplementary Table S2**)(Garcia-Pelaez et al. 2021b). The importance of *ATM* variants in GC has been a question in debate, since several cancer patients have been reported to carry both germline and somatic *ATM* variants, namely lymphomas, leukaemia, hepatocellular carcinoma, CRC,

GC, BC, among others (Navrkalova et al. 2013; Beà et al. 2013; Tung et al. 2015, 2016; Hansford et al. 2015; Choi et al. 2016; Pearlman et al. 2017; Lin et al. 2019). Although these data support *ATM* involvement in HDGC, *ATM* recognized pleiotropy hampers further conclusions.

The *PRF1* mapped in chromosome 10 is an additional candidate gene. It encodes Perforin 1, an important protein to resolve virus infections and to remove malignant cells from the body through a Ca^{2+}-dependent pore-forming mechanism (Law et al. 2010; Praper et al. 2011). *PRF1* variants are considered predisposing to hereditary cancer predisposition and lymphoid diseases (OMIM Nº 609135; 603553; 605027), although further studies are required to explore the role of *PRF1* in GC predisposition. Besides a P variant herein found in an HDGC family (**Table 3**), other GC cases were described in families harbouring *PRF1* germline variants (El Abed et al. 2011; Chaudhry et al. 2016). Recently, *PRF1* levels were found to be reduced in DGC patients and associated with a less functional $CD8^+$T cells population in DGC tumours (Li et al. 2020). The frequency of *PRF1* variants is similar between sporadic DGC (4%; 4/106) and IGC (6%; 1/18) (**Table 4, Supplementary Table S2**) (Garcia-Pelaez et al. 2021b)This single finding does not allow a conclusion on the involvement of *PRF1* in HDGC but collected evidence worth further research on this candidate gene.

We also found five P/LP variants in *BRCA2*, one occurring in a family fulfilling HDGC criteria and presenting BC involvement; and four fulfilling FIGC criteria from which two had BC involvement. Germline *BRCA2* (13q13.1) and *BRCA1* (17q21.31) alterations are the hallmark of autosomal dominant HBOC predisposition (**Table 2**; OMIM Nº 604370, 612555, 114480), and associated with autosomal recessive Fanconi Anaemia (OMIM Nº 617883, 605724). Further, both genes have been linked with increased risk of developing GC before 65 years of age (Johannsson et al. 1999; Jakubowska et al. 2002; Roy et al. 2011; Takaoka and Miki 2018). The fact that there are 4/23 (17%) FIGC families carrying *BRCA2* P/LP variants, supports *BRCA2* as a predisposing gene in FIGC (**Table 3**). Further supporting this relationship, is the fact that *BRCA2* somatic variants are found in 9% (25/279) of sporadic IGC and only in 5% (15/298) of sporadic DGC (**Table 4**, **Supplementary Table S2**) (Garcia-Pelaez et al. 2021b)

Regarding *BRCA1*, we found two P/LP variants in two families, one HDGC family without BC involvement and one FIGC family (**Table 3**). In fact, this FIGC family harbours co-occurring P/LP variants in *BRCA1* and *BRCA2*, where the proband was a 66 years old IGC patient, but also displaying a family history compatible with HBOC, as a family member developed BC at 37 years of age (**Table 3**). This specific case poses an important question, since a different proband choice (for instance the 37-years BC case) would have potentially triggered the suspicious of a different TRS and, consequently, a different clinical management. A similar outcome was identified in a *MSH2* variant for LS, where a deeper pedigree and segregation analysis revealed co-segregation with different cancers of the LS spectrum, including IGC and CRC (**Table 3**) (Tedaldi et al. 2019). Further, our analysis identified a single HDGC family carrying a LP in *PMS2*, and lacking CRC

in the reported family pedigree. It is possible that the DGC proband carrying the *PMS2* LP variant and its relative with GC, belong to a branch of a LS family, where GC presents as part of the spectrum, but such information is not available in the original publication (**Table 3**). Additionally, this could also be the case of both, early-onset DGC and BC diagnosis, in a proband with a TP53 P variant, (**Table 3**). This well-known tumour suppressor located in chromosome 17 is the cause of LFS. Its encoded protein is responsible for the regulation of cellular responses to stress, including DNA repair, cell cycle arrest and apoptosis(Vaseva et al. 2012). Several human cancers have been associated with *TP53* variants (OMIM Nº 114480, 114550, 607107, 260350) including hereditary cancers (**Table 2**, OMIM Nº 202300, 614740, 260500, 114500, 137800, 259500, 618165, 151623). Regarding our analysis, as *TP53* is considered a high penetrant susceptibility gene for BC, LFS syndrome could be implied (Daly et al. 2021). Given the lack of further family information, we support the possibility of a *TP53* variant in the context of early-onset BC or LFS, with involvement of DGC, rather than claiming *TP53* as a candidate gene for HDGC (Kim et al. 2004).

Altogether, these observations reinforce the usefulness of panel testing/exome sequencing for syndromes with overlapping cancer spectra, and the need for collection and publication of extended family pedigrees.

Finally, an IGC case at 19 years of age was found in a proband carrying a *BRIP1* (17q23.2) P variant (**Table 3**). *BRIP1* is also known to be associated with Fanconi Anaemia in accordance with OMIM (OMIM Nº 609054). This gene interacts with *BRCA1* and regulates *TP53*. In sporadic GC cases, *BRIP1* is specifically mutated in DGC cases (3%; 1/34), but not in IGC, which does not add to its potential role in FIGC (**Table 4**, **Supplementary Table S2**) (Garcia-Pelaez et al. 2021).

In summary, 16 genes bearing 28 P/LP variants were found in HDGC and/or FIGC representing potential candidate genes, and mainly associated with DNA repair pathways. Homologous recombination related genes were most often targeted than mismatch repair pathway gene in both GC-TRS. Moreover, *PALB2* and *BRCA2* were the most frequently and consistently mutated candidate genes in HDGC (2%) and FIGC (17%), respectively. Evidence for a direct correlation with either FIGC or HDGC for other genes was not evident from our analysis and supports the possibility of GC developing in the frame of families with unreported susceptibility to another TRS, and with GC involvement.

Impact of candidate gene analysis for clinical management of carriers

This study highlights a reasonable number of GC cases in families bearing germline variants in genes from the DNA repair machinery. This potential susceptibility may be a challenge for management of these families, as it adds a new organ to be surveilled. This would be the case for *PALB2*, *ATM*, and *BRCA2* variant carriers, in case these become formally associated with GC risk in larger and deeper studies. As we stand, it may be adequate to offer gastric surveillance with annual endoscopy in families bearing P/LP *PALB2*, *ATM*, and *BRCA2* in

case family history reveals a high number of GC cases, independently of the histological subtype and especially in countries of high incidence.

Conclusion

The acknowledgement of hereditary GC is crucial to reduce GC late diagnosis and, consequently mortality. While SNVs and CNVs have been identified in HDGC (*CDH1* and *CTNNA1* genes) and GAPPS families (*APC* promoter 1B), FIGC families remain genetically undetermined. Thus, efforts should be made to identify the underlying FIGC germline cause, establish a genetic screening protocol, and update clinical guidelines, to improve the management of these families. For HDGC and GAPPS, further research is still crucial to identify variant-specific penetrance and the significance of genetic modifiers, epigenetic alterations, or environmental factors in disease-associated risk.

Considering the complexity of these diseases independently of the GC syndrome, it is relevant to reinforce the need for patients' management to include multidisciplinary teams.

PALB2 was the most promising candidate gene in HDGC, while *BRCA2* was the most promising candidate gene in FIGC. Nevertheless, the association of these candidate genes with these TRS needs further validation, including segregation analysis, second somatic hit analysis, penetrance, and life-time risk calculations.

As we anticipated, the lack of extended family pedigrees prior to genetic testing prescription likely hampers the correct diagnosis of a TRS. This is likely the reason why so many different genes are found apparently mutated in small HDGC and FIGC families or isolated early-onset GC cases, which likely represent branches of larger families with different TRS.

This review highlights the importance of collecting extended and updated family histories, to decide on the most adequate genetic testing to be offered and the most adequate post-testing management. For FIGC, as a purer cause remains to be identified, it is important to consider these families for panel-testing including DNA damage repair genes, as these emerge as great candidates in this disease.

Acknowledgments

This research is supported by the European Reference Network on Genetic Tumour Risk Syndromes (ERN GENTURIS) – Project ID No. 739547. ERN GENTURIS is partly co-funded by the European Union within the framework of the Third Health Programme “ERN-2016 – Framework Partnership Agreement 2017–2021”.

Funding

This research is supported by the European Reference Network on Genetic Tumor Risk Syndromes (ERN GENTURIS) – Project ID No. 739547. ERN GENTURIS is partly co-funded by the European Union within the framework of the Third Health Programme “ERN-2016 – Framework Partnership Agreement 2017–2021”. Research was also funded by the European Union's Horizon 2020 research and innovation programme under grant agreement No. 779257 (Solve-RD: Ref. H2020-SC1-2017-Single-Stage-RTD) providing salary support through a Ph.D studentship to José G. Pelaez; by FEDER - Fundo Europeu de Desenvolvimento Regional funds through the COMPETE 2020-Operacional Programme for Competitiveness and Internationalisation (POCI), Portugal 2020, and by Portuguese funds through FCT/ Ministério da Ciência, Tecnologia e Inovação in the framework of the projects Ref. PTDC/BTM-TEC/30164/2017 and Ref. PTDC/BTM-TEC/6706/2020. This work was further supported by Fundação para a Ciência e Tecnologia (FCT) with salary support for the PhD studentships to Rita Barbosa-Matos, Ref. SFRH/BD/145132/2019 and Celina S. José, Ref. SFRH/BD/140796/2018, and to. University of Porto supported this work through the doctoral programme in biomedicine of the faculty of medicine to José G. Pelaez and Celina S. José, and through the international doctoral programme in molecular and cellular biotechnology applied to health sciences from Institute of Biomedical Sciences Abel Salazar to Rita Barbosa-Matos.

Conflict of interest: The authors disclosure no conflicts of interest.

References

Antoniou AC, Casadei S, Heikkinen T, et al (2014) Breast-cancer risk in families with mutations in PALB2. N Engl J Med 371:497–506. https://doi.org/10.1056/NEJMOA1400382

Beà S, Valdés-Mas R, Navarro A, et al (2013) Landscape of somatic mutations and clonal evolution in mantle cell lymphoma. Proc Natl Acad Sci U S A 110:18250–18255. https://doi.org/10.1073/PNAS.1314608110

Blair VR, McLeod M, Carneiro F, et al (2020) Hereditary diffuse gastric cancer: updated clinical practice guidelines. Lancet Oncol 21:e386–e397. https://doi.org/10.1016/S1470-2045(20)30219-9

Carvalho J, Oliveira P, Senz J, et al (2021) Redefinition of familial intestinal gastric cancer: clinical and genetic perspectives. J Med Genet 58:1–11. https://doi.org/10.1136/JMEDGENET-2019-106346

Chaudhry MS, Gilmour KC, House IG, et al (2016) Missense mutations in the perforin (PRF1) gene as a cause of hereditary cancer predisposition. Oncoimmunology 5:. https://doi.org/10.1080/2162402X.2016.1179415

Choi M, Kipps T, Kurzrock R (2016) ATM Mutations in Cancer: Therapeutic Implications. Mol Cancer Ther 15:1781–1791. https://doi.org/10.1158/1535-7163.MCT-15-0945

Daly MB, Pal T, Berry MP, et al (2021) Genetic/Familial High-Risk Assessment: Breast, Ovarian, and Pancreatic, Version 2.2021, NCCN Clinical Practice Guidelines in Oncology. J Natl Compr Canc Netw 19:77–102. https://doi.org/10.6004/JNCCN.2021.0001

Donner I, Kiviluoto T, Ristimäki A, et al (2015) Exome sequencing reveals three novel candidate predisposition genes for diffuse gastric cancer. Fam Cancer 14:241–246. https://doi.org/10.1007/S10689-015-9778-Z

El Abed R, Bourdon V, Voskoboinik I, et al (2011) Molecular study of the perforin gene in familial hematological malignancies. Hered Cancer Clin Pract 9:. https://doi.org/10.1186/1897-4287-9-9

Fewings E, Larionov A, Redman J, et al (2018) Germline pathogenic variants in PALB2 and other cancer-predisposing genes in families with hereditary diffuse gastric cancer without *CDH1* mutation: a whole-exome sequencing study. Lancet Gastroenterol Hepatol 3:489–498. https://doi.org/10.1016/S2468-1253(18)30079-7

Garcia-Pelaez J, Barbosa-Matos R, Gullo I, et al (2021) Histological and mutational profile of diffuse gastric cancer: current knowledge and future challenges. Mol Oncol 15:2841–2867. https://doi.org/10.1002/1878-0261.12948

Gaston D, Hansford S, Oliveira C, et al (2014) Germline mutations in MAP3K6 are associated with familial gastric cancer. PLoS Genet 10:. https://doi.org/10.1371/JOURNAL.PGEN.1004669

Hansford S, Kaurah P, Li-Chang H, et al (2015) Hereditary Diffuse Gastric Cancer Syndrome: *CDH1* Mutations and Beyond. JAMA Oncol 1:23–32. https://doi.org/10.1001/JAMAONCOL.2014.168

Herrera-Pariente C, Capó-García R, Díaz-Gay M, et al (2021) Identification of New Genes Involved in Germline Predisposition to Early-Onset Gastric Cancer. Int J Mol Sci 22:1–14. https://doi.org/10.3390/IJMS22031310

Huang K lin, Mashl RJ, Wu Y, et al (2018) Pathogenic Germline Variants in 10,389 Adult Cancers. Cell 173:355-370.e14. https://doi.org/10.1016/J.CELL.2018.03.039

Ichikawa H, Wakai T, Nagahashi M, et al (2018) Pathogenic germline BRCA1/2 mutations and familial predisposition to gastric cancer. JCO Precis Oncol 2:1–8. https://doi.org/10.1200/PO.18.00097

Jakubowska A, Nej K, Huzarski T, et al (2002) BRCA2 gene mutations in families with aggregations of breast and stomach cancers. Br J Cancer 87:888–891. https://doi.org/10.1038/SJ.BJC.6600562

Jin MH, Oh DY (2019) ATM in DNA repair in cancer. Pharmacol Ther 203:. https://doi.org/10.1016/J.PHARMTHERA.2019.07.002

Johannsson O, Loman N, Möller T, et al (1999) Incidence of malignant tumours in relatives of BRCA1 and BRCA2 germline mutation carriers. Eur J Cancer 35:1248–1257. https://doi.org/10.1016/S0959-8049(99)00135-5

Keller G, Vogelsang H, Becker I, et al (2004) Germline mutations of the E-cadherin(*CDH1*) and TP53 genes, rather than of RUNX3 and HPP1, contribute to genetic predisposition in German gastric cancer patients. J Med Genet 41:. https://doi.org/10.1136/JMG.2003.015594

Kim IJ, Kang HC, Shin Y, et al (2004) A TP53-truncating germline mutation (E287X) in a family with characteristics of both hereditary diffuse gastric cancer and Li-Fraumeni syndrome. J Hum Genet 49:591–595. https://doi.org/10.1007/S10038-004-0193-9

Kim IJ, Park JH, Kang HC, et al (2003) A novel germline mutation in the MET extracellular domain in a Korean patient with the diffuse type of familial gastric cancer. J Med Genet 40:. https://doi.org/10.1136/JMG.40.8.E97

Law RHP, Lukoyanova N, Voskoboinik I, et al (2010) The structural basis for membrane binding and pore formation by lymphocyte perforin. Nature 468:447–451. https://doi.org/10.1038/NATURE09518

Lee JH, Han SU, Cho H, et al (2000) A novel germ line juxtamembrane Met mutation in human gastric cancer. Oncogene 19:4947–4953. https://doi.org/10.1038/SJ.ONC.1203874

Li R, Zhang H, Cao Y, et al (2020) Lauren classification identifies distinct prognostic value and functional status of intratumoral CD8+ T cells in gastric cancer. Cancer Immunol Immunother 69:1327–1336. https://doi.org/10.1007/S00262-020-02550-7

Lin J, Shi J, Guo H, et al (2019) Alterations in DNA Damage Repair Genes in Primary Liver Cancer. Clin Cancer Res 25:4701–4711. https://doi.org/10.1158/1078-0432.CCR-19-0127

Lu C, Xie M, Wendl MC, et al (2015) Patterns and functional implications of rare germline variants across 12 cancer types. Nat Commun 6:. https://doi.org/10.1038/NCOMMS10086

Navrkalova V, Sebejova L, Zemanova J, et al (2013) ATM mutations uniformly lead to ATM dysfunction in chronic lymphocytic leukemia: application of functional test using doxorubicin. Haematologica 98:1124–1131. https://doi.org/10.3324/HAEMATOL.2012.081620

Pearlman R, Frankel WL, Swanson B, et al (2017) Prevalence and Spectrum of Germline Cancer Susceptibility Gene Mutations Among Patients With Early-Onset Colorectal Cancer. JAMA Oncol 3:464–471. https://doi.org/10.1001/JAMAONCOL.2016.5194

Praper T, Sonnen A, Viero G, et al (2011) Human perforin employs different avenues to damage membranes. J Biol Chem 286:2946–2955. https://doi.org/10.1074/JBC.M110.169417

Richards S, Aziz N, Bale S, et al (2015) Standards and guidelines for the interpretation of sequence variants: a joint consensus recommendation of the American College of Medical Genetics and Genomics and the Association for Molecular Pathology. Genet Med 17:405–424. https://doi.org/10.1038/GIM.2015.30

Roy R, Chun J, Powell SN (2011) BRCA1 and BRCA2: different roles in a common pathway of genome protection. Nat Rev Cancer 12:68–78. https://doi.org/10.1038/NRC3181

Sahasrabudhe R, Lott P, Bohorquez M, et al (2017) Germline Mutations in PALB2, BRCA1, and RAD51C,

Which Regulate DNA Recombination Repair, in Patients With Gastric Cancer. Gastroenterology 152:983-986.e6. https://doi.org/10.1053/J.GASTRO.2016.12.010
Savitsky K, Bar-Shira A, Gilad S, et al (1995) A single ataxia telangiectasia gene with a product similar to PI-3 kinase. Science 268:1749–1753. https://doi.org/10.1126/SCIENCE.7792600
Setia N, Wang CX, Lager A, et al (2021) Morphologic and molecular analysis of early-onset gastric cancer. Cancer 127:103–114. https://doi.org/10.1002/CNCR.33213
Slavin T, Neuhausen SL, Rybak C, et al (2017) Genetic Gastric Cancer Susceptibility in the International Clinical Cancer Genomics Community Research Network. Cancer Genet 216–217:111–119. https://doi.org/10.1016/J.CANCERGEN.2017.08.001
Takaoka M, Miki Y (2018) BRCA1 gene: function and deficiency. Int J Clin Oncol 23:36–44. https://doi.org/10.1007/S10147-017-1182-2
te Paske IBAW, Garcia-Pelaez J, Sommer AK, et al (2021) A mosaic PIK3CA variant in a young adult with diffuse gastric cancer: case report. Eur J Hum Genet 29:1354–1358. https://doi.org/10.1038/S41431-021-00853-6
Tedaldi G, Pirini F, Tebaldi M, et al (2019) Multigene Panel Testing Increases the Number of Loci Associated with Gastric Cancer Predisposition. Cancers (Basel) 11:. https://doi.org/10.3390/CANCERS11091340
Tischkowitz M, Xia B (2010) PALB2/FANCN: recombining cancer and Fanconi anemia. Cancer Res 70:7353–7359. https://doi.org/10.1158/0008-5472.CAN-10-1012
Tung N, Battelli C, Allen B, et al (2015) Frequency of mutations in individuals with breast cancer referred for BRCA1 and BRCA2 testing using next-generation sequencing with a 25-gene panel. Cancer 121:25–33. https://doi.org/10.1002/CNCR.29010
Tung N, Lin NU, Kidd J, et al (2016) Frequency of Germline Mutations in 25 Cancer Susceptibility Genes in a Sequential Series of Patients With Breast Cancer. J Clin Oncol 34:1460–1468. https://doi.org/10.1200/JCO.2015.65.0747
Vaseva A V., Marchenko ND, Ji K, et al (2012) p53 opens the mitochondrial permeability transition pore to trigger necrosis. Cell 149:1536–1548. https://doi.org/10.1016/J.CELL.2012.05.014
Velázquez C, Esteban-Cardeñosa EM, Lastra E, et al (2019) A PALB2 truncating mutation: Implication in cancer prevention and therapy of Hereditary Breast and Ovarian Cancer. Breast 43:91–96. https://doi.org/10.1016/J.BREAST.2018.11.010
Vogelaar IP, Van Der Post RS, Van Krieken JHJM, et al (2017) Unraveling genetic predisposition to familial or early onset gastric cancer using germline whole-exome sequencing. Eur J Hum Genet 25:1246–1252. https://doi.org/10.1038/EJHG.2017.138
Weren RDA, Van Der Post RS, Vogelaar IP, et al (2018) Role of germline aberrations affecting *CTNNA1*, MAP3K6 and MYD88 in gastric cancer susceptibility. J Med Genet 55:669–674. https://doi.org/10.1136/JMEDGENET-2017-104962
Yang C, Arnold AG, Trottier M, et al (2016) Characterization of a novel germline PALB2 duplication in a hereditary breast and ovarian cancer family. Breast Cancer Res Treat 160:447–456. https://doi.org/10.1007/S10549-016-4021-7
Zhou J, Zhao Z, Zhang Y, et al (2020) Pathogenic Germline Mutations in Chinese Patients with Gastric Cancer Identified by Next-Generation Sequencing. Oncology 98:583–588. https://doi.org/10.1159/000505664

Original paper 1

Combined analysis of chromatin accessibility and genome-wide loss of function variants reveals missing heritability in cancer

Celina São José[1,2,3], Marta Ferreira[1,2], Pelaez-Garcia J.[1,2,3], Ana André[1,2], Senz J.[4], Kaurah P.[4], Cordova L.[4], Schrader K.[4], Huntsman D.[4,5], Fiona Puntieri[7], Juliane Glasser[7], Stefan Mundlos[7,8,9], Carla Oliveira[1,2,10]*

[1] i3S, Institute for Research and Innovation in Health, University of Porto, Porto, Portugal

[2] IPATIMUP, Institute of Molecular Pathology and Immunology, University of Porto, Porto, Portugal

[3] Doctoral Programme in Biomedicine, Faculty of Medicine, University of Porto, Porto, Portugal

[4] Department of Molecular Oncology, British Columbia Cancer Research Centre

[5] Department of Pathology and Laboratory Medicine, University of British Columbia, Vancouver, British Columbia

[6] Department of Pathology and Laboratory Medicine, University of British Columbia, Vancouver, British Columbia

[7] Max Planck Institute for Molecular Genetics, RG Development & Disease, Berlin, Germany

[8] Institute of Medical and Human Genetics, Charité Universitätsmedizin Berlin, Berlin, Germany

[9] Berlin-Brandenburg Center for Regenerative Therapies (BCRT), Charité Universitätsmedizin Berlin, Berlin, Germany

[10] Faculty of Medicine of the University of Porto, Porto, Portugal

Manuscript *in preparation*.

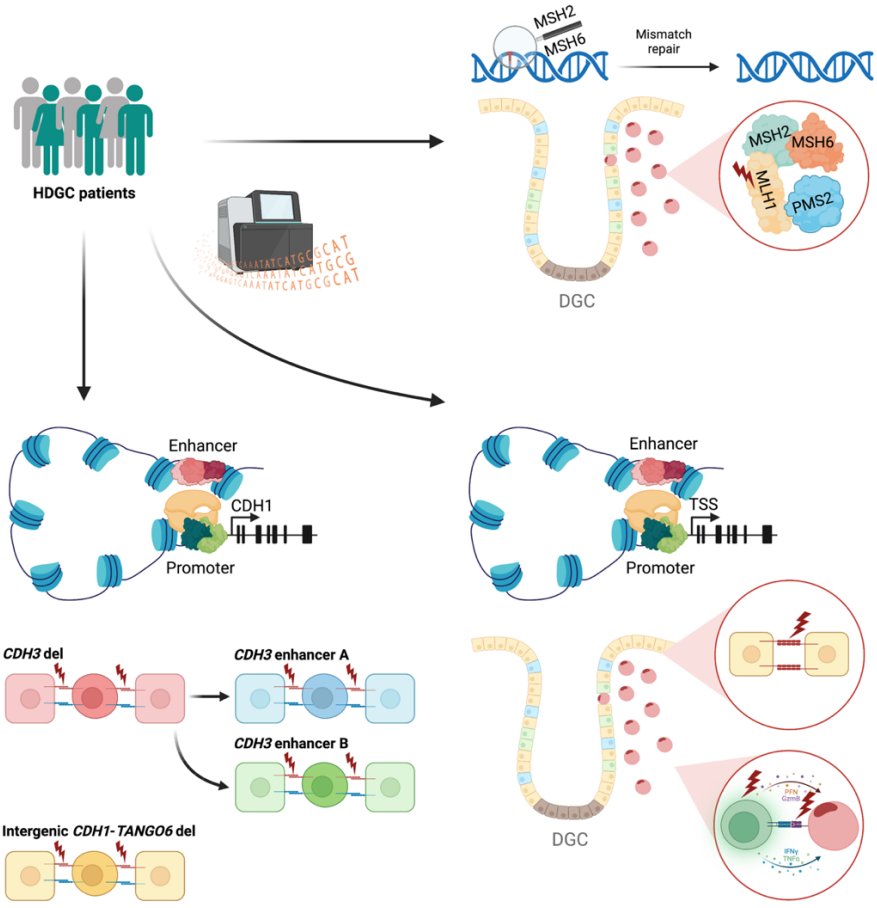
Author statement

Celina São José: Investigation, methodology, conceptualization, formal analysis, *in vitro* experimental work, bioinformatic analysis, *in vivo* experimental work, writing – original draft, writing – review and editing.

Unveil the germline landscape of HDGC patients, pinpointing *CDH1* potential regulatory elements

In an attempt to solve the missing heritability in HDGC patients, we WGS sequenced the gDNA of HDGC-like individuals lacking a genetic diagnosis. The absence of *CDH1* and *CTNNA1* genetic defects pinpointed regulatory elements as the main HDGC trigger in these families. Briefly, we found a *MLH1* coding deletion in a family with a bias for GC development. This deletion encompasses a putative regulatory element present in normal stomach but absent in normal colonic mucosa. Within the *CDH1* TAD, we found a large deletion in the *CDH3* upstream gene, encompassing a *CDH1*-interactor with accessible chromatin and a *CDH1*-eQTL, whose CRISPR-Cas9 deletion impacts *CDH1*/E-cadherin expression. A small deletion was also identified in the *CDH1-TANGO6* downstream intergenic region, which deleted downregulates *CDH1*/E-cadherin expression. The germline landscape of HDGC-like families displays downregulation of immune-associated pathways, such as MHCI and natural killer mediated immunity, that may increase immune evasion. Further, young affected individuals in HDGC-like families display increased impact genome-wide stomach regulatory elements, than older affected individuals.

In summary, we found the missing heritability in 3 out of 21 HDGC-like families, corresponding to 14% of the entire cohort and pinpointed *CDH1* putative regulatory elements as regulators of *CDH1*/E-cadherin expression impacted in HDGC families.



Abstract

Germline defects in *CDH1* and *CTNNA1* genes predispose to hereditary diffuse gastric cancer (HDGC), a syndrome that confers risk to develop diffuse gastric cancer (DGC) and lobular breast cancer (LBC). However, most HDGC-like remain genetically unexplained, despite the presence of *CDH1* monoallelic downregulation, a common feature observed in families bearing coding variants in the *CDH1* gene. Working under the hypothesis that regulatory elements controlling *CDH1* expression may predispose to HDGC, we whole genome sequenced the gDNA of 17 HDGC-like families and searched for copy number variants (CNV) overlapping the coding region and truncating single nucleotide variants (SNV) in cancer syndrome-associated genes.

We found a very rare *MLH1* germline deletion in a family with multiple cases of DGC, an no other cancer types. The CNV overlaps enhancer chromatin marks in normal stomach, absent in normal colon mucosa, pinpointing this region as a gastric-specific regulatory element that may explain the presence of GC alone in this family. To further dissect the causing germline defects in HDGC-like families, we explored the *CDH1 cis*-regulatory network. We found a 20kb deletion in the upstream gene *CDH3*, comprising a *CDH1*-interactor with accessible chromatin and *CDH1*-eQTL. CRISPR-Cas9 deletion of *CDH1*-interactor and *CDH1*-eQTL putative regulatory elements impact *CDH1* expression in 50%. In addition, the *CDH1*-eQTL displays strong tissue specific enhancer activity in forebrain and mild activity in the stomach. The cancer types observed in this family may be associated with *CDH1* or *CDH3* expression loss, a result from regulatory element loss or protein loss itself, respectively. We found a small deletion within the *CDH1-TANGO6* intergenic region in a family with late onset DGC, and the CRISPR-Cas9 deletion of the entire intergenic region results in 30% *CDH1* downregulation. The germline landscape of young DGC patients in HDGC-like families encompass immune-associated pathways downregulation, namely MHC I and natural killer immunity, that may play a role in immune evasion. These patients also display multiple deletions overlapping genome-wide open chromatin region in normal stomach.

In conclusion, we found the missing heritability in 3 out of 21 HDGC-like families, corresponding to 14% of the entire cohort, and pinpointed *CDH1* putative regulatory elements as regulators of *CDH1*/E-cadherin expression impacted in HDGC families. The germline landscape of younger DGC patients includes downregulation of immune-associated pathways, that may play a role in immune evasion.

Introduction

Hereditary diffuse gastric cancer (HDGC) predisposes to early onset diffuse gastric cancer (DGC) and lobular breast cancer (LBC) development (Caldas et al. 1999). Single nucleotide variants (SNV) in *CDH1* and *CTNNA1* and copy number variants (CNV) in *CDH1* are the main genetic triggers for HDGC genetic predisposition (Guilford et al. 1998b; Majewski et al. 2013b). *CDH1* and *CTNNA1* genes encode E-cadherin and αE-catenin proteins, respectively, both adhesion partners in the cell adhesion complex. Their impairment in HDGC context results in cell architecture loss and decreased differentiation, characterized by the presence of signet-ring cells, a DGC signature (Gullo et al. 2020). A positive result triggers cascade genetic testing and life-saving measures, encompassing endoscopic surveillance or prophylactic surgery of stomach and breast for at-risk individuals (Blair et al. 2020a).

Despite the presence of signet-ring cells and *CDH1* monoallelic downregulation, pinpointing *CDH1* as the major gene triggering HDGC, most HDGC families remain genetically undiagnosed (Pinheiro et al. 2010). Thus, we hypothesize that regulatory elements may regulate *CDH1* gene expression, beyond their coding sequence.

Regulatory elements are instrumental for proper gene expression in time and space, for which enhancers play an important role (Spitz and Furlong 2012b). Their mode of action is usually redundant, modular, orientation independent and through the binding of tissue-specific transcription factors (Banerji et al. 1981; Lettice et al. 2003; Will et al. 2017). Enhancer-promoter communication potentiates gene expression, which may occur through long-physical distances in the DNA sequence, but in close spatial proximity in 3D-chromatin folding (Bonev and Cavalli 2016b; Cruz-Molina et al. 2017). Topologically-associating domains (TAD) are the functional unit of 3D-chromatin architecture that retains most enhancer-promoter interactions (Dekker et al. 2002). CNVs encompassing one enhancer are generally inefficient in producing gene expression loss, yet deletion of several enhancers may lead to downregulation of their target gene (Will et al. 2017). Yet, assigning function and pathogenicity to non-coding variants in the next-generation sequencing era remains challenging. Thus, unravelling the function of enhancers overlapping CNVs will improve their variant classification. Herein, we propose that regulatory elements control *CDH1* expression, and their impairment may result in HDGC development.

Materials and Methods

Patient selection

Nineteen HDGC patients from various genetic backgrounds were admitted in British Columbia and selected based on 1999 HDGC clinical criteria (Caldas et al. 1999). The 21 patients tested negative for *CDH1* SNVs and CNVs.

Whole genome sequencing

Genomic DNA of 21 HDGC patients was sequenced using illumina platform. Sequencing reads were mapped to the human reference genome (GRCh38) using BWA mem (v.0.7.15). Duplicated reads were marked, and their base quality score was recalibrated with GATK (v4.1.9.0, HTSJDK 2.23.0 and Picard 2.23.3).

Variant calling

SNVs were called with GATK – HaplotypeCaller tool (v4.1.9.0, HTSJDK 2.23.0 and Picard 2.23.3), and Depp Variant (Docker image: 1.3.0) (Poplin et al. 2018). Called SNVs were filtered according to quality control metrics, read depth (≥8) and genotype quality (≥20), and annotated with VEP (v.107). SNVs were prioritized if called by (1) both callers, from which HaplotypeCaller had high quality (≥200), (2A) Depp Variant, or (2B) HaplotypeCaller with extremely high quality (≥500). CNVs calling was performed with LUMPY (v 0.2.13) (Layer et al. 2014), DELLY (v.0.9.1) (Rausch et al. 2012) and GRIDSS (v.2.13.2) (Cameron et al. 2017), and annotated with AnnotSV (v. 2.5.2, Tcl (v. 8.6)) (Geoffroy et al. 2018). Structural variants with ≥1Mb length, inter-chromosomal events and variants lacking 'PASS' flag were discarded. CNVs were prioritized based on the following criteria: (1) LUMPY and DELLY with ≥90% overlap, and GRIDSS confirmed breakpoints with ≤4bp difference, (2A) DELLY and GRIDSS confirmed breakpoints with ≤4bp difference, (2B) LUMPY and GRIDSS confirmed breakpoints with ≤4bp difference, and (2C) LUMPY and DELLY with ≥90% overlap.

Chromatin available data

ChIP-seq data of H3K27Ac from normal stomach and colonic mucosa was collected from phase 3 ENCODE project (Abascal et al. 2020). GeneHancer enhancers and eQTLs were collected from the GeneHancer project (Fishilevich et al. 2017).

Cell lines and culture conditions

Human cell line MKN74 was purchased from the JCRB Cell Bank. MKN74 cell lines were cultured in RPMI medium supplemented with 10% foetal bovine serum (FBS) and maintained at 37ºC and 5% CO_2 in a high humidity atmosphere. Cell identification was confirmed by STR analysis and cells were confirmed to be free of mycoplasma contamination.

In vitro enhancer assay by CRISPR-Cas9

To test enhancer function, each region was targeted with CRISPR-Cas9. CRISPR-Cas9 editing. sgRNAs were designed using Benchling online platform (**supplementary Table 1**). Individual sgRNAs (Invitrogen) were cloned in LentiCRISPRv2GFP (addgene 82416) or LentiCRISPRv2-mCherry (addgene 99154) vectors using BsmBIv2 (New England Biolabs). Plasmids were transformed into Stbl3 competent cells and colonies were sequenced (**supplementary Table 1**). Lentiviral particles were produced resourcing to HEK293T cell line with pMD2.G (addgene 12259) and pCMV-dR8.91 (addgene) vectors, following Lipofectamine 3000 manufacture's protocol (Invitrogen) and collected at 48h. MKN74 was infected with pairs of lentivirus particles in medium supplemented with 10µg/µl hexadimethrine bromide (Merck Life Science S.L.U.) for 48h. Transduced cells were selected for GFP and mCherry positive expression at 7 days post-infection using FACS ARIA (BD Biosciences).

Genotyping of edited clones

gDNA was extracted using NZY Tissue gDNA isolation kit (NZYTech), according to the manufacturers' protocol. gDNA was amplified using primers flanking the edition sites (**supplementary Table 1**) and Multiplex PCR kit (Qiagen). PCR products were analysed in gel electrophoresis and Sanger sequenced on an ABI-3130 Genetic Analyzer (Applied Biosystems).

CDH1/E-cadherin and CDH3/P-cadherin expression analysis

CDH1/*3* mRNA expression was assessed by qPCR in triplicates. RNA was extracted using mirVana RNA Isolation Kit (Invitrogen), according to manufacturers' protocol. cDNA was synthesized using SuperScriptII reverse transcriptase (Invitrogen), according to the manufacturers' protocol. *CDH1* mRNA expression was analyzed by qPCR with KAPA PROBE FAST qPCR Master Mix (2X) Kit (Sigma-Aldrich) and probes for *CDH1* (Hs.PT.58.3324071, TaqMan), *CDH3* (Hs.PT.51.5028751, IDT) and 18S (custom assay, IDT). Reactions were sequenced on a 7500 Real-Time PCR System (Applied Biosystems). Relative expression was normalized for the endogenous 18S control and quantified using the $2^{-\Delta\Delta Ct}$ method.

E-cadherin and P-cadherin expression was assessed by flow cytometry in triplicates. Cells were detached with Versene (Gibco) and blocked with 3% bovine serum albumin-phosphatase buffer saline. Cells were incubated with primary mouse monoclonal antibody HECD-1 or P-cadherin (1:100 dilution; 1h at 4ºC; Invitrogen), washed and incubated with secondary antibody anti-mouse Alexa Fluor 647 (Invitrogen). Fluorescence was measured using FACS Fortessa (BD Biosciences) and Flow Jo version 10 software was used to analyse the data.

Statistical analysis

Statistical analysis was performed using GraphPad Prism version 7.00 software (GraphPad Software Inc.). A T-student test was used for comparison analysis, assuming equal variance between clones and parental samples. Differences were considered significant when p-value<0.05.

Mouse LacZ Enhancer Assay

Enhancer activity was studied resourcing to a LacZ reporting vector. Briefly, selected regions were PCR amplified from human gDNA, using PrimeSTAR GXL DNA Polymerase (Takara Bio) with overhang primers appropriate for Gibson assembly (**supplementary Table 2**). Enhancer candidate regions were cloned into a LacZ expressing vector with Gibson assembly and transfected into mouse embryonic stem cells containing an attP integration site and targeted by PhiC31. Successfully transfected cells were selected with puromycin. Transgenic embryos were generated from positive genetically modified mouse embryonic stem cells by diploid aggregation (Artus and Hadjantonakis 2011) and tested for LacZ reporter activity at E14.5. CD1 female mice were used as foster mothers. Tissues from E14.5 mouse embryos were dissected, fixed in 4% PFA for 10 minutes and stained in X-gal solution for a few hours to overnight, as described previously (Lobe et al. 1999). Dissected organs were imaged using a ZEISS SteREO Discovery.V12 with cold light source CL9000 microscope and Leica DFC420 digital camera (Leica Microsystems). Embryos were stored at 4 °C in 4% PFA in 1x PBS.

4C-seq

Fresh gastric tissue from bariatric surgeries was collected and washed in HBSS 1x (Gibco). Gastric mucosa cells were scraped and enzymatically dissociated with collagenase type I (Merck), dispase (Merck), trypsin inhibitor (Sigma), BSA (Nzytech), dithiothreitol (Invitrogen) and HBSS 1x (Gibco) at 37ºC and 150rpm for at least 1h. 4C-seq libraries were generated, as previously described (Van De Werken et al. 2012; Krijger et al. 2020), containing $1x10^7$ cells, cross-linked in 2% paraformaldehyde and lysed. Nuclei suspensions were digested with DnpII (New England Biolabs) as primary and Csp6I (New England Biolabs) as secondary restriction enzymes and re-circularized with T4 DNA Ligase (Thermo Fisher Scientific). 4C-seq libraries were purified using Amicon Ultra-15 10 kDa (MWCO) (Millipore) and PCR-amplified with 3.2µg per reaction (**supplementary Table 3**). Samples were paired-end sequenced on Illumina HiSeqX technology (150bp reads) according to standard protocols. *CDH1* interactions on a genome scale were mapped from sequenced 4C libraries, using a bioinformatics pipeline based on Pipe4C (Krijger et al. 2020) and PeakC (Geeven et al. 2018) with window size 2, alpha fdr 0.1 and minimal distance 500.

Results

Cancer-associated genes may predispose to HDGC

Coding SNVs and CNVs in the *CDH1* gene and SNVs in the *CTNNA1* gene are the main germline defects causing HDGC. Thus, we sequenced the gDNA of 21 HDGC patients meeting 2020 clinical criteria. Description of the cohort is depicted in **supplementary Figure 1**. However, all 21 HDGC patients lack coding germline variants in the *CDH1* and *CTNNA1* genes.

In an attempt to solve the missing heritability in these HDGC families, we resourced to the WGS sequencing of these HDGC patients lacking known HDGC causing defects and merged ultra rare (AF<0.00001) SNVs and CNVs with cancer syndrome-associated genes depicted in **supplementary Table 4**. We found a deletion in *MLH1* exon 13 (out-of-frame) and part of introns 12 and 13 (**Figure 1A,B**). *MLH1* is essential for mismatch repair of double-stranded breaks and a well-described gene in lynch syndrome, that predisposes to colorectal cancer, endometrial cancer and in a small proportion gastric cancer, corresponding to 13% cumulative risk by 75 years of age (Møller et al. 2018).

As the pedigree evidences, this family displays two very early-onset DGCs, at 34 and 40 years of age, and one gastric cancer that died at 43 years of age, indicative of being a mutation carrier (**Figure 1C**). The high gastric cancer frequency in this family, lacking colorectal cancer development, prompt the next question of this study, related with the role of stomach-specific regulatory elements in cancer development in individuals bearing this CNV. Indeed, the CNV encompasses H3K27Ac peaks in normal stomach, chromatin marks of promoters and enhancers, absent in normal colon epithelia (**Figure 1B**). These marks hint this region as a *MLH1* potential tissue-specific regulatory element in stomach and pinpoint a biological mechanism for gastric cancer development bias in this family, despite an impact in the *MLH1* coding sequence (Llorens-Bobadilla et al. 2023). Thus, this family likely displays lynch syndrome with a biased predisposition to the stomach.

A

Gene	Syndrome	Chr	Start hg38	End hg38	SV type	AF
MLH1	Lynch syndrome	3	37039111	37041898	DEL	novel

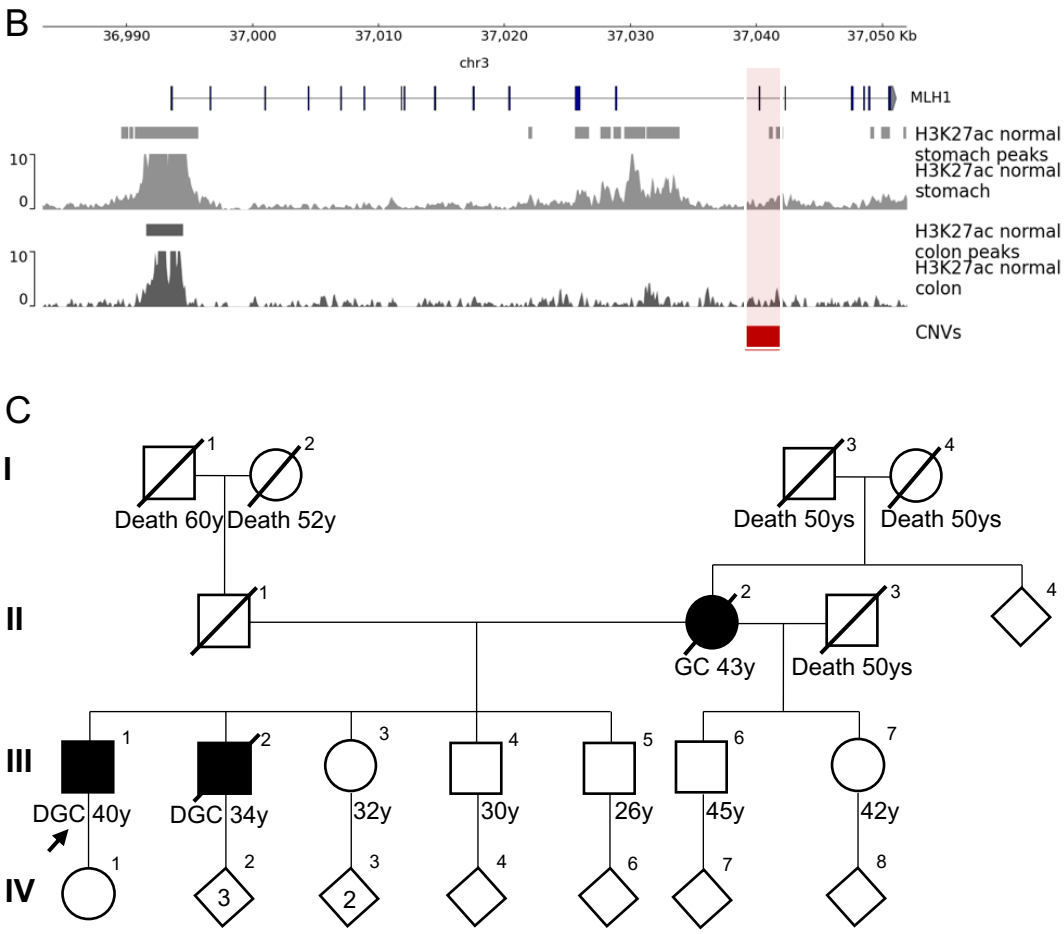


Figure 1. Genomic CNVs in cancer associated syndromes. *A) Description of CNV identified, B) Chromatin marks and accessibility in normal stomach and colon tissue from ENCODE, C) Family pedigree. Black fill represent gastric cancer development. DGC: diffuse gastric cancer, GC: gastric cancer.*

CDH1 regulatory network solves the HDGC missing heritability

Given the lack of variants in HDGC- or other cancer syndromes-associated genes in most families, we explored the *CDH1* regulatory network within its TAD for CNVs overlapping *CDH1* putative regulatory elements. TADs display increased enhancer-promoter communications, thus displaying an increased potential for regulatory elements identification. Herein, we found a new coding deletion that encompasses the *CDH3* gene between *CDH3* exon 1 and part of intron 2, and two inversions in the boundaries of this deletion in the same patient (**Figure 2A,B**). While validating the three genetic defects by Sanger sequencing, we found evidence of a WT allele (a), a 20kb-deleted allele (b) and an allele with two inversions in the boundaries of the deletion, but with a WT sequence in between

(c) (**Figure 2B**). While the WT and the deleted alleles (a and b) have been validated resourcing to Pacbio long-read sequencing and sanger sequencing, the inverted allele breakpoints have not yet been validated. The *CDH3* gene encodes for P-cadherin, a protein with similar functions to E-cadherin but with distinct expression patterns in different epithelial tissues. While E-cadherin is mainly expressed in stomach and intestine epithelia, P-cadherin expression is restricted to organ regions, such as myoepithelial cells of the breast or mesothelia (Daniel et al. 1995). In an attempt to dissect the CNV impact in *CDH1* expression and HDGC development, we mimicked the 20kb deletion within the *CDH3* gene resourcing to CRISPR-Cas9. We found complete *CDH1*/E-cadherin and *CDH3*/P-cadherin expression loss both at the mRNA and protein levels (**Figure 2C,D, supplementary Figure 2A,B**). This result pinpoint this *CDH3* 20kb region as a potential regulator of *CDH1* expression, that may compose enhancer regions within its coding or non-coding regions. Indeed, we found two potential *CDH1* regulatory elements within *CDH3* intron 2, an accessible chromatin region interacting with the *CDH1* promoter in normal stomach epithelia and an eQTL with a *CDH1* promoter interaction (**Figure 2C**). CRISPR-Cas9 deletion of the accessible chromatin region leads to 60% *CDH1*/E-cadherin loss at the RNA and protein levels, representing half of the impact of deleting the complete 20kb region (**Figure 2D**). While the eQTL deletion promotes 40% *CDH1*/E-cadherin downregulation at the RNA and protein levels (**Figure 2D**). Enhancer reporter assays in mouse embryos reveal accessible chromatin region and eQTL strong enhancer activity in forebrain (**Figure 2E**). While the accessible chromatin region has strong enhancer activity in the closing palate, the eQTL has mild activity in the stomach, correlating well with both tissues expressing *CDH1* (**Figure 2E**). Interestingly, this family displays three gastric cancers, at the ages of 38 and 39, one central nervous system tumour at 40s, one breast cancer, one ovarian cancer at 27 and one skin cancer at 52 years of age (**Figure 3A**). We may hypothesize that the 20kb deletion in the *CDH3* gene may lead to cancer development in this family from either or both *CDH1* and *CDH3* loss of expression through the deletion of a *CDH1* regulatory element and *CDH3* promoter, exon 1 and 2, respectively. From the *CDH1* regulation standpoint, we may obtain *CDH1* expression loss and, consequently, gastric or breast cancer development, while from the *CDH3* gene expression standpoint, we may find *CDH3* expression loss and CNS, ovarian and skin cancer development. From a heritability standpoint, we expect that inheriting the 1) WT and the deleted alleles (a and b) results in increased risk for DGC and *CDH3*-associated cancers development, 2) WT and inverted alleles (a and c) leads to *CDH3*-associated cancers development increased risk alone, and 3) deleted and inverted alleles (b and c) increases the risk for DGC and *CDH3*-associated cancers development (**Figure 3D**).

We found a second family displaying later-onset gastric cancer development, encompassing 4 gastric cancers, at 63, 69, 42, 67 years of age, and 1 colorectal cancer at age 55 (**Figure 3B**). This family carries a 40bp deletion within the intergenic region of the *CDH1* and *TANGO6* genes, bearing *CDH1*-promotor interactors in normal stomach (**Figure 2A,B**). Deletion of the entire *CDH1*-

TANGO6 intergenic region leads to 30% *CDH1*/E-cadherin downregulation (**Figure 2D**). Interestingly, a family with a highly penetrant and aggressive HDGC phenotype with earlier-onset DGC development carrying a *CDH1-TANGO6* deletion has been identified previously (**original paper 2**). The authors propose a role beyond the *CDH1 locus* due to the massive impact of the *CDH1-TANGO6* deletion in the *CDH1* expression itself, in immune, proliferation and adhesion-associated pathways compared with the *CDH1* deletion alone (**original paper 2**). Herein, we suggest that regulatory elements present in the *CDH1-TANGO6* intergenic region and controlling *CDH1* expression may predispose the present family to HDGC development at a latter age of onset and increase disease penetrance, decreasing age of onset in the previously identified family.

So far, we solved the missing heritability in 3 out of 21 HDGC-like patients, representing 14% of the cohort, resourcing to either a cancer syndromes-associated genes approach or a *CDH1* regulatory network approach.

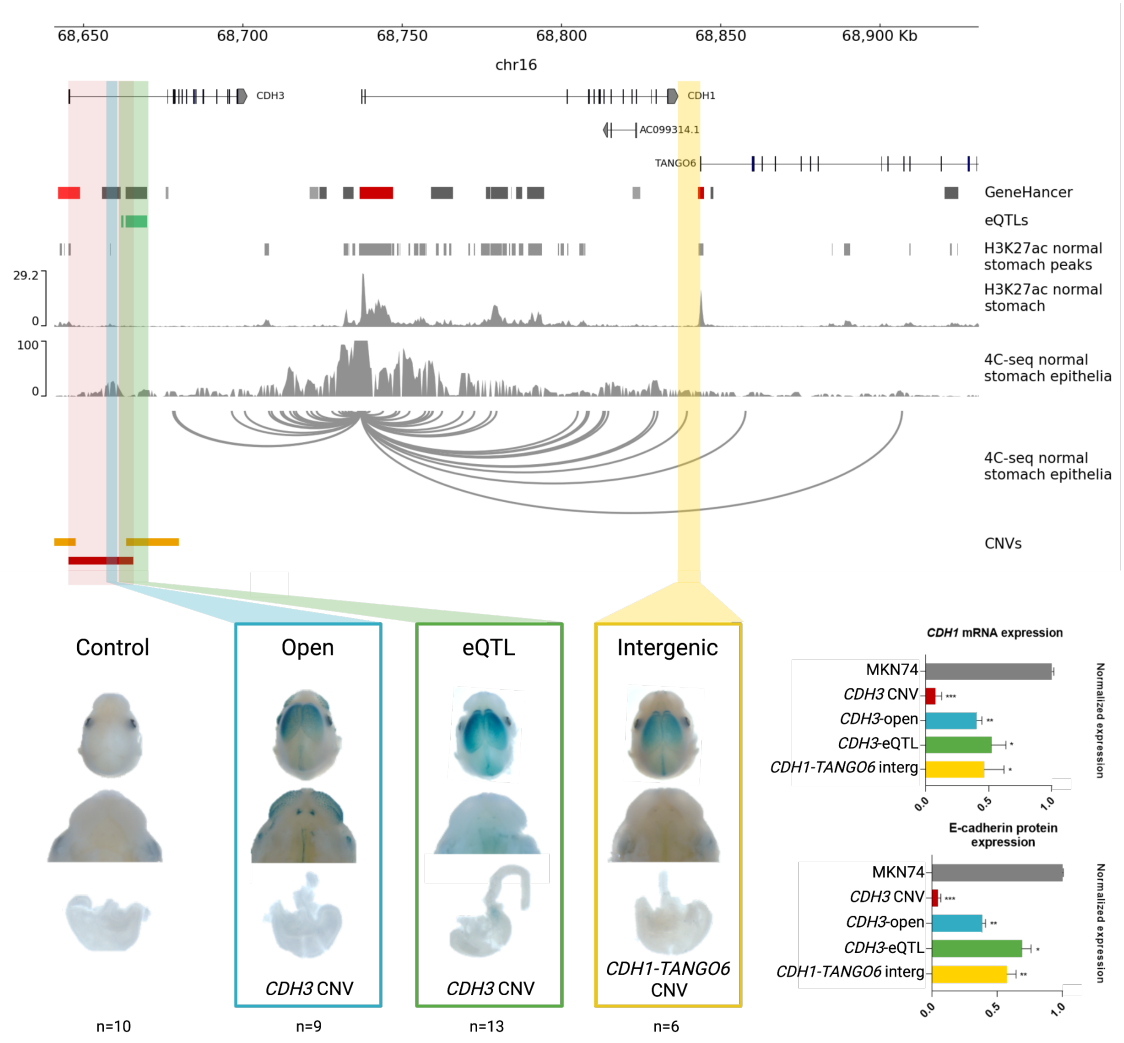


***Figure 2. Genomic CNVs in CDH1 TAD**. A) Description of identified CNVs, B) Validated alleles by Sanger sequencing, C) Putative enhancers, chromatin marks, and CDH1 interactors in normal stomach tissue, D) CDH1 mRNA and protein expression in edited clones, E) Mouse enhancer assay of putative regulatory elements.*

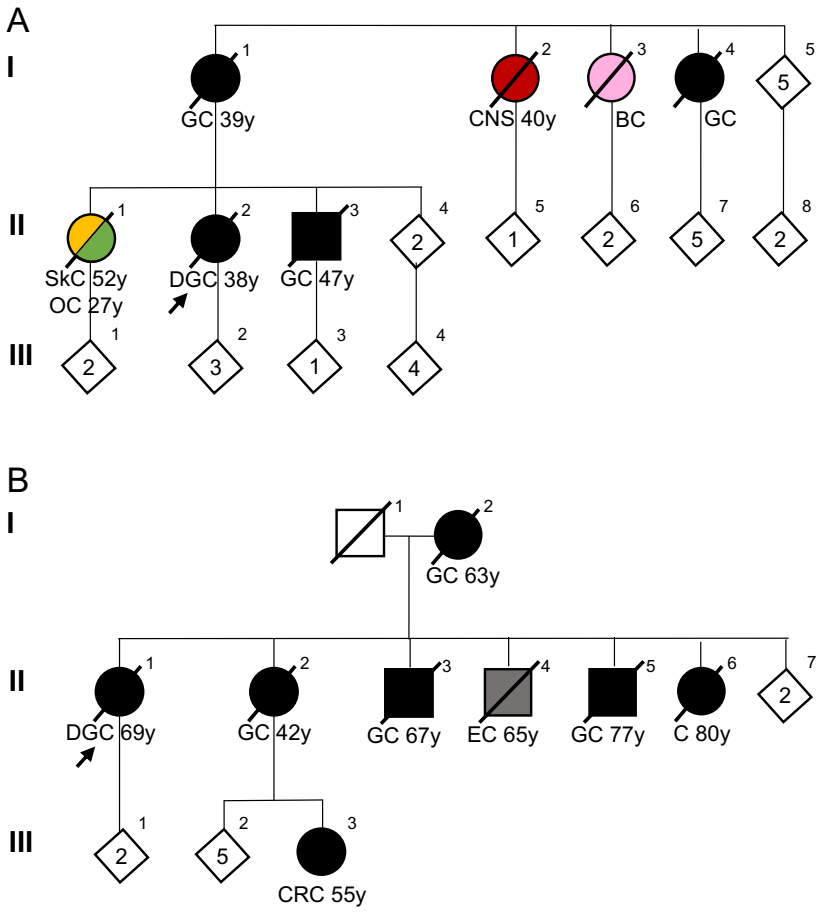


Figure 3. Family pedigrees. *A) Family pedigree of CDH3 CNV family, B) Family pedigree of CDH1-TANGO6 intergenic CNV. Black fill represents gastric cancer, red represent central nervous system cancer, light blue represents breast cancer, yellow represent skin cancer, green represent ovarian cancer, grey represent endometrial cancer and blue represent colorectal cancer. DGC: diffuse gastric cancer, GC: gastric cancer, CNS: central nervous system cancer, BC: breast cancer, SC: skin cancer, OC: ovarian cancer, EC: endometrial cancer, CRC: colorectal cancer*

HDGC germline landscape

Besides gastrointestinal syndromes-associated genes, we explored the germline landscape of HDGC patients in an attempt to identify novel germline signatures that might play a milder role in HDGC development.

Genomic alterations that have a straight-forward impact on disease encompass CNVs (deletions) and truncating SNVs. Considering these in each sample, we selected variants prioritized by at least two callers, ultra-rare in the general population (AF<0.00001) and relatively common in the cohort (<6 patients, <24%). We found several genes and pathways impaired upon these genomic variants (**Figure 4A,B**). Among these pathways, we found a cluster impacted in most HDGC patients (**Figure 4B**). Immune-associated pathways, encompassing antigen-presenting through major histocompatibility complex (MHC) I, natural killer, and T cell mediated immunity have been found impaired in HDGC-like families (**Figure 4B**). These pathways were absent for most families when

considering CNVs or SNVs alone, suggesting a synergistic role of both in HDGC germline landscape (**supplementary Figure 3**). Interestingly, immune-associated pathways have been found downregulated in clones mimicking a *CDH1-TANGO6* large deletion present in patients diagnosed with earlier onset DGC, with mean age of cancer onset of 29±7.55 years (**original paper 2**). Downregulation of MHC I and natural killer mediated immunity play a role in cancer evading immune system (Shen et al. 2005; Lee et al. 2011; Na et al. 2021). This may result in absence of immune response in early stages of cancer development and, therefore, DGC earlier onset, in the presence of a still unknown genetic trigger. Altogether, the combination of these downregulated pathways may create a genomic background favourable for DGC development, representing a potential germline signature of HDGC patients.

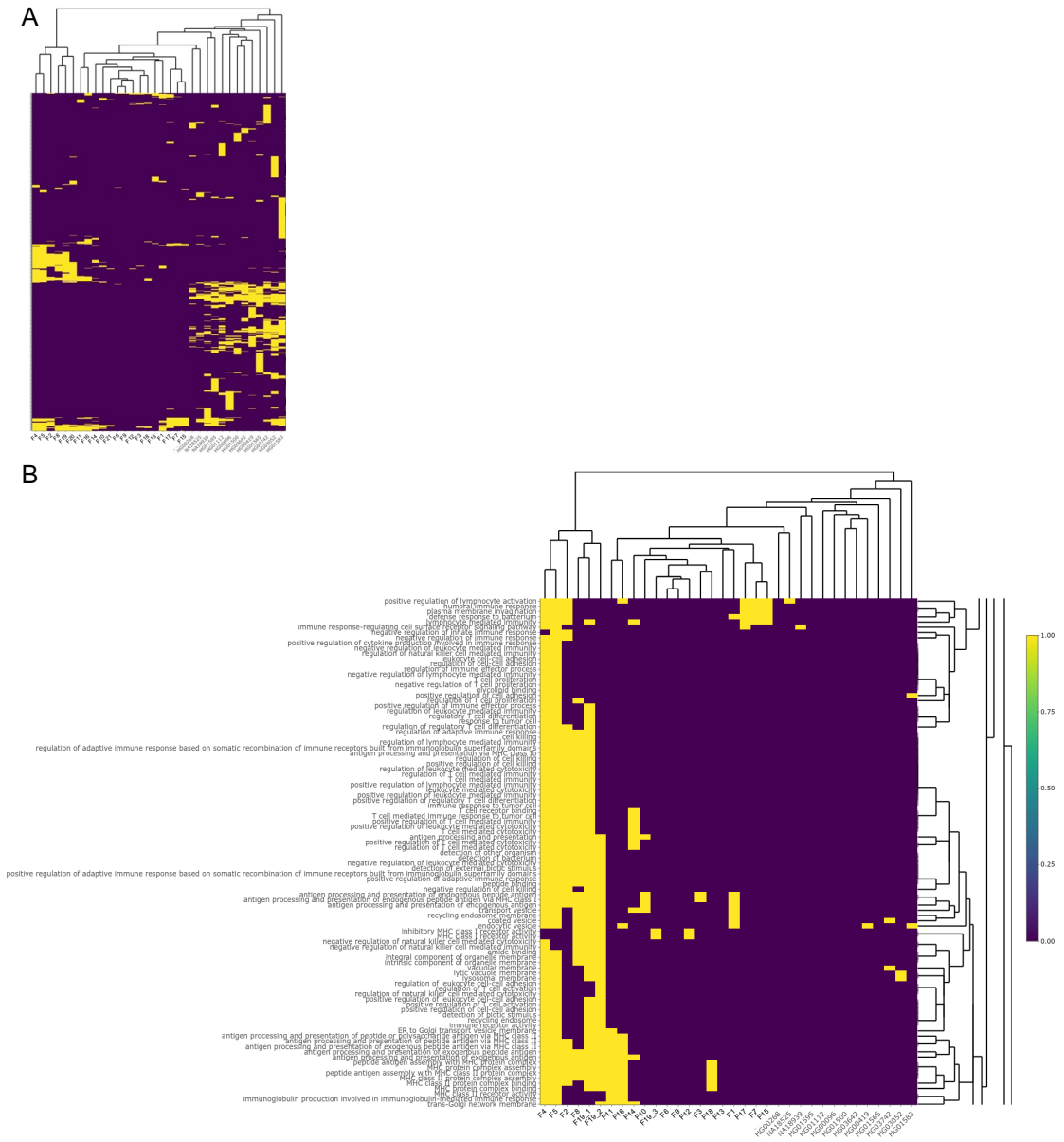


Figure 4. CNVs and SNVs downregulated pathways in HDGC-like patients. *A) Heatmap representing all impaired pathways in HDGC-like patients, B) Cluster of impaired pathways in HDGC-like patients and absent in control individuals, C) Heatmap representing immune-associated impaired pathways, from which genes were impacted in at least 3 patients.*

Regulatory germline landscape of HDGC patients

The germline landscape of HDGC patients is characterized by several impaired genes and pathways that may create the appropriate genetic background for DGC development, unrelated with an increased risk conferred by *CDH1* or *CTNNA1* loss of function variants. Thus, we hypothesized that deletions in regulatory elements controlling *CDH1* or *CTNNA1* expression increase risk for DGC development.

Herein, we focused on deletions prioritized by at least two callers, ultra-rare in the general population (AF<0.00001) and relatively common in the cohort (<6 patients, <24%) that overlapped accessible chromatin regions in normal stomach tissue. We found 6642 deletions, from which 415 were overlapping accessible chromatin regions. We divided the patients according to their DGC age of onset, with a cut-off at 50 years of age. DGC patients before 50 years of age display increased affected accessible chromatin regions upon germline deletions, in comparison with later on-set DGC patients (**Figure 5A**). In an attempt to characterize these regions, we mined the Ensembl regulatory build database and searched for putative regulatory elements. Despite no differences between the young-affected DGC, later on-set DGC and controls groups in Ensembl defined regulatory elements, young-affected DGC carry deletions in putative regulatory elements in immune-associated pathways (**Figure 5B,C,D**).

This data may pinpoint a regulatory impact upon deletion of gastric accessible chromatin in earlier on-set DGC development, driving an immune germline signature.

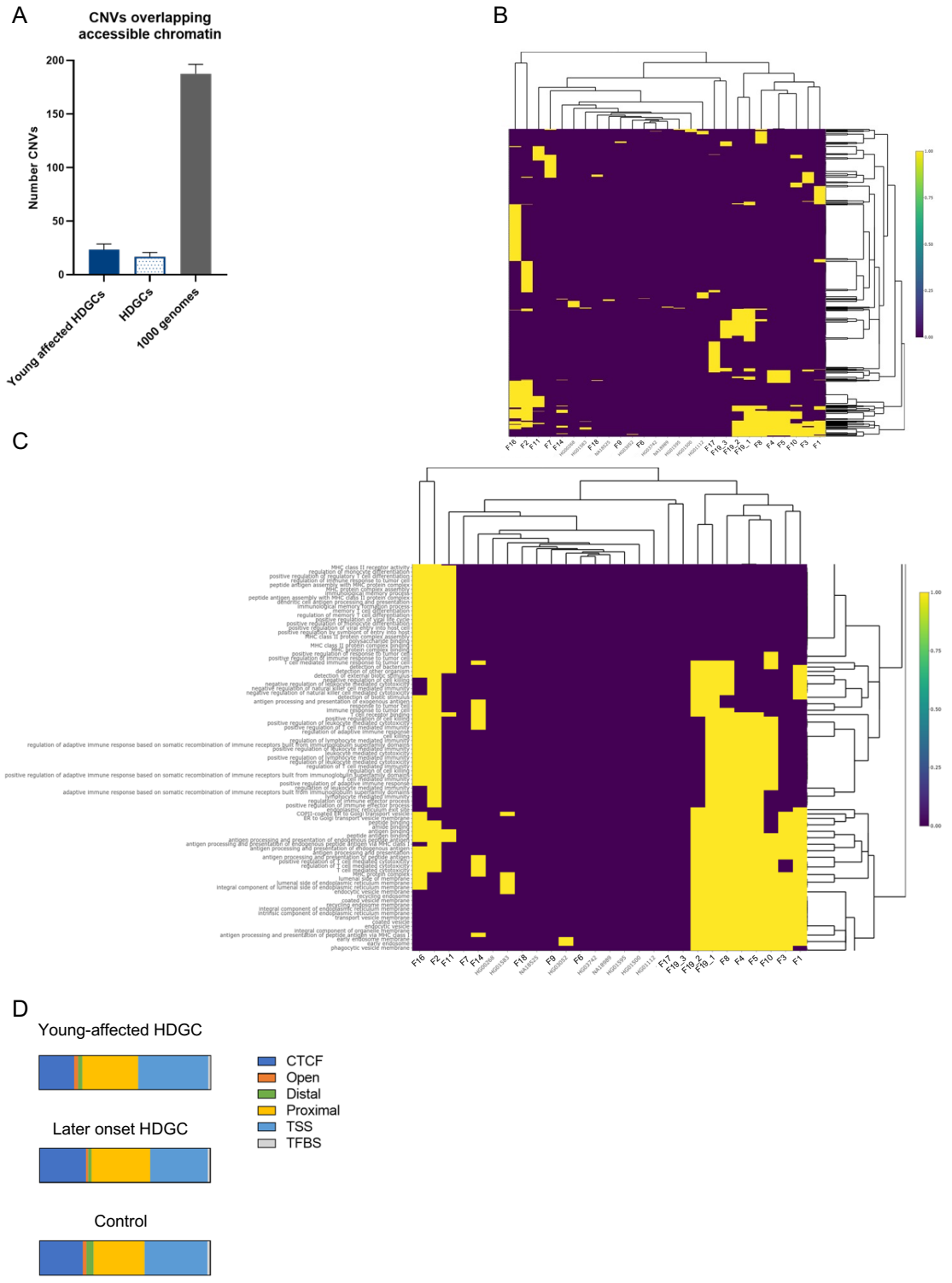


Figure 5. CNVs overlapping accessible chromatin regions. *A) CNVs overlapping accessible chromatin in normal stomach tissue of young-affected HDGC patients, older-affected HDGCs patients and controls from 1000genomes project, B) Impaired pathways of CNVs overlapping accessible chromatin regions, C) Cluster of CNVs overlapping accessible chromatin regions, D) Characterization of CNVs overlapping accessible chromatin regions in ensembl regulatory build.*

Discussion

Germline loss of function defects in *CDH1* and *CTNNA1* genes predispose to HDGC, and, therefore, DGC and LBC development(Guilford et al. 1998a; Majewski et al. 2013a). Nevertheless, nearly half of families meeting HDGC clinical criteria still lack a genetic diagnosis, denominated HDGC-like families (Blair et al. 2020a), despite most still display *CDH1* mono-allelic RNA downregulation (Pinheiro et al. 2010). Altogether, these results pinpoint *CDH1* as the main gene triggering HDGC. Working under the hypothesis that regulatory elements control *CDH1* expression, beyond their coding sequence, we aimed at solving the missing heritability in HDGC-like families.

In an attempt to solve the missing heritability in HDGC-like families, lacking *CDH1* and *CTNNA1* loss of function variants, we sequenced the gDNA of 21 HDGC-like families.

In a knowledge-based approach, we explored ultra-rare CNVs and SNVs in cancer syndromes-associated genes and found a *MLH1* germline deletion. *MLH1* predisposes to lynch syndrome, harbouring 13% risk of developing gastric cancer (Møller et al. 2018). This CNV encompasses enhancer/promoter chromatin marks in normal stomach tissue, absent in normal colon mucosa. This region may act as a tissue-specific putative regulatory element in stomach, beyond the impact in *MLH1* coding sequence, and may suggest a biological mechanism behind gastric cancer development alone. To our knowledge this is the first report of a *MLH1* exon 13 deletion in a gastric cancer biased family, that may be driven by stomach-specific regulatory elements.

Given the lack of variants in HDGC- or other cancer syndromes-associated genes in the remaining families, we dissected the *CDH1* regulatory network. We found a deletion of the upstream gene *CDH3* that encompass a *CDH1*-interactor with accessible chromatin and *CDH1*-eQTL, both impacting *CDH1* expression with tissue specific enhancer activity in forebrain (accessible chromatin region and *CDH1*-eQTL), palate (accessible chromatin region) and stomach (*CDH1*-eQTL). Remarkably, all cancer types in the family may be associated with either *CDH1* or *CDH3* expression loss, a result from regulatory element loss or protein loss itself, respectively.

Further, we found a small downstream deletion within *CDH1-TANGO6* intergenic region in a family with later on-set DGC, which deletion of the entire intergenic region results in *CDH1* downregulation. São José and colleagues showed an increased impact of the *CDH1-TANGO6* deletion in *CDH1* expression itself, immune, proliferation and adhesion-associated pathways compared with the *CDH1* deletion alone (**original paper 2**). Altogether we propose that regulatory elements in *CDH1-TANGO6* intergenic region control *CDH1* expression, driving later on-set DGC in the present family and increased age of onset and disease penetrance when combined with a *CDH1* deletion, as observed for the previously identified family (**original paper 2**).

The germline landscape of HDGC-like families encompasses downregulation of immune-associated pathways. A specific cluster was found in younger affected individuals with downregulation of pathways associated with MHC I and natural killer immunity, absent in healthy controls, which play a role in gastric cancer immune evasion (Shen et al. 2005; Na et al. 2021). Interestingly, earlier on-set DGC patients displays increased deletions in genome-wide open chromatin regions in normal stomach tissue. These regions encompass immune-affected pathways, recapitulating the germline landscape and pinpointing stomach regulatory elements as possible genetic drivers. Altogether, the observed immune impairment in HDGC families may reflect the lack of immunological surveillance of intramucosal signet-ring cell carcinoma *foci*.

The HDGC phenotype observed in families F15 and F9 is solved by germline *MLH1* and *CDH1-TANGO6* intergenic CNVs alone. Besides the germline CNV in the upstream *CDH3* gene, family F4 is solved by impaired immune-associated genes/pathways (**Figure 6**). Besides immune genes, mucins were also found impaired in HDGC-like patients. Indeed, *MUC3A*, *MUC4* and *MUC12* were overexpressed in several epithelial tumours with vessel invasion and lymph node metastasis (Tamura et al. 2012; Gao et al. 2020; Su et al. 2022). Their impairment, through truncating SNVs or deletions, correlates well with most HDGC tumours that mainly infiltrate in the peritoneum (Oliveira et al. 2015). Reduced *MUC4* expression was also found in poorly differentiated gastric cancers (Tamura et al. 2012), such as signet-ring cell carcinoma, characteristic of HDGC tumours (Carneiro et al. 1995; van der Post et al. 2016). As *MUC6* is expressed in normal stomach epithelia(Babu et al. 2006), its downregulation coordinates with advanced gastric cancer stage and poor prognosis (Shi and Xi 2021), and may recapitulate the observed phenotype in HDGC. Altogether, impairment of immune- and mucins-associated pathways/genes may compose a germline genetic background capable of regulating HDGC development in the presence of a triggering mechanism.

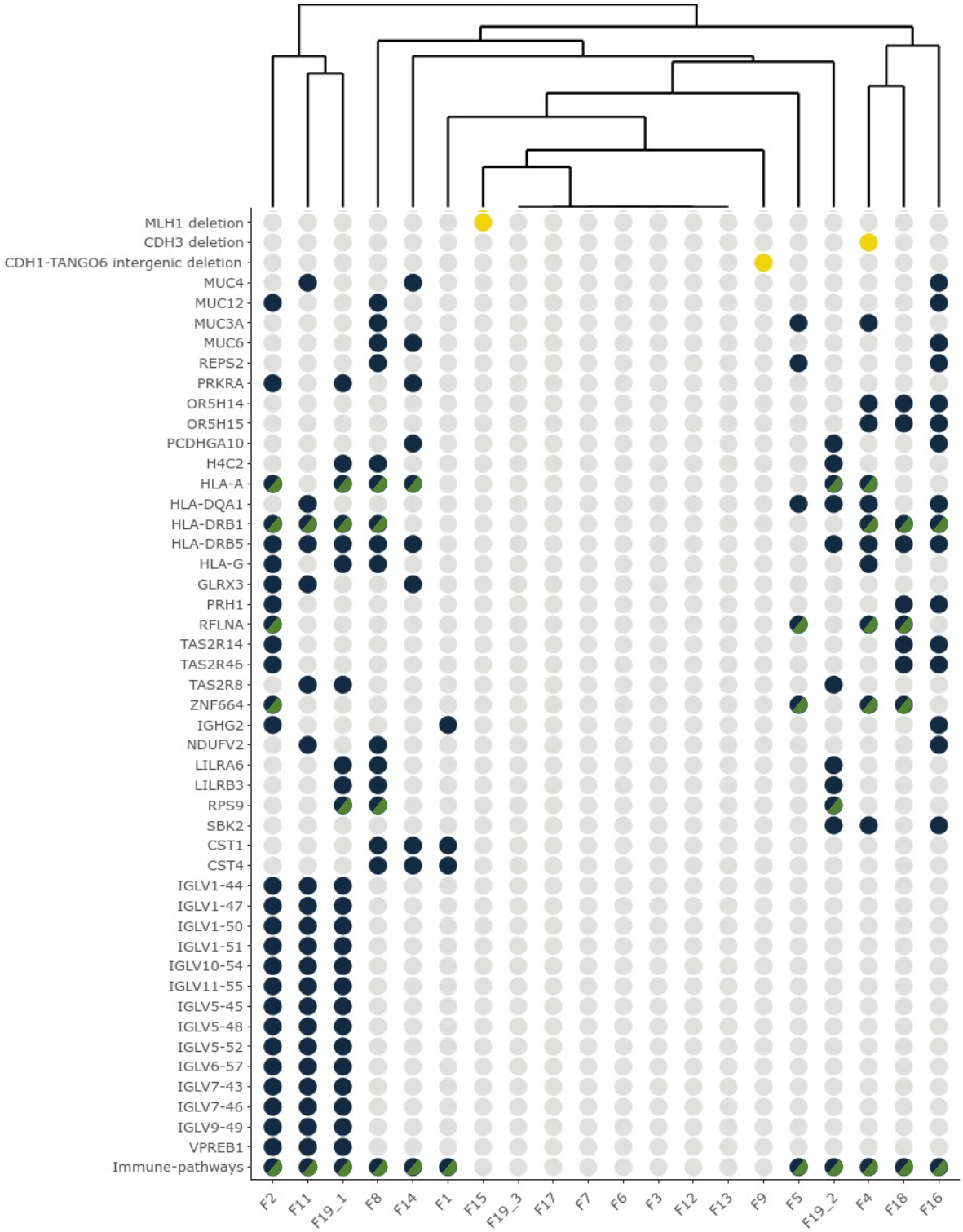


Figure 6. Main findings solving the missing heritability in HDGC patients. *Blue represents ultra-rare (AF<0.00001) coding CNVs and truncating SNVs found from WGS sequencing of HDGC patients. Yellow represent CNVs found overlapping CDH1-promoter interactors. Green represents pathways overlapping accessible chromatin in normal stomach.*

In conclusion, we found the missing heritability in 3 out of 21 HDGC-like families, corresponding to 14% of the entire cohort and pinpointed *CDH1* putative regulatory elements as regulators of *CDH1*/E-cadherin expression impacted in HDGC families. The germline landscape of younger DGC patients includes downregulation of immune-associated pathways, that may play a role in immune surveillance and DGC development when triggered by a still unknown mechanism.

Acknowledgments

This work was supported by: (1) Doctoral Programme in Biomedicine - Faculty of Medicine, University of Porto, to CSJ and JG-P; (2) Doctoral Programme in Computational Sciences, Faculty of Sciences, University of Porto, to MF.

Funding

This work received funding from: (1) the European Union's Horizon 2020 research and innovation programme under grant agreement No 779257 (Solve-RD: H2020-SC1-2017-Single-Stage-RTD), which provided Ph.D. salaries to JG-P; (2) The European Regional Development Fund (ERDF) through the COMPETE 2020-Operacional Programme for Competitiveness and Internationalisation (POCI), Portugal 2020, and by Portuguese funds through Fundação para a Ciência e Tecnologia (FCT)/Ministério da Ciência, Tecnologia e Inovação (3DChroMe: PTDC/BTM-TEC/30164/2017, LEGOH: PTDC/BTM-TEC/6706/2020); (3) FCT with salary support for PhD studentships to CSJ (Ref. SFRH/BD/140796/2018) and MF (Ref. 2020.05763.BD).

Conflict of interest: The authors disclosure no conflicts of interest.

References

Abascal F, Acosta R, Addleman NJ, et al (2020) Expanded encyclopaedias of DNA elements in the human and mouse genomes. Nature 583:699–710. https://doi.org/10.1038/S41586-020-2493-4

Artus J, Hadjantonakis A-K (2011) Generation of chimeras by aggregation of embryonic stem cells with diploid or tetraploid mouse embryos. Methods Mol Biol 693:37–56. https://doi.org/10.1007/978-1-60761-974-1_3

Babu SD, Jayanthi V, Devaraj N, et al (2006) Expression profile of mucins (MUC2, MUC5AC and MUC6) in Helicobacter pylori infected pre-neoplastic and neoplastic human gastric epithelium. Mol Cancer 5:. https://doi.org/10.1186/1476-4598-5-10

Banerji J, Rusconi S, Schaffner W (1981) Expression of a beta-globin gene is enhanced by remote SV40 DNA sequences. Cell 27:299–308. https://doi.org/10.1016/0092-8674(81)90413-X

Blair VR, McLeod M, Carneiro F, et al (2020) Hereditary diffuse gastric cancer: updated clinical practice guidelines. Lancet Oncol 21:e386–e397. https://doi.org/10.1016/S1470-2045(20)30219-9

Bonev B, Cavalli G (2016) Organization and function of the 3D genome. Nat Rev Genet 17:661–678. https://doi.org/10.1038/NRG.2016.112

Caldas C, Carneiro F, Lynch HT, et al (1999) Familial gastric cancer: overview and guidelines for management. J Med Genet 36:873–880

Cameron DL, Schröder J, Penington JS, et al (2017) GRIDSS: sensitive and specific genomic rearrangement detection using positional de Bruijn graph assembly. Genome Res 27:2050–2060. https://doi.org/10.1101/GR.222109.117

Carneiro F, Seixas M, Sobrinho-Simões M (1995) New elements for an updated classification of the carcinomas of the stomach. Pathol Res Pract 191:571–584. https://doi.org/10.1016/S0344-0338(11)80878-2

Cruz-Molina S, Respuela P, Tebartz C, et al (2017) PRC2 Facilitates the Regulatory Topology Required for Poised Enhancer Function during Pluripotent Stem Cell Differentiation. Cell Stem Cell 20:689-705.e9. https://doi.org/10.1016/J.STEM.2017.02.004

Daniel CW, Strickland P, Friedmann Y (1995) Expression and functional role of E- and P-cadherins in mouse mammary ductal morphogenesis and growth. Dev Biol 169:511–519. https://doi.org/10.1006/DBIO.1995.1165

Dekker J, Rippe K, Dekker M, Kleckner N (2002) Capturing chromosome conformation. Science 295:1306–1311. https://doi.org/10.1126/SCIENCE.1067799

Fishilevich S, Nudel R, Rappaport N, et al (2017) GeneHancer: genome-wide integration of enhancers and target genes in GeneCards. Database (Oxford) 2017:1–17. https://doi.org/10.1093/database/bax028

Gao SL, Yin R, Zhang LF, et al (2020) The oncogenic role of MUC12 in RCC progression depends on c-Jun/TGF-β signalling. J Cell Mol Med 24:8789–8802. https://doi.org/10.1111/JCMM.15515

Geeven G, Teunissen H, de Laat W, de Wit E (2018) peakC: a flexible, non-parametric peak calling package for 4C and Capture-C data. Nucleic Acids Res 46:E91. https://doi.org/10.1093/NAR/GKY443

Geoffroy V, Herenger Y, Kress A, et al (2018) AnnotSV: an integrated tool for structural variations annotation. Bioinformatics 34:3572–3574. https://doi.org/10.1093/BIOINFORMATICS/BTY304

Guilford P, Hopkins J, Harraway J, et al (1998a) E-cadherin germline mutations in familial gastric cancer. Nature 392:402–405. https://doi.org/10.1038/32918

Guilford P, Hopkins J, Harraway J, et al (1998b) E-cadherin germline mutations in familial gastric cancer. Nature 392:402–405. https://doi.org/10.1038/32918

Gullo I, Grillo F, Mastracci L, et al (2020) Precancerous lesions of the stomach, gastric cancer and hereditary gastric cancer syndromes. Pathologica 112:166–185. https://doi.org/10.32074/1591-951X-166

Krijger PHL, Geeven G, Bianchi V, et al (2020) 4C-seq from beginning to end: A detailed protocol for sample preparation and data analysis. Methods 170:17–32. https://doi.org/10.1016/J.YMETH.2019.07.014

Layer RM, Chiang C, Quinlan AR, Hall IM (2014) LUMPY: a probabilistic framework for structural variant discovery. Genome Biol 15:. https://doi.org/10.1186/GB-2014-15-6-R84

Lee HW, Min SK, Ju YS, et al (2011) Prognostic significance of HLA class I expressing in gastric carcinoma defined by monoclonal anti-pan HLA class I antibody, EMR8-5. J Gastrointest Surg 15:1336–1343. https://doi.org/10.1007/S11605-011-1545-3

Lettice LA, Heaney SJH, Purdie LA, et al (2003) A long-range Shh enhancer regulates expression in the developing limb and fin and is associated with preaxial polydactyly. Hum Mol Genet 12:1725–1735. https://doi.org/10.1093/HMG/DDG180

Llorens-Bobadilla E, Zamboni M, Marklund M, et al (2023) Solid-phase capture and profiling of open chromatin by spatial ATAC. Nat Biotechnol. https://doi.org/10.1038/S41587-022-01603-9

Lobe CG, Koop KE, Kreppner W, et al (1999) Z/AP, a double reporter for cre-mediated recombination. Dev Biol 208:281–292. https://doi.org/10.1006/dbio.1999.9209

Majewski IJ, Kluijt I, Cats A, et al (2013a) An α-E-catenin (*CTNNA1*) mutation in hereditary diffuse gastric cancer. J Pathol 229:621–629. https://doi.org/10.1002/path.4152

Majewski IJ, Kluijt I, Cats A, et al (2013b) An alpha-E-catenin (*CTNNA1*) mutation in hereditary diffuse gastric cancer. J Pathol 229:621–629. https://doi.org/10.1002/path.4152

Møller P, Seppälä TT, Bernstein I, et al (2018) Cancer risk and survival in path_MMR carriers by gene and gender up to 75 years of age: a report from the Prospective Lynch Syndrome Database. Gut 67:1306–1316. https://doi.org/10.1136/GUTJNL-2017-314057

Na HY, Park Y, Nam SK, et al (2021) Prognostic significance of natural killer cell-associated markers in gastric cancer: quantitative analysis using multiplex immunohistochemistry. J Transl Med 19:. https://doi.org/10.1186/S12967-021-03203-8

Oliveira C, Pinheiro H, Figueiredo J, et al (2015) Familial gastric cancer: genetic susceptibility, pathology, and implications for management. Lancet Oncol 16:e60-70. https://doi.org/10.1016/S1470-2045(14)71016-2

Pinheiro H, Bordeira-Carrico R, Seixas S, et al (2010) Allele-specific *CDH1* downregulation and hereditary diffuse gastric cancer. Hum Mol Genet 19:943–952. https://doi.org/10.1093/hmg/ddp537

Poplin R, Chang PC, Alexander D, et al (2018) A universal SNP and small-indel variant caller using deep neural networks. Nat Biotechnol 36:983. https://doi.org/10.1038/NBT.4235

Rausch T, Zichner T, Schlattl A, et al (2012) DELLY: structural variant discovery by integrated paired-end and split-read analysis. Bioinformatics 28:. https://doi.org/10.1093/BIOINFORMATICS/BTS378

Shen YQ, Zhang JQ, Miao FQ, et al (2005) Relationship between the downregulation of HLA class I antigen and clinicopathological significance in gastric cancer. World J Gastroenterol 11:3628–3631. https://doi.org/10.3748/WJG.V11.I23.3628

Shi D, Xi XX (2021) Regulation of MUC6 Methylation Correlates with Progression of Gastric Cancer. Yonsei Med J 62:1005–1015. https://doi.org/10.3349/YMJ.2021.62.11.1005

Spitz F, Furlong EEM (2012) Transcription factors: from enhancer binding to developmental control. Nat Rev Genet 13:613–626. https://doi.org/10.1038/NRG3207

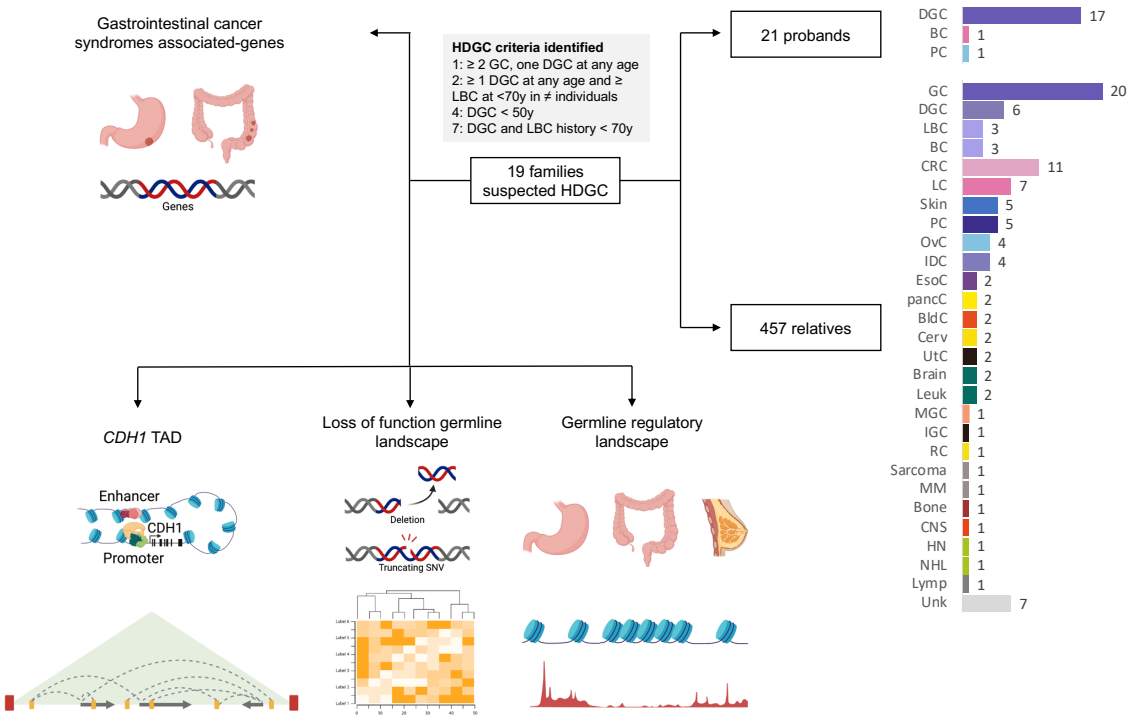
Su W, Feng B, Hu L, et al (2022) MUC3A promotes the progression of colorectal cancer through the PI3K/Akt/mTOR pathway. BMC Cancer 22:. https://doi.org/10.1186/S12885-022-09709-8

Tamura Y, Higashi M, Kitamoto S, et al (2012) MUC4 and MUC1 expression in adenocarcinoma of the stomach correlates with vessel invasion and lymph node metastasis: an immunohistochemical study of early gastric cancer. PLoS One 7:. https://doi.org/10.1371/JOURNAL.PONE.0049251

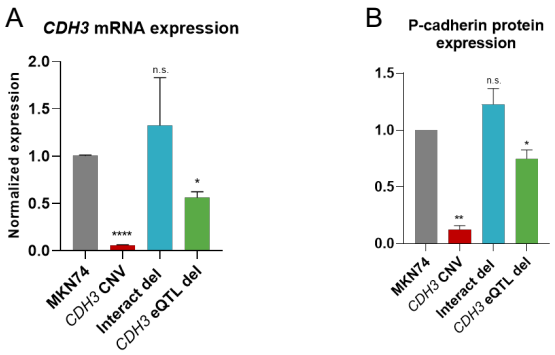
Van De Werken HJG, De Vree PJP, Splinter E, et al (2012) 4C technology: protocols and data analysis. Methods Enzymol 513:89–112. https://doi.org/10.1016/B978-0-12-391938-0.00004-5

van der Post RS, Gullo I, Oliveira C, et al (2016) Histopathological, Molecular, and Genetic Profile of Hereditary Diffuse Gastric Cancer: Current Knowledge and Challenges for the Future. Adv Exp Med Biol 908:371–391. https://doi.org/10.1007/978-3-319-41388-4_18

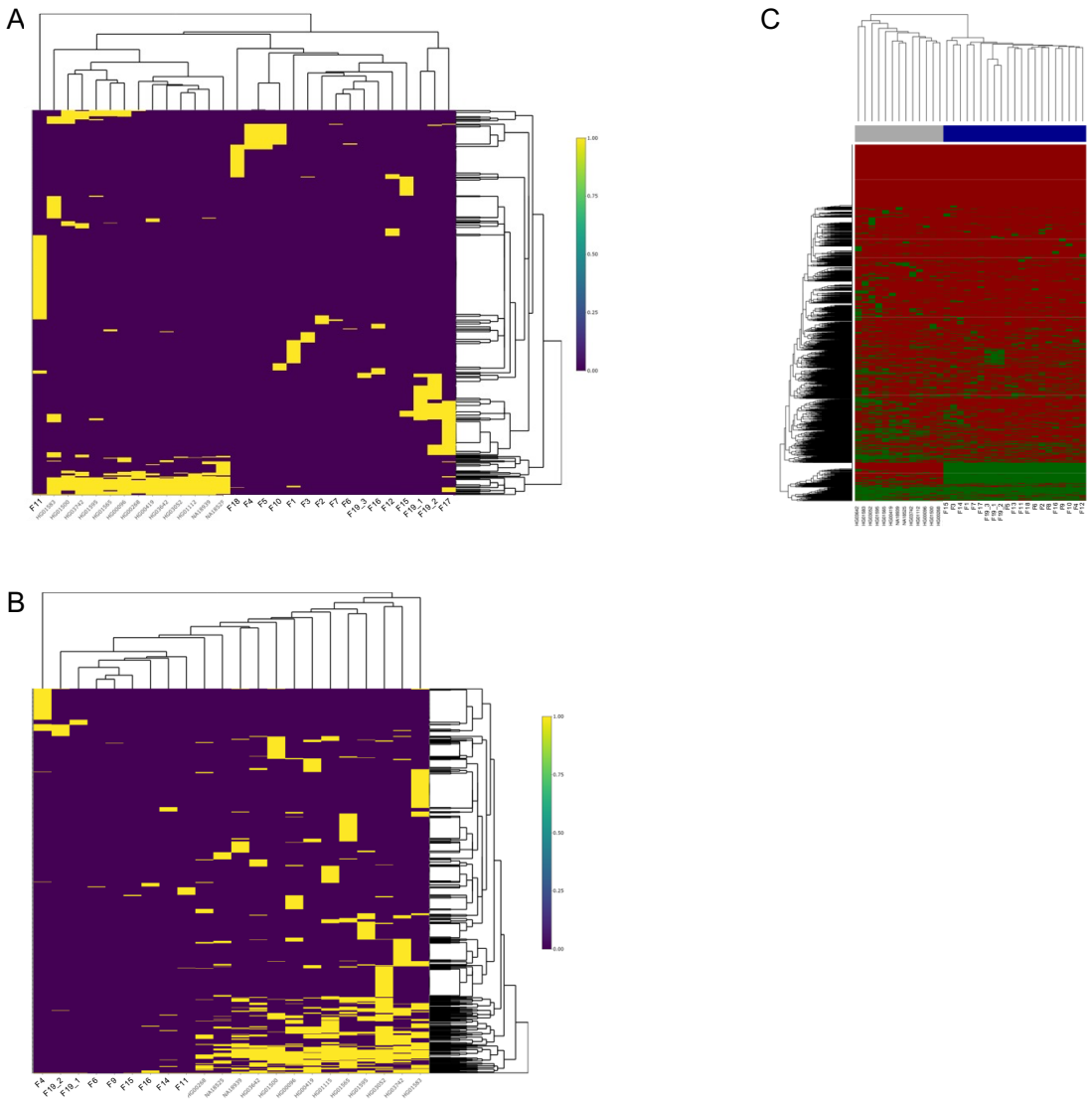
Will AJ, Cova G, Osterwalder M, et al (2017) Composition and dosage of a multipartite enhancer cluster control developmental expression of Ihh (Indian hedgehog). Nat Genet 49:1539–1545. https://doi.org/10.1038/NG.3939



Supplementary Figure 1. Strategy overview and cohort description. *DGC: diffuse gastric cancer, BC: breast cancer, PC: prostate cancer, GC: gastric cancer, LBC: lobular breast cancer, CRC: colorectal cancer, LC: lung cancer, OvC: ovarian cancer, IDC: invasive ductal breast cancer, EsoC: esophageal cancer, pancC: pancreatic cancer, BldC: bladder cancer, Cerv: cervical cancer, UtC: uterine cancer, Leuk: leukemia, MGC: mixed-type gastric cancer, IGC: intestinal gastric cancer, RC: renal cancer, MM: malignant melanoma, CNS: central nervous system cancer, HN: NHL: non-Hodgkins lymphoma, Lymp: lymphoma, Unk: unknown.*



Supplementary Figure 2. Characterization of CDH3/P-cadherin expression. *A) CDH3 mRNA expression in edited clones, B) P-cadherin protein expression in edited clones.*



Supplementary Figure 3. Impaired pathways in HDGC-like patients. *A) Ultra-rare coding deletions present in 5 or less patients and absent in controls, B) Ultra-rare truncating SNVs present in 5 or less patients and absent in controls, C) Frequent CNVs and SNVs between HDGCs and controls.*

Supplementary table 1. sgRNAs and primers used.

Primer	Sequence	Orientation	Used for
CDH3_CNV_5_F1	CACCGCTGATTGGTCCAGCGCCACG	-	sgRNA
CDH3_CNV_5_R1	aaacCGTGGCGCTGGACCAATCAGC	-	sgRNA
CDH3_CNV_3_F1	CACCGCACTCTGAGAGTTTGGAACG	-	sgRNA
CDH3_CNV_3_R1	aaacCGTTCCAAACTCTCAGAGTGC	-	sgRNA
CDH3_CNV_F	ACTGAGGAGGCTGAGGAGTG	Forward	primer
CDH3_CNV_R	TCTCCAGGTACTCCACAGCC	Reverse	primer
CDH3_eQTL_5_F1	CACCGCCAAACGggctcggtgtga	-	sgRNA
CDH3_eQTL_5_R1	aaactcacaccgagccCGTTTGGC	-	sgRNA
CDH3_eQTL_3_F1	CACCGtcccatctactctactcagg	-	sgRNA
CDH3_eQTL_3_R1	aaaccctgagtagagtagatgggaC	-	sgRNA
CDH3_eQTL_F	gagagccagtgaggcatagg	Forward	primer
CDH3_eQTL_R	ACCACCCTGTCCTAGGCTTT	Reverse	primer
CDH3_ATAC_5'_F1	CACCGtgagaagcaaattctagtgg	-	sgRNA
CDH3_ATAC_5'_R1	aaacccactagaatttgcttctcaC	-	sgRNA
CDH3_ATAC_3'_F1	CACCGaatgacgtgaacctagaagg	-	sgRNA
CDH3_ATAC_3'_R1	aaacccttctaggttcacgtcattC	-	sgRNA
CDH3_ATAC_F	AGTGGCAGGAGGAATAGCAG	Forward	primer
CDH3_ATAC_R	ACCGGGAATGTGTAGACCCCCA	Reverse	primer
CDH1_3UTR_F1	CACCGAGATGGCCTCAGTCAAAACG	-	sgRNA
CDH1_3UTR_R1	aaacCGTTTTGACTGAGGCCATCTC	-	sgRNA
Tango6_5UTR_F1	CACCgcttctgaaattgtgcgcgt	-	sgRNA
Tango6_5UTR_R1	aaacacgcgcacaatttcagaagc	-	sgRNA
CDH1-TANGO_F	ACCCTGCAATCACTTTTTGG	Forward	primer
CDH1-TANGO_R	ccatatcggacaggttggtc	Reverse	primer

Supplementary table 2. primers used for mouse enhancer assay.

Primer	Sequence
Interg_enh_Gib_F2	gtatacattatacgaagttatTTAATTGGAAGGAATGGAGGAGTCTCAACAT
Interg_enh_Gib_R2	GCGGCTTCGAGTTAATAATCATGTTAAGTGTGGCAGAAGGCTCT
eQTL_enh_Gib_F2	gtatacattatacgaagttatTTAATGTGTCTGGGTTGAATCTGTCCACTT
eQTL_enh_Gib_R2	GCGGCTTCGAGTTAATccccagaaccttacagaccattacc
ATAC_enh_Gib_F2	GATCTTAATTAACAGCAACCCTTCTCAACTCC
ATAC_enh_Gib_R2	gtatacattatacgaagttatTTAATTGGGGAAAGAAAGCCAAACG

Supplementary table 3. primers used for 4C-seq.

Primer	Sequence
4C_prom_P5_N502	AATGATACGGCGACCACCGAGATCTACACCTCTCTATTCGTCGGCAGCGTCAGATGTGTATAAGAGACAGGTGAACCCTCAGCCAATCAG
4C_prom_P5_N504	AATGATACGGCGACCACCGAGATCTACACAGAGTAGATCGTCGGCAGCGTCAGATGTGTATAAGAGACAGGTGAACCCTCAGCCAATCAG
4C_prom_P5_N503	AATGATACGGCGACCACCGAGATCTACACTATCCTCTTCGTCGGCAGCGTCAGATGTGTATAAGAGACAGGTGAACCCTCAGCCAATCAG
4C_prom_P5_N501	AATGATACGGCGACCACCGAGATCTACACTAGATCGCTCGTCGGCAGCGTCAGATGTGTATAAGAGACAGGTGAACCCTCAGCCAATCAG
4C_prom_P7_N705	CAAGCAGAAGACGGCATACGAGATAGGAGTCCGTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGGATCCCAGGTCTTAGTGAGCCA
4C_prom_P7_N707	CAAGCAGAAGACGGCATACGAGATGTAGAGAGGTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGGATCCCAGGTCTTAGTGAGCCA
4C_prom_P7_N706	CAAGCAGAAGACGGCATACGAGATCATGCCTAGTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGGATCCCAGGTCTTAGTGAGCCA
4C_prom_P7_N708	CAAGCAGAAGACGGCATACGAGATCCTCTCTGGTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGGATCCCAGGTCTTAGTGAGCCA

Supplementary table 4. Genes associated with gastrointestinal syndromes

Gene name	Reference sequence
A2ML1	NM_144670.6
ACD	NM_001082486.2
ADM	NM_001124.3
AIP	NM_003977.4
AKT1	NM_001014431.2
ALK	NM_004304.5
ANKRD26	NM_014915.2
APC	NM_001354896.2
ARMC5	NM_001288767.2
ASXL1	NM_015338.6
ATM	NM_000051.3
ATR	NM_001184.4
AXIN1	NM_003502.4
AXIN2	NM_004655.4
BAP1	NM_004656.4
BARD1	NM_000465.4
BLM	NM_000057.4
BMPR1A	NM_004329.2
BRAF	NM_001354609.2
BRCA1	NM_007300.4
BRCA2	NM_000059.3
BRIP1	NM_032043.3
BUB1	NM_004336.5
BUB1B	NM_001211.5
BUB3	NM_004725.4
CBL	NM_005188.4
CDC73	NM_024529.4
CDH1	NM_004360.5
CDK4	NM_000075.4
CDKN1A	NM_001291549.2
CDKN1B	NM_004064.4
CDKN1C	NM_000076.2
CDKN2A	NM_001195132.1
CDKN2B	NM_004936.4
CDKN2C	NM_001262.2
CEBPA	NM_001287424.2
CHEK1	NM_001330427.1
CHEK2	NM_001005735.2
CREBBP	NM_004380.3
CTC1	NM_025099.6
CTNNA1	NM_001323982.2
CTR9	NM_014633.5
CYLD	NM_015247.2
DDB2	NM_000107.2
DDX11	NM_001257144.2
DICER1	NM_001271282.3
DIS3L2	NM_152383.4
DKC1	NM_001363.5
DNAJC21	NM_194283.4
EGFR	NM_005228.5
ELANE	NM_001972.4
ENG	NM_001114753.2
EPCAM	NM_002354.3

Gene name	Reference sequence
ERCC1	NM_001369408.1
ERCC2	NM_000400.3
ERCC3	NM_000122.2
ERCC4	NM_005236.3
ERCC5	NM_000123.3
ERCC6	NM_000124.4
ESR2	NM_001437.2
ETV6	NM_001987.5
EXO1	NM_006027.4
EXT1	NM_000127.2
EXT2	NM_000401.3
EZH2	NM_004456.5
FAN1	NM_014967.5
FANCA	NM_000135.4
FANCB	NM_001018113.3
FANCC	NM_000136.3
FANCD2	NM_033084.5
FANCE	NM_021922.3
FANCF	NM_022725.4
FANCG	NM_004629.1
FANCI	NM_001113378.1
FANCL	NM_001114636.1
FANCM	NM_020937.4
FAS	NM_000043.6
FBXW7	NM_001349798.2
FH	NM_000143.3
FLCN	NM_001353229.2
FOXO3	NM_001455.4
G6PC3	NM_138387.3
GALNT12	NM_024642.5
GATA2	NM_001145661.2
GDNF	NM_001190468.1
GFI1	NM_001127215.2
GPC3	NM_001164617.2
GREM1	NM_001368719.1
GRHL2	NM_024915.4
HABP2	NM_004132.5
HAX1	NM_006118.4
HNF1A	NM_001306179.2
HOXB13	NM_006361.5
IPMK	NM_152230.5
KIF1B	NM_001365951.2
KIT	NM_000222.2
KLLN	NM_001126049.1
KRAS	NM_001369786.1
LIG4	NM_001352604.1
LZTR1	NM_006767.4
MAP2K1	NM_002755.3
MAP2K2	NM_030662.3
MAX	NM_002382.5
MDH2	NM_005918.4
MEN1	NM_001370251.1
MET	NM_001127500.3
MITF	NM_001354604.2
MLH1	NM_000249.3
MLH3	NM_001040108.1
MPL	NM_005373.3

Supplementary table 4. Genes associated with gastrointestinal syndromes

Gene name	Reference sequence
MRE11A	NM_005591.3
MSH2	NM_000251.3
MSH3	NM_002439.5
MSH6	NM_000179.2
MUC5B	NM_002458.3
MUTYH	NM_001128425.1
NBN	NM_002485.4
NF1	NM_001042492.3
NF2	NM_000268.3
NHP2	NM_017838.3
NOP10	NM_018648.3
NPM1	NM_001355006.1
NRAS	NM_002524.5
NSD1	NM_022455.4
NTHL1	NM_002528.7
OGG1	NM_016821.2
PALB2	NM_024675.4
PARK2	NM_004562.3
PARN	NM_002582.4
PAX5	NM_016734.3
PDGFB	NM_002608.4
PDGFRA	NM_001347828.2
PHOX2B	NM_003924.4
PIK3CA	NM_006218.4
PMS1	NM_000534.4
PMS2	NM_001322014.2
PMS2CL	NR_002217.1
POLD1	NM_001308632.1
POLD2	NM_006230.3
POLD3	NM_006591.3
POLD4	NM_021173.5
POLE	NM_006231.4
POLE2	NM_002692.4
POLE3	NM_001278255.1
POLE4	NM_019896.4
POLH	NM_006502.3
POT1	NM_015450.3
POU6F2	NM_001370959.1
PPM1D	NM_003620.4
PRF1	NM_001083116.3
PRKAR1A	NM_001276289.1
PRSS1	NM_002769.5
PTCH1	NM_000264.5
PTCH2	NM_003738.5
PTEN	NM_001304717.5
PTPN11	NM_001330437.1
PTPN12	NM_002835.4
PTPRJ	NM_002843.4
RAD50	NM_005732.4
RAD51C	NM_058216.3
RAD51D	NM_001142571.2
RAF1	NM_001354689.3
RB1	NM_000321.2
RECQL	NM_002907.4
RECQL4	NM_004260.3
REST	NM_001193508.1
RET	NM_020975.6
RHBDF2	NM_024599.5

Gene name	Reference sequence
RINT1	NM_021930.6
RIT1	NM_001256821.2
RNF43	NM_001305544.2
RPL11	NM_000975.5
RPL35A	NM_000996.4
RPL5	NM_000969.5
RPS10	NM_001014.5
RPS17	NM_001021.6
RPS19	NM_001321485.2.2
RPS20	NM_001146227.2
RPS24	NM_001142285.2
RPS26	NM_001029.5
RPS7	NM_001011.4
RTEL1	NM_001283009.2
RUNX1	NM_001754.4
SAMD9	NM_001193307.1
SAMD9L	NM_001303496.3
SBDS	NM_016038.4
SDHA	NM_004168.4
SDHAF2	NM_017841.2
SDHB	NM_003000.3
SDHC	NM_003001.3
SDHD	NM_003002.4
SEC23B	NM_001172745.3
SEMA4A	NM_001193300.2
SFTPA1	NM_001093770.3
SFTPA2	NM_001320814.1
SH2B3	NM_005475.3
SHOC2	NM_001324336.1
SLC26A3	NM_000111.3
SLX4	NM_032444.4
SMAD4	NM_005359.6
SMAD9	NM_001127217.2
SMARCA4	NM_001128849.2
SMARCB1	NM_001362877.2
SMARCE1	NM_003079.5
SOS1	NM_005633.3
SPINK1	NM_001354966.1
SPRED1	NM_152594.3
SQSTM1	NM_003900.5
SRC	NM_005417.4
STK11	NM_000455.5
SUFU	NM_016169.3
TERC	NR_001566.1
TERF2IP	NM_018975.4
TERT	NM_198253.3
TINF2	NM_001099274.3
TMEM127	NM_001193304.3
TNFRSF11A	NM_003839.4
TP53	NM_000546.5
TRIP13	NM_004237.4
TSC1	NM_000368.5
TSC2	NM_000548.5
USB1	NM_024598.4
VHL	NM_000551.3
WAS	NM_000377.3
WRAP53	NM_001143990.1
WRN	NM_000553.6
WT1	NM_024426.6
XPA	NM_000380.3
XPC	NM_004628.4

Original paper 2

Combined loss of *CDH1* and downstream regulatory sequences drive early-onset diffuse gastric cancer and increases penetrance of Hereditary Diffuse Gastric Cancer

Celina São José[1,2,3]†, José Garcia-Pelaez[1,2,3]†, Marta Ferreira[1,2,4], Oscar Arrieta[5], Ana André[1,2], Nelson Martins[1,2,6], Samantha Solís[7], Braulio Martínez-Benítez[8], Mª Luisa Ordóñez-Sánchez[7], Maribel Rodríguez-Torres[7], Anna K. Sommer[9], Iris B.A.W. te Paske[10], Carlos Caldas[11,12,13], Marc Tischkowitz[14], Maria Teresa Tusié[7], Solve-RD DITF-GENTURIS, Nicoline Hoogerbrugge[10], German Demidov[15], Richarda M. de Voer[10], Steve Laurie[16], Carla Oliveira[1,2,17]

† These authors contributed equally to this work

[1] Instituto de Investigação e Inovação em Saúde, Porto, Portugal

[2] Instituto de Patologia e Imunologia Molecular da Universidade do Porto, Porto, Portugal

[3] Doctoral Programme in Biomedicine, Faculty of medicine, University of Porto, Porto, Portugal

[4] Department Computer Science Faculty of Science, University of Porto, Porto, Portugal;

[5] Thoracic Oncology Unit, Department of Thoracic Oncology, Instituto Nacional de Cancerología, Mexico City, Mexico

[6] Master Programme in Molecular Medicine and Oncology, Faculty of Medicine, University of Porto, Porto, Portugal

[7] Unidad de Biología Molecular y Medicina Genómica Instituto Nacional de Ciencias Médicas y Nutrición Salvador Zubirán, INCMNSZ/Instituto de Investigaciones Biomédicas, UNAM Mexico City, Mexico

[8] Pathology Department Instituto Nacional de Ciencias Médicas y Nutrición Salvador Zubirán, INCMNSZ Mexico City, Mexico

[9] Institute of Human Genetics, Medical Faculty, University of Bonn, Bonn, Germany

[10] Research Institute for Medical Innovation, Department of Human Genetics, Radboud university medical center, Nijmegen, the Netherlands

[11] Cancer Research UK Cambridge Institute, University of Cambridge, Li Ka Shing Centre, Cambridge, UK

[12] Department of Oncology, University of Cambridge, Cambridge, UK

[13] CRUK Cambridge Centre, Cambridge Experimental Cancer Medicine Centre (ECMC) and NIHR Cambridge Biomedical Research Centre, University of Cambridge and Cambridge University Hospitals NHS Foundation Trust, Cambridge, UK

[14] Department of Medical Genetics, National Institute for Health Research Cambridge Biomedical Research Centre, University of Cambridge, Cambridge, UK

[15] Institute of Medical Genetics and Applied Genomics, Tübingen, Germany

[16] The Barcelona Institute of Science and Technology, CNAG-CRG, Centre for Genomic Regulation (CRG), Barcelona, Spain

[17] Faculty of Medicine of the University of Porto, Porto, Portugal

* Correspondence: carlaol@ipatimup.pt; +351-225570785

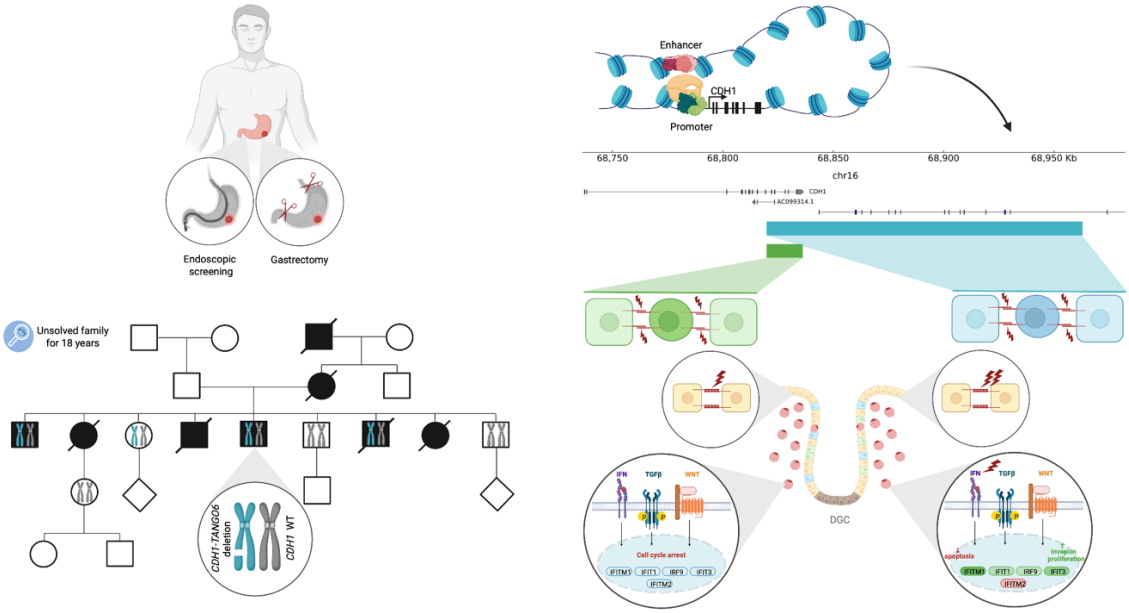
Gastric cancer (appendix B).

Author statement

Celina São José: Investigation, experimental work, methodology, conceptualization, formal analysis, writing – original draft, writing – review and editing.

Disclose the missing heritability in HDGC patients, beyond *CDH1* coding variants

Genetic diagnosis remains challenging for most of the HDGC-suspected families considered genetically undiagnosed and frequently lost for follow-up. Germline WES data re-analysis, encompassing copy number variants (CNVs) screening by multiple variant callers was performed to identify the main genetic defect in this family. Herein, we describe the diagnostic odyssey of an HDGC family identified in 2003 and genetically unresolved for almost twenty years after several attempts for diagnosis. CNVs screening revealed a germline heterozygous deletion encompassing the last fraction of the *CDH1* gene, the downstream intergenic region, and the beginning of the neighbouring gene (*TANGO6*) in four individuals. Three of these, as well as another five additional cases in the family, presented very early onset DGC. This high penetrance, aggressiveness and early development of the disease may be influenced by the deletion of the *CDH1* downstream region, altering *CDH1* regulatory elements. CRISPR-cas9 clones mimicking a homozygous deletion of only the *CDH1* fraction and the family CNV, revealed an almost two-fold *CDH1* mRNA expression downregulation induced by the *CDH1-TANGO6* deletion compared to the *CDH1* deletion alone. Gene ontology analysis highlighted an increased impact on adhesion-associated pathways of the *CDH1-TANGO6* deletion, compared with the *CDH1* deletion alone. Moreover, a specific misregulation towards the *CDH1-TANGO6* deletion in immune-associated pathways, response to virus and interferon-alpha/-beta was shown. This clinical report highlights the importance of retesting families with suspected tumour-risk syndromes due to the improvement of technical approaches and knowledge on these genetic disorders to confer proper management and care. Moreover, it demonstrates the impact of deleting regulatory regions to increase the *CDH1* penetrance in HDGC context.



Abstract

Germline *CDH1* pathogenic or likely pathogenic variants cause hereditary diffuse gastric cancer (HDGC). Once a genetic cause is identified, stomachs' and breasts' surveillance and/or prophylactic surgery is offered to asymptomatic *CDH1*-carriers, which is life-saving. Herein, we characterized a inherited mechanism responsible for extremely early-onset gastric cancer and atypical HDGC high-penetrance. Whole-exome sequencing (WES) re-analysis was performed in an usolved HDGC family. Accessible chromatin and *CDH1*-promoter interactors were evaluated in normal stomach by ATAC-seq and 4C-seq, and functional analysis was performed using CRISPR-Cas9, RNA-seq and pathway analysis.

We identified a germline heterozygous 23Kb *CDH1-TANGO6* deletion in a family with eight diffuse gastric cancers, six before age 30. Atypical HDGC high-penetrance and young cancer-onset argued towards a role for the deleted region downstream of *CDH1*, which we proved to present accessible chromatin, and *CDH1*-promoter interactors in normal stomach. CRISPR-Cas9 edited cells mimicking the *CDH1-TANGO6* deletion display strongest *CDH1* mRNA downregulation, more impacted adhesion-associated, type-I interferon immune-associated and oncogenic signalling pathways, compared to wild-type or *CDH1*-deleted cells. This finding solved a 18-years family odyssey and engaged carrier family members in a cancer prevention pathway of care.

In this work, we demonstrated that regulatory elements lying dowstream of *CDH1* are part of a chromatin network that control *CDH1* expression and influence cell transcriptome and associated signalling pathways, likely explaining high disease-penetrance and very young cancer-onset. This study highlights the importance of incorporating scientific-technological updates and clinical guidelines in routine diagnosis, given their impact in timely genetic diagnosis and disease prevention.

Introduction

Pathogenic and likely pathogenic germline truncating single-nucleotide (SNVs) and copy-number variants (CNVs) affecting the *CDH1* coding sequence [1-3], as well as, truncating *CTNNA1* SNVs [4, 5] cause hereditary diffuse gastric cancer (HDGC) syndrome [3, 5, 6]. *CDH1* and *CTNNA1* genes encode epithelial adherens-junctions proteins, E-cadherin and αE-catenin respectively [7, 8], whose loss of function reduces cell-cell adhesion and increases invasiveness and cell survival, among other consequences [7, 9]. HDGC is autosomal dominant and predisposes to diffuse gastric cancer (DGC) and lobular breast cancer (LBC) [1, 3].

In *CDH1*-associated HDGC, the cumulative DGC incidence ranges from 42% to 72% in males [10, 11], and from 33% to 55% in females, while the risk of LBC ranges from 39% to 55% in females, by 80 years of age [10-13]. Even taking into account that estimates are likely to be biased, either due to patient geographical origin or due to clinical ascertainment [11], the risk of developing DGC or LBC remains high across different cohorts and studies [3]. Individuals with suspected-HDGC are identified and referred for genetic testing, according to guidelines and clinical criteria, which have been evolving since 1999 (**supplementary Figure 1**) [14-17], based on technological advances and recognition of novel HDGC-related clinical presentations [3]. Identification of a *CDH1* variant in the proband, classified as pathogenic or likely pathogenic, according to the American College of Medical Genetics and Genomics and the Association for Molecular Pathology (ACMG/AMP) variant curation guidelines for *CDH1* germline variants [18], leads to cascade genetic testing in other family members [17]. Once carrier status is confirmed in asymptomatic relatives, endoscopic surveillance and/or prophylactic surgery are offered according to guidelines to prevent DGC in both genders and/or LBC in females [17]. (Blair et al. 2020b)

Many families fulfilling the 2020 HDGC clinical criteria remain genetically unexplained, despite the use of candidate-gene, whole-exome (WES), and whole-genome (WGS) sequencing approaches [6, 19]. Missing hereditability in such families may derive from screening approaches excluding CNV analysis or by technical limitations still existing in next generation sequencing (NGS) protocols and analyses pipelines [20, 21]. In addition, missing hereditability may be hidden in regulatory elements or unscreened genomic regions [22](Kamps et al. 2017; Vos et al. 2019). The 3D chromatin architecture, and enhancers, may drive gene expression across large genomic distances through physical proximity to the promoter [23]. These mechanisms are often cell-type specific and control overall gene expression, including that of important disease-causing genes [24, 25](Andrey and Mundlos 2017b). In fact, the gene expression profile of some disease-causing genes is controlled by regulatory elements located in gene bodies, intergenic regions in their vicinity, and other regions of the causal-gene topological-associating domain [26]. CNVs or other type of aberrations affecting these regulatory regions, which consequently interfere with the normal 3D-

chromatin architecture, may lead to misexpression patterns, causing or aggravating disease [26].

Given the advances in knowledge on all above-described layers, retesting families who tested negative prior to a syndrome-specific breakthrough, implementation of relevant technological improvements, or integration of well-validated novel disease-causative mechanisms in routine testing, are important aspects in diagnosis and management [27-29].

Herein, we wish to characterize a novel inherited mechanism, that is simultameously disease causing and likely responsible for extremely early-onset gastric cancer and atypical high-penetrance of HDGC.

Materials/Subjects and Methods

Subjects

A HDGC-suspected family from Mexico was enrolled for meeting the 1999 HDGC clinical criteria [14]. The family presents eight DGC with signet-ring cells. WES of individuals III-1, III-3, III-5, III-6, III-7, III-9 were available for CNVs re-analysis. Multiplex ligation-dependent probe amplification (MLPA) validation was performed for individuals III-1, III-5, III-6, III-9, IV-1, IV-3.

Whole exome sequencing (WES) re-analysis

WES from HDGC-suspected family [19] was re-analyzed and screened for CNVs using three variant callers (ExomeDepth [30], ClinCNV [31] and Manta [32]) in tumour risk syndromes-associated genes (supplementary **Table 1**) and annotated using AnnotSV [33]. Candidate calls were prioritized considering quality scores, calling by more than one caller and visual evaluation using Integrated Genome Viewer (IGV).

Multiplex ligation-dependent probe amplification (MLPA)

Specific-*CDH1* probes (SALSA MLPA-Probemix P083 *CDH1*, MRC-Holland) were used to identify gene dosage alterations, as previously described [2, 34] and following manufacturer's recommendations. Probemix include 20 *CDH1*-specific probes, 2 probes in *CDH1* flanking genes and 13 reference probes in relatively copy number stable regions. Data analysis was performed using Coffalyser.Net (MRC-Holland).

ATAC-seq and 4C-seq

ATAC-seq data was collected from phase 3 ENCODE project [35]. For 4C-seq, gastric mucosa cells from bariatric surgeries were scraped and enzymatically dissociated. 4C-seq libraries were generated, as previously described [36, 37], digested with DnpII (New England Biolabs) and Csp6I (New England Biolabs) and re-circularized with T4 DNA Ligase (Thermo Fisher Scientific). 4C-seq libraries were PCR-amplified (**supplementary Table 2**) and sequenced on Illumina HiSeqX technology.

CDH1 interactions on a genome scale were mapped, based on Pipe4C [37] and PeakC [38] with window size 2, alpha fdr 0.1 and minimal distance 500.

Cell culture

MKN74 human gastric cancer cell line was purchased from the japanese collection of research bioresources cell bank. MKN74 and isogenic clones were cultured in RPMI medium (Gibco), supplemented with 10% fetal bovine serum (Biowest) and 1% penicillin streptomycin (Gibco), and maintained at 37°C under 5% CO_2 humidified atmosphere.

CRISPR-Cas9 editing

sgRNAs were designed to target the family CNV or the *CDH1* portion of that CNV using Benchling online platform (**supplementary Table 3**). Individual sgRNAs (Invitrogen) were cloned in LentiCRISPRv2GFP (addgene 82416) or LentiCRISPRv2-mCherry (addgene 99154) vectors using BsmBIv2 (New England Biolabs). Plasmids were transformed into Stbl3 competent cells and colonies were sequenced (supplementary **Table 4**). Lentiviral particles were produced resourcing to HEK293T cell line with pMD2.G (addgene 12259) and pCMV-dR8.91 (addgene) vectors, following Lipofectamine 3000 manufacture's protocol (Invitrogen) and collected at 48h. MKN74 was infected with pairs of lentivirus particles in medium supplemented with 10µg/µl hexadimethrine bromide (Merck Life Science S.L.U.) for 48h. Transduced cells were selected for GFP and mCherry positive expression at 7 days post-infection using FACS ARIA (BD Biosciences).

Genotyping of edited clones

gDNA was extracted using NZY Tissue gDNA isolation kit (NZYTech), according to the manufacturers' protocol. gDNA was amplified using primers flanking the edition sites (**supplementary Table 4**) and Multiplex PCR kit (Qiagen). PCR products were analysed in gel electrophoresis and Sanger sequenced on an ABI-3130 Genetic Analyzer (Applied Biosystems).

CDH1/E-cadherin expression analysis

CDH1 mRNA expression was assessed by qPCR in triplicates. RNA was extracted using *mir*Vana RNA Isolation Kit (Invitrogen), according to manufacturers' protocol. cDNA was synthesized using SuperScriptII reverse transcriptase (Invitrogen), according to the manufacturers' protocol. *CDH1* mRNA expression was analyzed by qPCR with KAPA PROBE FAST qPCR Master Mix (2X) Kit (Sigma-Aldrich) and probes for *CDH1* (Hs.PT.58.3324071, TaqMan) and 18S (custom assay, IDT). Reactions were sequenced on a 7500 Real-Time PCR System (Applied Biosystems). Relative expression was normalized for the endogenous 18S control and quantified using the $2^{-\Delta\Delta Ct}$ method.

E-cadherin expression was assessed by flow cytometry in triplicates. Cells were detached with Versene (Gibco) and blocked with 3% bovine serum albumin-phosphatase buffer saline. Cells were incubated with primary mouse monoclonal antibody HECD-1 (Invitrogen), washed and incubated with secondary antibody anti-mouse Alexa Fluor 647 (Invitrogen). Fluorescence was measured using FACS ARIA (BD Biosciences) and Flow Jo version 10 software was used to analyse the data.

Statistical analysis

Statistical analysis was performed using GraphPad Prism version 7.00 software (GraphPad Software Inc.). A T-student test was used for comparison analysis, assuming equal variance between clones and WT samples. Differences were considered significant when p-value<0.05.

RNA-seq and *whole transcriptome sequencing bioinformatics analysis*

Duplicated RNA samples collected from control and CRISPR-cas9 edited cell lines were sequenced with TruSeq stranded total RNA with Ribo-Zero Gold library in NovaSeq6000. Unique reads were mapped to GRCh38 human genome using STAR and RSeQC and annotated with USCS GRCh38 (Genecode v36). Batch effect was assessed using a PCA and corrected with ComBat_seq from sba (v.3.44) R package for replicates and samples. Differential expression analysis was performed using Deseq2 (v.1.36.0) R package and canonical transcripts were selected considering |log2fold-change|≥1 and corrected p-value<0.05. ClusterProfiler (v.4.4.4) R package was used for enrichment GO terms analysis (q-value<0.05). Statistics were performed using R (v.4.2.0).

Results

The genetic odyssey of a HDGC suspected family ends with the finding of a CDH1-TANGO6 deletion

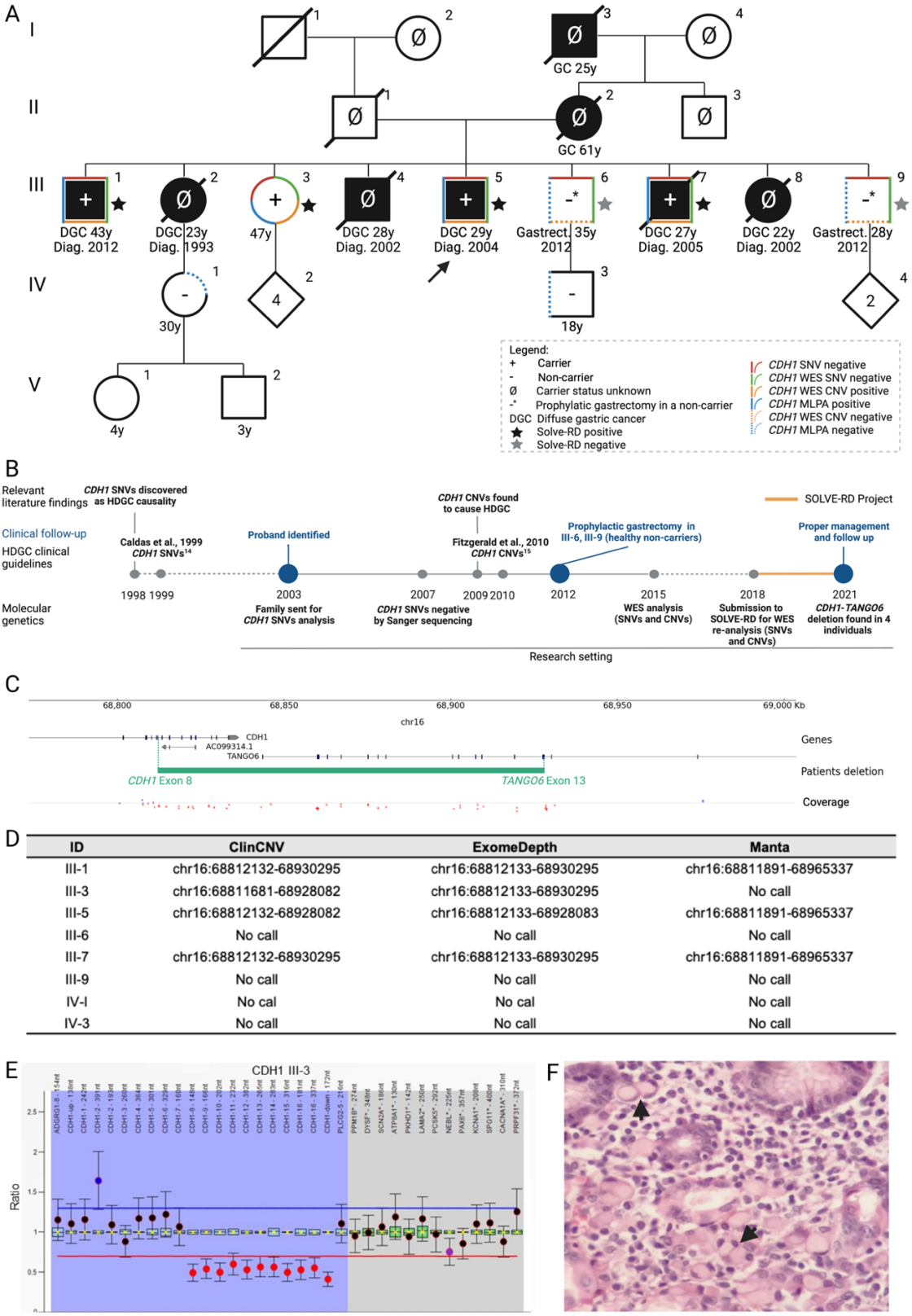
A family of Mexican origin, meeting the 1999 HDGC clinical criteria [14], was identified in 2003 showing five DGC in three consecutive generations, three occurring in the same generation at ages 23 (III-2), 28 (III-4) and 22 (III-8) (**Figure 1A**). In addition, individuals III-5 and III-7 also developed DGC at age 29 and 27, in 2004 and 2005, respectively. Given the highly suggestive HDGC pedigree, samples from family members that remained alive (III-1, III-3, III-5, III-6, III-7 and III-9), were all sent out for *CDH1* Sanger sequencing, under a research protocol (**Figure 1A,B**). The genetic test was negative for the presence of *CDH1* SNVs in 2007. Meanwhile, all asymptomatic family members were integrated into a gastric surveillance protocol, similar to that offered to genetically unsolved HDGC families [14]. In 2012, at age 43, individual III-1 developed DGC and underwent gastric surgery with curative intent. Due to the absence of a genetic diagnosis to find individuals at higher risk of DGC, the clinical team decided to offer (presumably) prophylactic total gastrectomy to asymptomatic individuals III-6 and III-9 (**Figure 1A**).

Although, *CDH1* CNVs were firstly described as a cause of HDGC in 2009 [2], and integrated in HDGC clinical guidelines in 2010 [15], samples from this family were not tested by MLPA. Samples were rather submitted to WES in 2015, again as part of a research protocol, and analysed for SNVs and CNVs which result turned out negative (**Figure 1B**), likely due to technical limitations in WES analysis or sample/sequencing quality [39].

WES re-analysis led to the identification of a heterozygous deletion involving half of the *CDH1* gene (starting downstream of *CDH1* exon 7) and the downstream *TANGO6* gene (ending upstream of *TANGO6* exon 14) (**Figure 1B,C,D**). This CNV was identified by several variant callers, namely 'ClinCNV' [31], 'Manta' [32](Chen et al. 2016) and 'ExomeDepth' [30] (**Figure 1C,D**), and supported by Integrative Genomics Viewer (IGV) visual inspection [40], despite the low quality of the sequencing data (**Figure 1C**). The CNV was after validated by MLPA (**Figure 1E**) in individuals III-1, III-3, III-5 and III-7 (**Figure 1A**). Individuals III-6 and III-9, who had opted for risk-reducing gastrectomy, were actually non-carriers of the CNV (**Figure 1A**).

As the pedigree evidences (**Figure 1A**), three out of four carriers in generation III, developed DGC with signet-ring cells at ages 27, 29 and 43 (**Figure 1F**). Individual III-3 is the only carrier to our knowledge that remains asymptomatic at the age of 47. Besides these family members, there are five not tested DGC patients, two (I-3 and II-2) died from DGC at ages 25 and 61, being in principle obligated carriers. There are three other relatives (III-2, III-4, and III-8) who died of DGC at age 23, 28 and 22 and were likely also carriers of this predisposing CNV (**Figure 1A**). After the genetic diagnosis, a blood sample was obtained from

individuals IV-1 and IV-3, who were old enough to be tested, and the CNV screening performed by MLPA returned a negative result for both (**Figure 1A**).



ID	ClinCNV	ExomeDepth	Manta
III-1	chr16:68812132-68930295	chr16:68812133-68930295	chr16:68811891-68965337
III-3	chr16:68811681-68928082	chr16:68812133-68930295	No call
III-5	chr16:68812132-68928082	chr16:68812133-68928083	chr16:68811891-68965337
III-6	No call	No call	No call
III-7	chr16:68812132-68930295	chr16:68812133-68930295	chr16:68811891-68965337
III-9	No call	No call	No call
IV-I	No cal	No cal	No call
IV-3	No call	No call	No call

Figure 1. *Diagnostic odyssey, pedigree, variant calling details and cancer histology of a HDGC family carrying a causative CDH1 CNV. A) Current family pedigree. Full black symbols: individuals with confirmed*

DGC; Red outline: negative for SNVs (Sanger sequencing); green outline: negative for SNVs (WES); orange outline: positive for CNVs (WES); blue outline: MLPA positive for the CDH1 deletion; dark grey: negative for CNVs (WES); purple: MLPA negative for the CDH1 deletion. All family member, submitted to multiple genetic tests herein described, signed informed consents. The research work has been approved by the Ethics Committee of Centro Hospitalar Universitário São João, in Porto, Portugal, in the frame of the Solve-RD project with the reference CHUSJ_445/2020; B) Timeline of events and diagnostic odyssey of the family; C) CDH1 CNV found by WES, encompassing part of the CDH1 and TANGO6 genes (represented in green) and IGV coverage; Green text represents first and last edited exons; D) Detected CDH1 CNV in ClinCNV [31] , ExomeDepth [30] and Manta [32]; E) MLPA analysis performed using SALSA MLPA-Probemix P083 CDH1 (MRC Holland) in patient III-3; Ratio above blue line indicate increased copy number, while ratio below red line indicate reduced copy number; Blue highlight represent CDH1 probes and 2 neighbour genes; Grey highlight represent reference probes; F) Haematoxylin and eosin staining of DGC depicting signet-ring cells (arrow heads) (100x magnification) from the proband III-5.

Chromatin conformation analysis highlights physical interactions between the CDH1 promoter and sequences within the deleted region in normal stomach

Although highly penetrant, HDGC caused by *CDH1* pathogenic or likely pathogenic variants is rarely as penetrant as in this family, which presents an uncommonly high number of individuals in a single generation affected by DGC (six out of seven carriers/presumed carriers). Additionally, five out of six developed DGC before age 29, and average age of cancer onset was 29±7.55 years in six carriers/presumed carriers from generation III. We therefore hypothesised that sequences within the deleted region, and extending into the downstream intergenic region, and the neighbouring *TANGO6* gene sequence, would affect *CDH1* expression and the *locus* 3D-chromatin structure more strongly, than deletion of the *CDH1* alone.

To address this hypothesis, we mined the GeneHancer database [41] and found that several enhancers within the deleted region, interact with either *TANGO6* or *CDH1* promoters (**Figure 2A**). Additionally, we used normal epithelial cells from bariatric surgeries to profile *CDH1*-promoter interactions using 4C-seq, and analysed accessible chromatin regions using ATAC-seq from the ENCODE project [35]. We found that the *CDH1* promoter interacts not only with downstream sequences within *CDH1*, but also *TANGO6* and the intergenic region, all deleted by the family CNV, and displaying regions of accessible chromatin (**Figure 2A**). These results suggest a role of *CDH1*/*TANGO6* co-regulation in stomach, besides gene physical proximity.

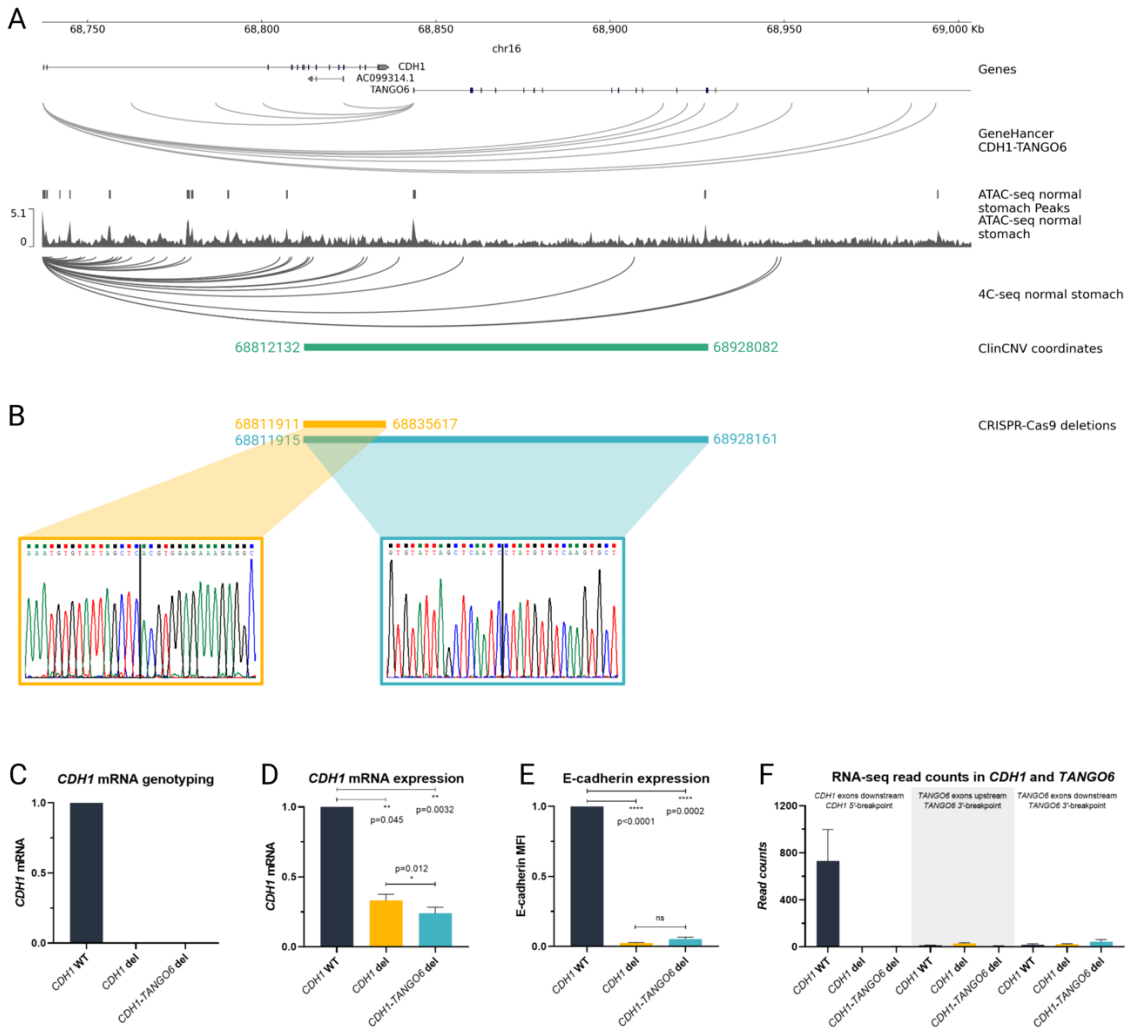


Figure 2. *CDH1 and TANGO6 regulatory network and characterization of CRISPR-cas9 edited clones. A) CDH1/TANGO6 potential enhancers depicted in GeneHancer, accessible chromatin and CDH1-promotor interactors in normal stomach tissue; B) Sanger sequencing genotyping of the CNV breakpoints, and deletion coordinates in CDH1 and CDH1-TANGO6 CRISPR-cas9 edited clones; C) mRNA genotyping of CDH1 and CDH1-TANGO6 edited clones resourcing to a probe located in the deleted CDH1 region; D) CDH1 mRNA expression of CDH1 and CDH1-TANGO6 edited clones resourcing to a probe located in exons 6-7 (Hs.PT.58.3324071, TaqMan); E) E-cadherin protein expression of CDH1 and CDH1-TANGO6 edited clones measured by flow cytometry (monoclonal antibody HECD-1, Invitrogen, 1:100 dilution; and mouse Alexa Fluor 647, Invitrogen, 1:500 dilution); F) RNA-seq expression read counts of CDH1 and TANGO6 genes. Data are represented as the mean ± SEM, MFI: median fluorescence intensity. Experiments depicted in panels C-E were performed in triplicates and differences considered statistically significant if p-value<0.05 in a t-test.*

Deletion of CDH1 and downstream regulatory sequences, present within the HDGC causal CNV, strongly impact CDH1 expression and specific signalling-pathways

To understand the impact of deleting several *CDH1* promoter interacting sequences, beyond the *CDH1 locus*, we CRISPR-Cas9 deleted either the *CDH1* portion of the CNV, or mimicked the family CNV. This was performed in a human

gastric cancer cell line bearing normal *CDH1*/E-cadherin expression and function [42]. We used sgRNAs flanking target regions (**supplementary Table 3**), and successful editing was confirmed by Sanger sequencing and qRT-PCR (**Figure 2B,C**). A clone bearing a 23.705bp homozygous deletion spanning from *CDH1* intron 7 and extending till exon 16 (chr16:68811911-68835617, hg38) and a clone bearing a 116.245bp homozygous deletion starting in *CDH1* intron 7 and ending after *TANGO6* intron 13 (chr16:68811915-68928161, hg38) were produced (**Figure 2B**).

We found that both *CDH1* and *CDH1-TANGO6* deletions led to a significant decrease in *CDH1* mRNA compared to the wild-type (WT) control (p-value=0.045 and p-value=0.0032, respectively), being this downregulation stronger for the *CDH1-TANGO6* deletion as compared to the *CDH1* deletion alone (p-value=0.012) (**Figure 2D**). The level of protein downregulation in relation to the WT control was equivalent in both deletion clones, as assessed by flow cytometry (p-value<0.0001 and p-value=0.0002, respectively) (**Figure 2E**).

We next explored the effect of both deletions in the cells' transcriptomes. For this, we evaluated the differential transcriptome between *CDH1* or *CDH1-TANGO6* deletion clones and WT cells, and between *CDH1-TANGO6* and *CDH1* deletion clones. We confirmed that *CDH1* mRNA expression was more strongly downregulated (3.4-fold) in the *CDH1-TANGO6* deletion than in the *CDH1* deletion alone (1.6-fold), when each was compared to *CDH1* mRNA expression in WT cells (p-value<0.0001) (**Figure 2F,3A,B**). We also verified that *TANGO6* is not expressed in WT cells, and its expression remains, as expected, unchanged after the 5'-deletion of the *TANGO6* gene (**Figure 2F**).

We also observed that the global pattern of differentially expressed genes clearly differ between clones bearing the *CDH1* deletion alone, the *CDH1-TANGO6* deletion and the *CDH1* WT cells (**Figure 3D**). There were 3 272 differentially expressed genes between the *CDH1* deleted clone and *CDH1* WT cells; 2 492 differentially expressed genes between the *CDH1-TANGO6* deleted clone and *CDH1* WT cells, and; 1 102 differentially expressed genes between the *CDH1-TANGO6* and the *CDH1* deleted clones. Gene ontology analysis further revealed that 'nucleosome assembly', 'chromatin remodelling', 'regulation of small GTPase mediated signal transduction', as well as pathways related to cyclin-dependent protein kinase activity and actin-cytoskeleton were impacted in both in *CDH1* alone and in the *CDH1-TANGO6* deletion clones (**Figure 3E**).

CDH1 deletion exclusive events encompass downregulation of 'mitotic nuclear division', 'regulation of spindle organization' and planar polarity-related pathways, as well as upregulation of 'lipid catabolic processes', among others (**Figure 3E**). Events exclusively altered in the *CDH1-TANGO6* deletion clone include mainly impairment of adhesion- and cytoskeleton-associated pathways (**Figure 3E**).

The direct comparison between the transcriptional profile of clones bearing the *CDH1-TANGO6* deletion or the *CDH1* deletion alone, emphasise the negative

impact in adhesion-, p53-mediated DNA damage response and cytoskeleton-associated pathways and the positive impact in MAPK cascade associated proliferation, presented by the *CDH1-TANGO6* deletion clone (**Figure 3E**). Additional positive and negative impacts were found in several immune-associated pathways, namely related to 'cellular response to type I interferon' and 'response to virus' in the *CDH1-TANGO6* deletion clone (**Figure 3E,F**). The expression of *IFITM* genes (*IFITM1*, *IFITM3*, *IFITM2*) and some of their upstream regulators (STAT1, *KLF4*) was particularly affected in the *CDH1-TANGO6* deletion clone (**Figure 3C,F**).

Altogether, and upon data integration, cancer cells arising in patients bearing the *CDH1-TANGO6* deletion are likely to present decreased apoptosis, increased proliferation and invasion, and no cell cycle arrest (**Figure 3F**).

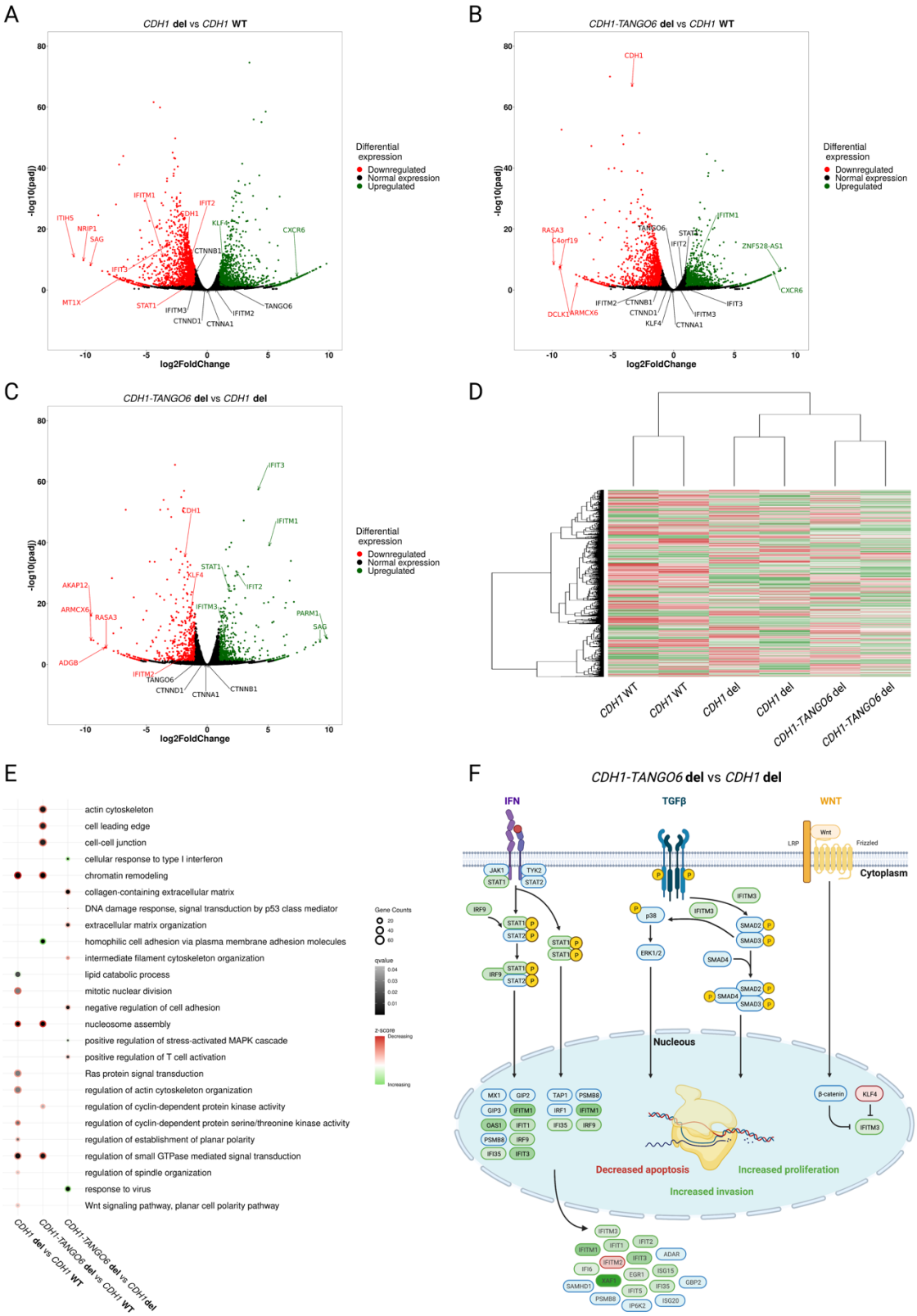


Figure 3. *Genome-wide transcriptome regulation of CDH1 CNV and CDH1-TANGO6 CNV. A) Volcano-plot illustrating CDH1 CNV differentially expressed genes with CDH1 WT; B) Volcano-plot illustrating CDH1-TANGO6 CNV differentially expressed genes with CDH1 WT; C) Volcano-plot illustrating CDH1-TANGO6 CNV differentially expressed genes with CDH1 CNV; D) Heatmap illustrating genome-wide differentially expressed genes; E) Selected genome-wide transcriptomic misregulated pathways of CDH1-TANGO6 CNV and CDH1 CNV. Grey scale represents q-value and red-green scale represents z-score (red for pathways*

with mainly downregulated genes, white for equally downregulated and upregulated genes and green for pathways with mainly upregulated genes); F) Response to interferon, TGF-β and Wnt signalling pathways misregulated in CDH1-TANGO6 CNV vs CDH1 CNV, created with BioRender.com. Red-green scale represents down to upregulated genes and blue represents normal expressed genes.

Discussion

In this work, we demonstrated that regulatory elements, lying within the *CDH1-TANGO6* deletion, are part of a chromatin network that, not only control *CDH1* expression, but also influence the cell transcriptome and associated signalling pathways. We believe this finding explains the high penetrance and extremely young age of cancer-onset in patients from this family. In addition, we describe the diagnostic odyssey of a HDGC-suspected family, that remained genetically unsolved for 18 years, to highlight the importance of incorporating scientific-technological updates and clinical guidelines in routine diagnosis. This is expected to ultimately improve clinical management and care of hereditary cancer patients and families.

A distinctive feature of this family, in comparison with other *CDH1*-positive HDGC families, is the high disease penetrance reflected in generation III, with six out of seven individuals being CNV carriers or presumed carriers and developing DGC. The other is the very early age of cancer-onset (29±7.55 years), compared to the average age reported for DGC in most HDGC families (46,7±16,6 years) [3, 11, 43]. In accordance with our findings, an additional HDGC family has been reported in the literature, bearing a 275kb *CDH1* (exon 7) and *TANGO6* (full gene) deletion, with four patients in four consecutive generations affected by gastric cancer, from which two were confirmed with signet ring cell carcinoma at ages 30 and 34 years old [44]. These features prompted the next question of this study, which was related to a potential role for the downstream intergenic region and the neighbouring *TANGO6* gene sequence, in modulating *CDH1* expression, the *CDH1 locus* 3D-chromatin structure and the transcriptome, in aggravating the clinical presentation in this family.

CRISPR-Cas9 clones mimicking the *CDH1* deletion alone or the family full CNV, demonstrated that the full CNV was more efficient in promoting downregulation of *CDH1* mRNA than the *CDH1* deletion alone. This finding supports the existence of regulatory DNA elements controlling *CDH1* expression, downstream of the *CDH1 locus*, and enclosed in the CNV, which we have also proved to interact with the *CDH1* promoter. Similar findings have been previously reported, as for instance those related to neighbour murine *Shh* and *Lmbr1* genes, in which the ZRS enhancer located within the *Lmbr1* gene, regulates *Shh* gene expression [45].

Our data on differential expression and gene ontology analysis of RNA-seq data from clones mimicking the *CDH1* deletion alone or the family full CNV in comparison to *CDH1* WT, highlighted shared biological terms between these

clones related to cell division and spindle orientation, and involving chromatin remodelling. These terms have been previously associated with E-cadherin function [46-49], and likely represent the contribution of the *CDH1* locus deletion itself to the differential transcriptome. The fact that adhesion-related ('negative regulation of cell adhesion') and actin-related pathways ('actin cytoskeleton') were found specifically downregulated upon *CDH1-TANGO6* deletion, and in comparison, to *CDH1* deleted cells, suggests that the perturbations induced by the *CDH1-TANGO6* deletion are greater than those induced by the *CDH1* deletion alone in these pathways. Although it is widely known that E-cadherin-deficiency leads to weaker adhesion and abnormal microtubule organization [50, 51], which may be involved in migration upon malignant transformation [52], our data specifically links homophilic cell adhesion to the longer *CDH1-TANGO6* deletion. Indeed, this clone presents upregulation of the pathway 'homophilic cell adhesion via plasma membrane adhesion molecules', which may reflect a yet to prove adhesion-compensatory mechanism, promoted by upregulation of protocadherins in the absence of E-cadherin, seen in our data. A similar switch has been previously demonstrated in lobular breast cancer, whereby E-cadherin is replaced by P-cadherin in cancer cells [53].

Immune-associated genes and pathways were particularly altered in the *CDH1-TANGO6* deletion clone, comparing to the clone bearing the *CDH1* deletion alone. *IFITM1* gene was particularly overexpressed in the *CDH1-TANGO6* deletion, a feature previously observed in several aggressive tumours, and shown to enhance tumour proliferation and invasion [42, 54]. In gastroesophageal tumours, for instance, *IFITM1* was found overexpressed comparing to normal adjacent tissues [55-57]. Additionally, *IFITM3* overexpression has been shown to promote gastric cancer growth, through the activation of several pathways, resulting in increased proliferation, invasion, and metastasis [58, 59]. IFITM3 has been shown to promote SMAD2/3/4 phophorilation and interact with STAT1/2 and PIP3/PI3K. This results in activation of downstream proliferation and oncogenic signalling pathways, such as p38/MAPK [60], JAK/STAT [61-63] and PI3K [64]. IFITM3 plays an additional role in the Wnt signalling pathway, being regulated by β-catenin [58] and KLF4 [65]. Indeed, KLF4, known to promote *IFITM3* transcriptional inhibition [65], is downregulated in the *CDH1-TANGO6* deletion clone providing a rationale for the *IFITM3* upregulation seen in this clone. Finally, *IFITM2,* a pro-apoptotic gene, encoding a protein which has been linked to protection against tumour proliferation [66], was found downregulated in the *CDH1-TANGO6* deletion clone. Most abovementioned observations support a relationship between *IFITM* proteins misregulation and the severe phenotype, observed in this HDGC family. Increased proliferation associated with MAPK cascade [67, 68] and decreased DNA damage response mediated by p53 signalling [69-71], also found in cells bearing the family CNV, may hint to more aggressive DGC, than the commonly described for HDGC patients bearing *CDH1* coding variants, who generally die by dissemination to the peritoneum [43]. Supporting this, is the fact that individual

III-4 died at age 28 with DGC metastases in the lung and bone, a very unusual event in patients with coding *CDH1* variants.

Our data also seems to indicate that while a role has not been found for the *TANGO6* mRNA itself, the genomic sequences downstream of the *CDH1 locus*, and within the sequence deleted in this family, host positive regulators of *CDH1* expression. This would explain not only the lower *CDH1* mRNA expression, but also the wider transcriptional and signalling impact, which likely contribute to the clinical presentation in this family. Only two HDGC families have been reported to carry large deletions involving specifically *CDH1* and *TANGO6*, both presenting early onset DGC and gastric cancer in consecutive generations [44]. Therefore, evidence for the role of regulatory *CDH1* downstream regions in early onset DGC is still limited. Yet, this study provides the first evidence on the role of *CDH1* downstream sequences in *CDH1* expression control, which may apply to other HDGC families harbouring similar deletions.

This HDGC-suspected family with a severe clinical presentation was somehow lost in genetic follow-up, as clinical standardized diagnostic tests were never used in their clinical path. Negative tests for *CDH1* coding SNVs and CNVs, both were obtained from research projects, although management guidelines proposed formal genetic testing with Sanger sequencing and MLPA since 2010 in families fulfilling HDGC clinical criteria [2, 15]. This highlights the risk associated with molecular diagnosis being performed in the frame of research studies, rather than in proper diagnostic and clinically approved laboratories, as part of patients' and families' pathways of care. In this case, the most obvious disease gene/causing mechanism was overlooked, as for research purposes, these should have been, in principle, excluded. The result of a WES data re-analysis 18-years after clinical diagnosis, revealed a CNV involving the most obvious HDGC-causing gene, *CDH1* [1, 3, 6, 17]. This case highlights the importance of revisiting, re-contacting, and retesting families fulfilling clinical criteria but lacking molecular diagnosis, in diagnostic and clinically approved laboratories, as novel and consolidated genetic susceptibility causes of hereditary syndromes arise, and both technologies and management guidelines evolve.

This family odyssey also illustrates the importance of timely genetic diagnosis. For this family, the greater consequences included two carriers developing cancer, two and nine years after the family was recognized clinically as an HDGC-suspected family; and two non-carrier siblings undergoing unnecessary and presumably prophylactic total gastrectomy, nine years after clinical diagnosis, based on family history only. The Solve-RD project (https://solve-rd.eu/) emerged in this case, as a solution to overcome an unfortunate diagnostic odyssey. The finding of a causal *CDH1* CNV was reported to the clinical team at the hospital of origin in 2021, which promoted re-engagement with the family and effective genetic counselling and disease risk management.

Taken together, our data suggests that patients bearing the *CDH1-TANGO6* deletion develop high penetrance, earlier-onset, and possibly more aggressive DGC, than patients with pathogenic variants confined to the *CDH1*, likely due to the impairment of regulatory elements able to activate oncogenic signalling pathways. It also illustrates the inequities and complexities of genetic diagnosis and integrated care in rare diseases, in a rapidly evolving field, which can impact timely genetic diagnosis and disease prevention.

Acknowledgments

This work was supported by: (1) the European Reference Network on Genetic Tumour Risk Syndromes (ERN GENTURIS) – Project ID No. 739547. ERN GENTURIS is partly co-funded by the European Union within the framework of the Third Health Programme “ERN-2016 – Framework Partnership Agreement 2017–2021”; (2) Doctoral Programme in Biomedicine - Faculty of Medicine, University of Porto, to JG-P and CSJ; (3) Doctoral Programme in Computational Sciences, Faculty of Sciences, University of Porto, to MF; (4) Masters in Molecular Oncology, Faculty of Medicine, University of Porto, to NM. The team would also like to thank Nuno Alves from i3S for fruitful discussions on immuno-oncology; Sónia Sousa, Cátia Pereira and Luis Cirnes from Ipatimup Diagnostics, Porto-Portugal for MLPA analysis.

Funding

This work received funding from: (1) the European Union’s Horizon 2020 research and innovation programme under grant agreement No 779257 (Solve-RD: H2020-SC1-2017-Single-Stage- RTD), which provided Ph.D. salaries to JG-P, AS, ItP; (2) The Euro- pean Regional Development Fund (ERDF) through the COMPETE 2020-Operacional Programme for Competitiveness and Internation- alisation (POCI), Portugal 2020, and by Portuguese funds through Fundação para a Ciência e Tecnologia (FCT)/Ministério da Ciência, Tecnologia e Inovação (LEGOH: PTDC/BTM-TEC/6706/2020); (3) FCT with salary support for Ph.D. studentships to CSJ (Ref. SFRH/BD/140796/2018), JG-P (Ref. 2022.11952.BD) and MF (Ref. 2020.05763.BD); (4) MT is supported by the NIHR Cambridge Bio- medical Research Centre (BRC-1215-20014). Open access funding provided by FCT|FCCN (b-on).

Conflict of interest: The authors disclosure no conflicts of interest.

References

1.Guilford P, Hopkins J, Harraway J, McLeod M, McLeod N, Harawira P, et al. E-cadherin germline mutations in familial gastric cancer. Nature. 1998;392(6674):402-5. doi: 10.1038/32918. PubMed PMID: 9537325.

2.Oliveira C, Senz J, Kaurah P, Pinheiro H, Sanges R, Haegert A, et al. Germline *CDH1* deletions in hereditary diffuse gastric cancer families. Hum Mol Genet. 2009;18(9):1545-55. Epub 20090124. doi: 10.1093/hmg/ddp046. PubMed PMID: 19168852; PubMed Central PMCID: PMC2667284.

3.Garcia-Pelaez J, Barbosa-Matos R, Lobo S, Dias A, Garrido L, Castedo S, et al. Genotype-first approach to identify associations between *CDH1* germline variants and cancer phenotypes: a multicentre study by the European Reference Network on Genetic Tumour Risk Syndromes. Lancet Oncol. 2022. Epub 20221124. doi: 10.1016/s1470-2045(22)00643-x. PubMed PMID: 36436516.

4.Majewski IJ, Kluijt I, Cats A, Scerri TS, de Jong D, Kluin RJ, et al. An α-E-catenin (*CTNNA1*) mutation in hereditary diffuse gastric cancer. J Pathol. 2013;229(4):621-9. doi: 10.1002/path.4152. PubMed PMID: 23208944.

5.Lobo S, Benusiglio PR, Coulet F, Boussemart L, Golmard L, Spier I, et al. Cancer predisposition and germline *CTNNA1* variants. Eur J Med Genet. 2021;64(10):104316. Epub 20210821. doi: 10.1016/j.ejmg.2021.104316. PubMed PMID: 34425242.

6.Garcia-Pelaez J, Barbosa-Matos R, São José C, Sousa S, Gullo I, Hoogerbrugge N, et al. Gastric cancer genetic predisposition and clinical presentations: Established heritable causes and potential candidate genes. Eur J Med Genet. 2022;65(1):104401. Epub 20211203. doi: 10.1016/j.ejmg.2021.104401. PubMed PMID: 34871783.

7.Harris TJ, Tepass U. Adherens junctions: from molecules to morphogenesis. Nat Rev Mol Cell Biol. 2010;11(7):502-14. doi: 10.1038/nrm2927. PubMed PMID: 20571587.

8.Takeichi M. Multiple functions of α-catenin beyond cell adhesion regulation. Curr Opin Cell Biol. 2018;54:24-9. Epub 20180308. doi: 10.1016/j.ceb.2018.02.014. PubMed PMID: 29525243.

9.Figueiredo J, Mercadillo F, Melo S, Barroso A, Gonçalves M, Díaz-Tasende J, et al. Germline *CDH1* G212E Missense Variant: Combining Clinical, In Vitro and In Vivo Strategies to Unravel Disease Burden. Cancers (Basel). 2021;13(17). Epub 20210828. doi: 10.3390/cancers13174359. PubMed PMID: 34503169; PubMed Central PMCID: PMC8430832.

10.Hansford S, Kaurah P, Li-Chang H, Woo M, Senz J, Pinheiro H, et al. Hereditary Diffuse Gastric Cancer Syndrome: *CDH1* Mutations and Beyond. JAMA Oncol. 2015;1(1):23-32. doi: 10.1001/jamaoncol.2014.168. PubMed PMID: 26182300.

11.Roberts ME, Ranola JMO, Marshall ML, Susswein LR, Graceffo S, Bohnert K, et al. Comparison of *CDH1* Penetrance Estimates in Clinically Ascertained Families vs Families Ascertained for Multiple Gastric Cancers. JAMA Oncol. 2019;5(9):1325-31. doi: 10.1001/jamaoncol.2019.1208. PubMed PMID: 31246251; PubMed Central PMCID: PMC6604087.

12.Pharoah PD, Guilford P, Caldas C. Incidence of gastric cancer and breast cancer in *CDH1* (E-cadherin) mutation carriers from hereditary diffuse gastric cancer families. Gastroenterology. 2001;121(6):1348-53. doi: 10.1053/gast.2001.29611. PubMed PMID: 11729114.

13.Xicola RM, Li S, Rodriguez N, Reinecke P, Karam R, Speare V, et al. Clinical features and cancer risk in families with pathogenic *CDH1* variants irrespective of clinical criteria. J Med Genet. 2019;56(12):838-43. Epub 20190711. doi: 10.1136/jmedgenet-2019-105991. PubMed PMID: 31296550.

14.Caldas C, Carneiro F, Lynch HT, Yokota J, Wiesner GL, Powell SM, et al. Familial gastric cancer: overview and guidelines for management. J Med Genet. 1999;36(12):873-80. Epub 1999/12/14. PubMed PMID: 10593993; PubMed Central PMCID: PMC1734270.

15.Fitzgerald RC, Hardwick R, Huntsman D, Carneiro F, Guilford P, Blair V, et al. Hereditary diffuse gastric cancer: updated consensus guidelines for clinical management and directions for future research. J Med Genet. 2010;47(7):436-44. doi: 10.1136/jmg.2009.074237. PubMed PMID: 20591882; PubMed Central PMCID: PMC2991043.

16.van der Post RS, Vogelaar IP, Carneiro F, Guilford P, Huntsman D, Hoogerbrugge N, et al. Hereditary diffuse gastric cancer: updated clinical guidelines with an emphasis on germline *CDH1* mutation carriers. J Med Genet. 2015;52(6):361-74.Epub 20150515. doi: 10.1136/jmedgenet-2015-103094. PubMed PMID: 25979631; PubMed Central PMCID: PMC4453626.

17.Blair VR, McLeod M, Carneiro F, Coit DG, D'Addario JL, van Dieren JM, et al. Hereditary diffuse gastric cancer: updated clinical practice guidelines. Lancet Oncol. 2020;21(8):e386-e97. doi: 10.1016/s1470-2045(20)30219-9. PubMed PMID: 32758476; PubMed Central PMCID: PMC7116190.

18.Lee K, Krempely K, Roberts ME, Anderson MJ, Carneiro F, Chao E, et al. Specifications of the ACMG/AMP variant curation guidelines for the analysis of germline *CDH1* sequence variants. Hum Mutat. 2018;39(11):1553-68. doi: 10.1002/humu.23650. PubMed PMID: 30311375; PubMed Central PMCID: PMC6188664.

19.Sommer AK, Te Paske I, Garcia-Pelaez J, Laner A, Holinski-Feder E, Steinke-Lange V, et al. Solving the genetic aetiology of hereditary gastrointestinal tumour syndromes- a collaborative multicentre endeavour within the project Solve-RD. Eur J Med Genet. 2022;65(5):104475. Epub 20220311. doi: 10.1016/j.ejmg.2022.104475. PubMed PMID: 35283344.

20.Kamps R, Brandão RD, Bosch BJ, Paulussen AD, Xanthoulea S, Blok MJ, et al. Next-Generation Sequencing in Oncology: Genetic Diagnosis, Risk Prediction and Cancer Classification. Int J Mol Sci. 2017;18(2). Epub 20170131. doi: 10.3390/ijms18020308. PubMed PMID: 28146134; PubMed Central PMCID: PMC5343844.

21.Vos JR, Giepmans L, Röhl C, Geverink N, Hoogerbrugge N. Boosting care and knowledge about hereditary cancer: European Reference Network on Genetic Tumour Risk Syndromes. Fam Cancer. 2019;18(2):281-4. doi: 10.1007/s10689-018-0110-6. PubMed PMID: 30302652; PubMed Central PMCID: PMC6449296.

22.Allou L, Balzano S, Magg A, Quinodoz M, Royer-Bertrand B, Schöpflin R, et al. Non-coding deletions identify Maenli lncRNA as a limb-specific En1 regulator. Nature. 2021;592(7852):93-8. Epub 2021/02/12. doi: 10.1038/s41586-021-03208-9. PubMed PMID: 33568816.

23.Andrey G, Mundlos S. The three-dimensional genome: regulating gene expression during pluripotency and development. Development. 2017;144(20):3646-58. doi: 10.1242/dev.148304. PubMed PMID: 29042476.

24.Cannavò E, Khoueiry P, Garfield DA, Geeleher P, Zichner T, Gustafson EH, et al. Shadow Enhancers Are Pervasive Features of Developmental Regulatory Networks. Curr Biol. 2016;26(1):38-51. Epub 20151210. doi: 10.1016/j.cub.2015.11.034. PubMed PMID: 26687625; PubMed Central PMCID: PMC4712172.

25.Kragesteen BK, Spielmann M, Paliou C, Heinrich V, Schöpflin R, Esposito A, et al. Dynamic 3D chromatin architecture contributes to enhancer specificity and limb morphogenesis. Nat Genet. 2018;50(10):1463-73. Epub 20180927. doi: 10.1038/s41588-018-0221-x. PubMed PMID: 30262816.

26.Will AJ, Cova G, Osterwalder M, Chan WL, Wittler L, Brieske N, et al. Composition and dosage of a multipartite enhancer cluster control developmental expression of Ihh (Indian hedgehog). Nat Genet. 2017;49(10):1539-45. Epub 20170828. doi: 10.1038/ng.3939. PubMed PMID: 28846100; PubMed Central PMCID: PMC5617800.

27.Deignan JL, Chung WK, Kearney HM, Monaghan KG, Rehder CW, Chao EC, et al. Points to consider in the reevaluation and reanalysis of genomictest results: a statement of the American College of Medical Genetics and Genomics(ACMG). Genetics in Medicine. 2019;21(6):1267-70. doi: 10.1038/s41436-019-0478-1.

28.Robertson AJ, Tan NB, Spurdle AB, Metke-Jimenez A, Sullivan C, Waddell N. Re-analysis of genomic data: An overview of the mechanisms and complexities of clinical adoption. Genet Med. 2022;24(4):798-810. Epub 20220120. doi: 10.1016/j.gim.2021.12.011. PubMed PMID: 35065883.

29.Dettwyler SA, Koeppe ES, Jacobs MF, Stoffel EM. Outcomes of retesting in patients with previously uninformative cancer genetics evaluations. Fam Cancer. 2022;21(3):375-85. Epub 20210921. doi: 10.1007/s10689-021-00276-8. PubMed PMID: 34545504; PubMed Central PMCID: PMC8934750.

30.Plagnol V, Curtis J, Epstein M, Mok KY, Stebbings E, Grigoriadou S, et al. A robust model for read count data in exome sequencing experiments and implications for copy number variant calling. Bioinformatics. 2012;28(21):2747-54. Epub 20120831. doi: 10.1093/bioinformatics/bts526. PubMed PMID: 22942019; PubMed Central PMCID: PMC3476336.

31.Demidov G, Sturm M, Ossowski S. ClinCNV: multi-sample germline CNV detection in NGS data. bioRxiv. 2022:2022.06.10.495642. doi: 10.1101/2022.06.10.495642.
32.Chen X, Schulz-Trieglaff O, Shaw R, Barnes B, Schlesinger F, Källberg M, et al. Manta: rapid detection of structural variants and indels for germline and cancer sequencing applications. Bioinformatics. 2016;32(8):1220-2. Epub 20151208. doi: 10.1093/bioinformatics/btv710. PubMed PMID: 26647377.
33.Geoffroy V, Herenger Y, Kress A, Stoetzel C, Piton A, Dollfus H, et al. AnnotSV: an integrated tool for structural variations annotation. Bioinformatics. 2018;34(20):3572-4. doi: 10.1093/bioinformatics/bty304. PubMed PMID: 29669011.
34.Schouten JP, McElgunn CJ, Waaijer R, Zwijnenburg D, Diepvens F, Pals G. Relative quantification of 40 nucleic acid sequences by multiplex ligation-dependent probe amplification. Nucleic Acids Res. 2002;30(12):e57. doi: 10.1093/nar/gnf056. PubMed PMID: 12060695; PubMed Central PMCID: PMC117299.
35.Moore JE, Purcaro MJ, Pratt HE, Epstein CB, Shoresh N, Adrian J, et al. Expanded encyclopaedias of DNA elements in the human and mouse genomes. Nature. 2020;583(7818):699-710. Epub 20200729. doi: 10.1038/s41586-020-2493-4. PubMed PMID: 32728249; PubMed Central PMCID: PMC7410828.
36.van de Werken HJ, de Vree PJ, Splinter E, Holwerda SJ, Klous P, de Wit E, et al. 4C technology: protocols and data analysis. Methods Enzymol. 2012;513:89-112. doi: 10.1016/b978-0-12-391938-0.00004-5. PubMed PMID: 22929766.
37.Krijger PHL, Geeven G, Bianchi V, Hilvering CRE, de Laat W. 4C-seq from beginning to end: A detailed protocol for sample preparation and data analysis. Methods. 2020;170:17-32. doi: https://doi.org/10.1016/j.ymeth.2019.07.014.
38.Geeven G, Teunissen H, de Laat W, de Wit E. peakC: a flexible, non-parametric peak calling package for 4C and Capture-C data. Nucleic Acids Res. 2018;46(15):e91. doi: 10.1093/nar/gky443. PubMed PMID: 29800273; PubMed Central PMCID: PMC6125690.
39.Fewings E, Larionov A, Redman J, Goldgraben MA, Scarth J, Richardson S, et al. Germline pathogenic variants in PALB2 and other cancer-predisposing genes in families with hereditary diffuse gastric cancer without *CDH1* mutation: a whole-exome sequencing study. Lancet Gastroenterol Hepatol. 2018;3(7):489-98. Epub 20180427. doi: 10.1016/s2468-1253(18)30079-7. PubMed PMID: 29706558; PubMed Central PMCID: PMC5992580.
40.Robinson JT, Thorvaldsdóttir H, Winckler W, Guttman M, Lander ES, Getz G, et al. Integrative genomics viewer. Nat Biotechnol. 2011;29(1):24-6. doi: 10.1038/nbt.1754. PubMed PMID: 21221095; PubMed Central PMCID: PMC3346182.
41.Fishilevich S, Nudel R, Rappaport N, Hadar R, Plaschkes I, Iny Stein T, et al. GeneHancer: genome-wide integration of enhancers and target genes in GeneCards. Database (Oxford). 2017;2017. doi: 10.1093/database/bax028. PubMed PMID: 28605766; PubMed Central PMCID: PMC5467550.
42.Pinheiro H, Carvalho J, Oliveira P, Ferreira D, Pinto MT, Osório H, et al. Transcription initiation arising from E-cadherin/*CDH1* intron2: a novel protein isoform that increases gastric cancer cell invasion and angiogenesis. Hum Mol Genet. 2012;21(19):4253-69. Epub 20120629. doi: 10.1093/hmg/dds248. PubMed PMID: 22752307.
43.Oliveira C, Pinheiro H, Figueiredo J, Seruca R, Carneiro F. Familial gastric cancer: genetic susceptibility, pathology, and implications for management. Lancet Oncol. 2015;16(2):e60-70. doi: 10.1016/s1470-2045(14)71016-2. PubMed PMID: 25638682.
44.Yamada M, Fukagawa T, Nakajima T, Asada K, Sekine S, Yamashita S, et al. Hereditary diffuse gastric cancer in a Japanese family with a large deletion involving *CDH1*. Gastric Cancer. 2014;17(4):750-6. Epub 20130915. doi: 10.1007/s10120-013-0298-y. PubMed PMID: 24037103.
45.Paliou C, Guckelberger P, Schöpflin R, Heinrich V, Esposito A, Chiariello AM, et al. Preformed chromatin topology assists transcriptional robustness of Shh during limb development. Proc Natl Acad Sci U S A. 2019;116(25):12390-9. Epub 2019/05/31. doi: 10.1073/pnas.1900672116. PubMed PMID: 31147463; PubMed Central PMCID: PMC6589666.
46.Gloerich M, Bianchini JM, Siemers KA, Cohen DJ, Nelson WJ. Cell division orientation is coupled to cell-cell adhesion by the E-cadherin/LGN complex. Nat Commun. 2017;8:13996. Epub 2017/01/04. doi: 10.1038/ncomms13996. PubMed PMID: 28045117; PubMed Central PMCID: PMC5216124.
47.Wang X, Dong B, Zhang K, Ji Z, Cheng C, Zhao H, et al. E-cadherin bridges cell polarity and spindle orientation to ensure prostate epithelial integrity and prevent carcinogenesis in vivo. PLoS Genet. 2018;14(8):e1007609. Epub 2018/08/18. doi: 10.1371/journal.pgen.1007609. PubMed PMID: 30118484; PubMed Central PMCID: PMC6115016.
48.Yan HB, Wang XF, Zhang Q, Tang ZQ, Jiang YH, Fan HZ, et al. Reduced expression of the chromatin remodeling gene ARID1A enhances gastric cancer cell migration and invasion via downregulation of E-cadherin transcription. Carcinogenesis. 2014;35(4):867-76. Epub 2013/12/03. doi: 10.1093/carcin/bgt398. PubMed PMID: 24293408.
49.Sánchez-Tilló E, Lázaro A, Torrent R, Cuatrecasas M, Vaquero EC, Castells A, et al. ZEB1 represses E-cadherin and induces an EMT by recruiting the SWI/SNF chromatin-remodeling protein BRG1. Oncogene. 2010;29(24):3490-500. Epub 2010/04/27. doi: 10.1038/onc.2010.102. PubMed PMID: 20418909.
50.Chen A, Beetham H, Black MA, Priya R, Telford BJ, Guest J, et al. E-cadherin loss alters cytoskeletal organization and adhesion in non-malignant breast cells but is insufficient to induce an epithelial-mesenchymal transition. BMC Cancer. 2014;14:552. Epub 2014/08/01. doi: 10.1186/1471-2407-14-552. PubMed PMID: 25079037; PubMed Central PMCID: PMC4131020.
51.Mestre T, Figueiredo J, Ribeiro AS, Paredes J, Seruca R, Sanches JM. Quantification of topological features in cell meshes to explore E-cadherin dysfunction. Sci Rep. 2016;6:25101. Epub 2016/05/07. doi: 10.1038/srep25101. PubMed PMID: 27151223; PubMed Central PMCID: PMC4858654.
52.Elisha Y, Kalchenko V, Kuznetsov Y, Geiger B. Dual role of E-cadherin in the regulation of invasive collective migration of mammary carcinoma cells. Sci Rep. 2018;8(1):4986. Epub 2018/03/23. doi: 10.1038/s41598-018-22940-3. PubMed PMID: 29563585; PubMed Central PMCID: PMC5862898.
53.Christgen M, Bartels S, van Luttikhuizen JL, Bublitz J, Rieger LU, Christgen H, et al. E-cadherin to P-cadherin switching in lobular breast cancer with tubular elements. Mod Pathol. 2020;33(12):2483-98. Epub 20200622. doi: 10.1038/s41379-020-0591-3. PubMed PMID: 32572153; PubMed Central PMCID: PMC7685979.
54.Lui AJ, Geanes ES, Ogony J, Behbod F, Marquess J, Valdez K, et al. IFITM1 suppression blocks proliferation and invasion of aromatase inhibitor-resistant breast cancer in vivo by JAK/STAT-mediated induction of p21. Cancer Lett. 2017;399:29-43. Epub 2017/04/16. doi: 10.1016/j.canlet.2017.04.005. PubMed PMID: 28411130; PubMed Central PMCID: PMC5530765.
55.Borg D, Hedner C, Gaber A, Nodin B, Fristedt R, Jirström K, et al. Expression of IFITM1 as a prognostic biomarker in resected gastric and esophageal adenocarcinoma. Biomark Res. 2016;4:10. Epub 2016/05/18. doi: 10.1186/s40364-016-0064-5. PubMed PMID: 27186374; PubMed Central PMCID: PMC4867989.
56.Lee J, Goh SH, Song N, Hwang JA, Nam S, Choi IJ, et al. Overexpression of IFITM1 has clinicopathologic effects on gastric cancer and is regulated by an epigenetic mechanism. Am J Pathol. 2012;181(1):43-52. Epub 20120518. doi: 10.1016/j.ajpath.2012.03.027. PubMed PMID: 22609115.
57.Yang Y, Lee JH, Kim KY, Song HK, Kim JK, Yoon SR, et al. The interferon-inducible 9-27 gene modulates the susceptibility to natural killer cells and the invasiveness of gastric cancer cells. Cancer Lett. 2005;221(2):191-200. doi: 10.1016/j.canlet.2004.08.022. PubMed PMID: 15808405.
58.Hu J, Wang S, Zhao Y, Guo Q, Zhang D, Chen J, et al. Mechanism and biological significance of the overexpression of IFITM3 in gastric cancer. Oncol Rep. 2014;32(6):2648-56. Epub 20141001. doi: 10.3892/or.2014.3522. PubMed PMID: 25270246.
59.Rajapaksa US, Jin C, Dong T. Malignancy and IFITM3: Friend or Foe? Front Oncol. 2020;10:593245. Epub 20201208. doi: 10.3389/fonc.2020.593245. PubMed PMID: 33364194; PubMed Central PMCID: PMC7753217.
60.Min J, Feng Q, Liao W, Liang Y, Gong C, Li E, et al. IFITM3 promotes hepatocellular carcinoma invasion and metastasis by regulating MMP9 through p38/MAPK signaling. FEBS Open Bio. 2018;8(8):1299-311. Epub 20180628. doi: 10.1002/2211-5463.12479. PubMed PMID: 30087833; PubMed Central PMCID: PMC6070650.
61.Liu X, Chen L, Fan Y, Hong Y, Yang X, Li Y, et al. IFITM3 promotes bone metastasis of prostate cancer cells by mediating activation of the TGF-β signaling pathway. Cell Death Dis. 2019;10(7):517. Epub 20190704. doi: 10.1038/s41419-019-1750-7. PubMed PMID: 31273201; PubMed Central PMCID: PMC6609682.

62.Wang Y, Wu C, Zhang C, Li Z, Zhu T, Chen J, et al. TGF-β-induced STAT3 overexpression promotes human head and neck squamous cell carcinoma invasion and metastasis through malat1/miR-30a interactions. Cancer Lett. 2018;436:52-62. Epub 20180814. doi: 10.1016/j.canlet.2018.08.009. PubMed PMID: 30118844.
63.Wang H, Tang F, Bian E, Zhang Y, Ji X, Yang Z, et al. IFITM3/STAT3 axis promotes glioma cells invasion and is modulated by TGF-β. Mol Biol Rep. 2020;47(1):433-41. Epub 20191021. doi: 10.1007/s11033-019-05146-2. PubMed PMID: 31637620.
64.Lee J, Robinson ME, Ma N, Artadji D, Ahmed MA, Xiao G, et al. IFITM3 functions as a PIP3 scaffold to amplify PI3K signalling in B cells. Nature. 2020;588(7838):491-7. Epub 20201104. doi: 10.1038/s41586-020-2884-6. PubMed PMID: 33149299; PubMed Central PMCID: PMC8087162.
65.Li D, Peng Z, Tang H, Wei P, Kong X, Yan D, et al. KLF4-mediated negative regulation of IFITM3 expression plays a critical role in colon cancer pathogenesis. Clin Cancer Res. 2011;17(11):3558-68. Epub 20110429. doi: 10.1158/1078-0432.Ccr-10-2729. PubMed PMID: 21531817; PubMed Central PMCID: PMC3107880.
66.Daniel-Carmi V, Makovitzki-Avraham E, Reuven EM, Goldstein I, Zilkha N, Rotter V, et al. The human 1-8D gene (IFITM2) is a novel p53 independent pro-apoptotic gene. Int J Cancer. 2009;125(12):2810-9. doi: 10.1002/ijc.24669. PubMed PMID: 19544527.
67.Bonni A, Brunet A, West AE, Datta SR, Takasu MA, Greenberg ME. Cell survival promoted by the Ras-MAPK signaling pathway by transcription-dependent and -independent mechanisms. Science. 1999;286(5443):1358-62. doi: 10.1126/science.286.5443.1358. PubMed PMID: 10558990.
68.Hou F, Shi DB, Chen YQ, Gao P. Human Epidermal Growth Factor Receptor-2 Promotes Invasion and Metastasis in Gastric Cancer by Activating Mitogen-activated Protein Kinase Signaling. Appl Immunohistochem Mol Morphol. 2019;27(7):529-34. doi: 10.1097/pai.0000000000000672. PubMed PMID: 29734245.
69.Ohtsuka J, Oshima H, Ezawa I, Abe R, Oshima M, Ohki R. Functional loss of p53 cooperates with the in vivo microenvironment to promote malignant progression of gastric cancers. Sci Rep. 2018;8(1):2291. Epub 20180202. doi: 10.1038/s41598-018-20572-1. PubMed PMID: 29396430; PubMed Central PMCID: PMC5797237.
70.Busuttil RA, Zapparoli GV, Haupt S, Fennell C, Wong SQ, Pang JM, et al. Role of p53 in the progression of gastric cancer. Oncotarget. 2014;5(23):12016-26. doi: 10.18632/oncotarget.2434. PubMed PMID: 25427447; PubMed Central PMCID: PMC4322971.
71.Kastan MB, Onyekwere O, Sidransky D, Vogelstein B, Craig RW. Participation of p53 protein in the cellular response to DNA damage. Cancer Res. 1991;51(23 Pt 1):6304-11. PubMed PMID: 1933891.

Supplementary table 1. Genes associated with gastrointestinal syndromes

Gene name	Reference sequence
A2ML1	NM_144670.6
ACD	NM_001082486.2
ADM	NM_001124.3
AIP	NM_003977.4
AKT1	NM_001014431.2
ALK	NM_004304.5
ANKRD26	NM_014915.2
APC	NM_001354896.2
ARMC5	NM_001288767.2
ASXL1	NM_015338.6
ATM	NM_000051.3
ATR	NM_001184.4
AXIN1	NM_003502.4
AXIN2	NM_004655.4
BAP1	NM_004656.4
BARD1	NM_000465.4
BLM	NM_000057.4
BMPR1A	NM_004329.2
BRAF	NM_001354609.2
BRCA1	NM_007300.4
BRCA2	NM_000059.3
BRIP1	NM_032043.3
BUB1	NM_004336.5
BUB1B	NM_001211.5
BUB3	NM_004725.4
CBL	NM_005188.4
CDC73	NM_024529.4
CDH1	NM_004360.5
CDK4	NM_000075.4
CDKN1A	NM_001291549.2
CDKN1B	NM_004064.4
CDKN1C	NM_000076.2
CDKN2A	NM_001195132.1
CDKN2B	NM_004936.4
CDKN2C	NM_001262.2
CEBPA	NM_001287424.2
CHEK1	NM_001330427.1
CHEK2	NM_001005735.2
CREBBP	NM_004380.3
CTC1	NM_025099.6
CTNNA1	NM_001323982.2
CTR9	NM_014633.5
CYLD	NM_015247.2
DDB2	NM_000107.2
DDX11	NM_001257144.2
DICER1	NM_001271282.3
DIS3L2	NM_152383.4
DKC1	NM_001363.5
DNAJC21	NM_194283.4
EGFR	NM_005228.5
ELANE	NM_001972.4
ENG	NM_001114753.2
EPCAM	NM_002354.3

Gene name	Reference sequence
ERCC1	NM_001369408.1
ERCC2	NM_000400.3
ERCC3	NM_000122.2
ERCC4	NM_005236.3
ERCC5	NM_000123.3
ERCC6	NM_000124.4
ESR2	NM_001437.2
ETV6	NM_001987.5
EXO1	NM_006027.4
EXT1	NM_000127.2
EXT2	NM_000401.3
EZH2	NM_004456.5
FAN1	NM_014967.5
FANCA	NM_000135.4
FANCB	NM_001018113.3
FANCC	NM_000136.3
FANCD2	NM_033084.5
FANCE	NM_021922.3
FANCF	NM_022725.4
FANCG	NM_004629.1
FANCI	NM_001113378.1
FANCL	NM_001114636.1
FANCM	NM_020937.4
FAS	NM_000043.6
FBXW7	NM_001349798.2
FH	NM_000143.3
FLCN	NM_001353229.2
FOXO3	NM_001455.4
G6PC3	NM_138387.3
GALNT12	NM_024642.5
GATA2	NM_001145661.2
GDNF	NM_001190468.1
GFI1	NM_001127215.2
GPC3	NM_001164617.2
GREM1	NM_001368719.1
GRHL2	NM_024915.4
HABP2	NM_004132.5
HAX1	NM_006118.4
HNF1A	NM_001306179.2
HOXB13	NM_006361.5
IPMK	NM_152230.5
KIF1B	NM_001365951.2
KIT	NM_000222.2
KLLN	NM_001126049.1
KRAS	NM_001369786.1
LIG4	NM_001352604.1
LZTR1	NM_006767.4
MAP2K1	NM_002755.3
MAP2K2	NM_030662.3
MAX	NM_002382.5
MDH2	NM_005918.4
MEN1	NM_001370251.1
MET	NM_001127500.3
MITF	NM_001354604.2
MLH1	NM_000249.3
MLH3	NM_001040108.1
MPL	NM_005373.3

Supplementary table 1. Genes associated with gastrointestinal syndromes

Gene name	Reference sequence
MRE11A	NM_005591.3
MSH2	NM_000251.3
MSH3	NM_002439.5
MSH6	NM_000179.2
MUC5B	NM_002458.3
MUTYH	NM_001128425.1
NBN	NM_002485.4
NF1	NM_001042492.3
NF2	NM_000268.3
NHP2	NM_017838.3
NOP10	NM_018648.3
NPM1	NM_001355006.1
NRAS	NM_002524.5
NSD1	NM_022455.4
NTHL1	NM_002528.7
OGG1	NM_016821.2
PALB2	NM_024675.4
PARK2	NM_004562.3
PARN	NM_002582.4
PAX5	NM_016734.3
PDGFB	NM_002608.4
PDGFRA	NM_001347828.2
PHOX2B	NM_003924.4
PIK3CA	NM_006218.4
PMS1	NM_000534.4
PMS2	NM_001322014.2
PMS2CL	NR_002217.1
POLD1	NM_001308632.1
POLD2	NM_006230.3
POLD3	NM_006591.3
POLD4	NM_021173.5
POLE	NM_006231.4
POLE2	NM_002692.4
POLE3	NM_001278255.1
POLE4	NM_019896.4
POLH	NM_006502.3
POT1	NM_015450.3
POU6F2	NM_001370959.1
PPM1D	NM_003620.4
PRF1	NM_001083116.3
PRKAR1A	NM_001276289.1
PRSS1	NM_002769.5
PTCH1	NM_000264.5
PTCH2	NM_003738.5
PTEN	NM_001304717.5
PTPN11	NM_001330437.1
PTPN12	NM_002835.4
PTPRJ	NM_002843.4
RAD50	NM_005732.4
RAD51C	NM_058216.3
RAD51D	NM_001142571.2
RAF1	NM_001354689.3
RB1	NM_000321.2
RECQL	NM_002907.4
RECQL4	NM_004260.3
REST	NM_001193508.1
RET	NM_020975.6
RHBDF2	NM_024599.5

Gene name	Reference sequence
RINT1	NM_021930.6
RIT1	NM_001256821.2
RNF43	NM_001305544.2
RPL11	NM_000975.5
RPL35A	NM_000996.4
RPL5	NM_000969.5
RPS10	NM_001014.5
RPS17	NM_001021.6
RPS19	NM_001321485.2.2
RPS20	NM_001146227.2
RPS24	NM_001142285.2
RPS26	NM_001029.5
RPS7	NM_001011.4
RTEL1	NM_001283009.2
RUNX1	NM_001754.4
SAMD9	NM_001193307.1
SAMD9L	NM_001303496.3
SBDS	NM_016038.4
SDHA	NM_004168.4
SDHAF2	NM_017841.2
SDHB	NM_003000.3
SDHC	NM_003001.3
SDHD	NM_003002.4
SEC23B	NM_001172745.3
SEMA4A	NM_001193300.2
SFTPA1	NM_001093770.3
SFTPA2	NM_001320814.1
SH2B3	NM_005475.3
SHOC2	NM_001324336.1
SLC26A3	NM_000111.3
SLX4	NM_032444.4
SMAD4	NM_005359.6
SMAD9	NM_001127217.2
SMARCA4	NM_001128849.2
SMARCB1	NM_001362877.2
SMARCE1	NM_003079.5
SOS1	NM_005633.3
SPINK1	NM_001354966.1
SPRED1	NM_152594.3
SQSTM1	NM_003900.5
SRC	NM_005417.4
STK11	NM_000455.5
SUFU	NM_016169.3
TERC	NR_001566.1
TERF2IP	NM_018975.4
TERT	NM_198253.3
TINF2	NM_001099274.3
TMEM127	NM_001193304.3
TNFRSF11A	NM_003839.4
TP53	NM_000546.5
TRIP13	NM_004237.4
TSC1	NM_000368.5
TSC2	NM_000548.5
USB1	NM_024598.4
VHL	NM_000551.3
WAS	NM_000377.3
WRAP53	NM_001143990.1
WRN	NM_000553.6
WT1	NM_024426.6
XPA	NM_000380.3
XPC	NM_004628.4

Supplementary table 2. Primers used for 4C-seq

Primer	Sequence
4C_prom_P5_N502	AATGATACGGCGACCACCGAGATCTACACCTCTCTATTCGTCGGCAGCGTCAGATGTGTATAAGAGACAGGTGAACCCTCAGCCAATCAG
4C_prom_P5_N504	AATGATACGGCGACCACCGAGATCTACACAGAGTAGATCGTCGGCAGCGTCAGATGTGTATAAGAGACAGGTGAACCCTCAGCCAATCAG
4C_prom_P5_N503	AATGATACGGCGACCACCGAGATCTACACTATCCTCTTCGTCGGCAGCGTCAGATGTGTATAAGAGACAGGTGAACCCTCAGCCAATCAG
4C_prom_P5_N501	AATGATACGGCGACCACCGAGATCTACACTAGATCGCTCGTCGGCAGCGTCAGATGTGTATAAGAGACAGGTGAACCCTCAGCCAATCAG
4C_prom_P7_N705	CAAGCAGAAGACGGCATACGAGATAGGAGTCCGTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGGATCCCAGGTCTTAGTGAGCCA
4C_prom_P7_N707	CAAGCAGAAGACGGCATACGAGATGTAGAGAGGTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGGATCCCAGGTCTTAGTGAGCCA
4C_prom_P7_N706	CAAGCAGAAGACGGCATACGAGATCATGCCTAGTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGGATCCCAGGTCTTAGTGAGCCA
4C_prom_P7_N708	CAAGCAGAAGACGGCATACGAGATCCTCTCTGGTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGGATCCCAGGTCTTAGTGAGCCA

Supplementary table 3. sgRNAs used for CRISPR-Cas9

sgRNA	Sequence
CDH1 intron 7	TGTGTATTAGCTCAATCCCG
CDH1 3UTR	AGATGGCCTCAGTCAAAACG
TANGO6 5UTR	GCTTCTGAAATTGTGCGCGT
TANGO6 intron 13	CAGCACTTGACACATAGTCC

Supplementary table 4. Primers used for CRISPR-Cas9

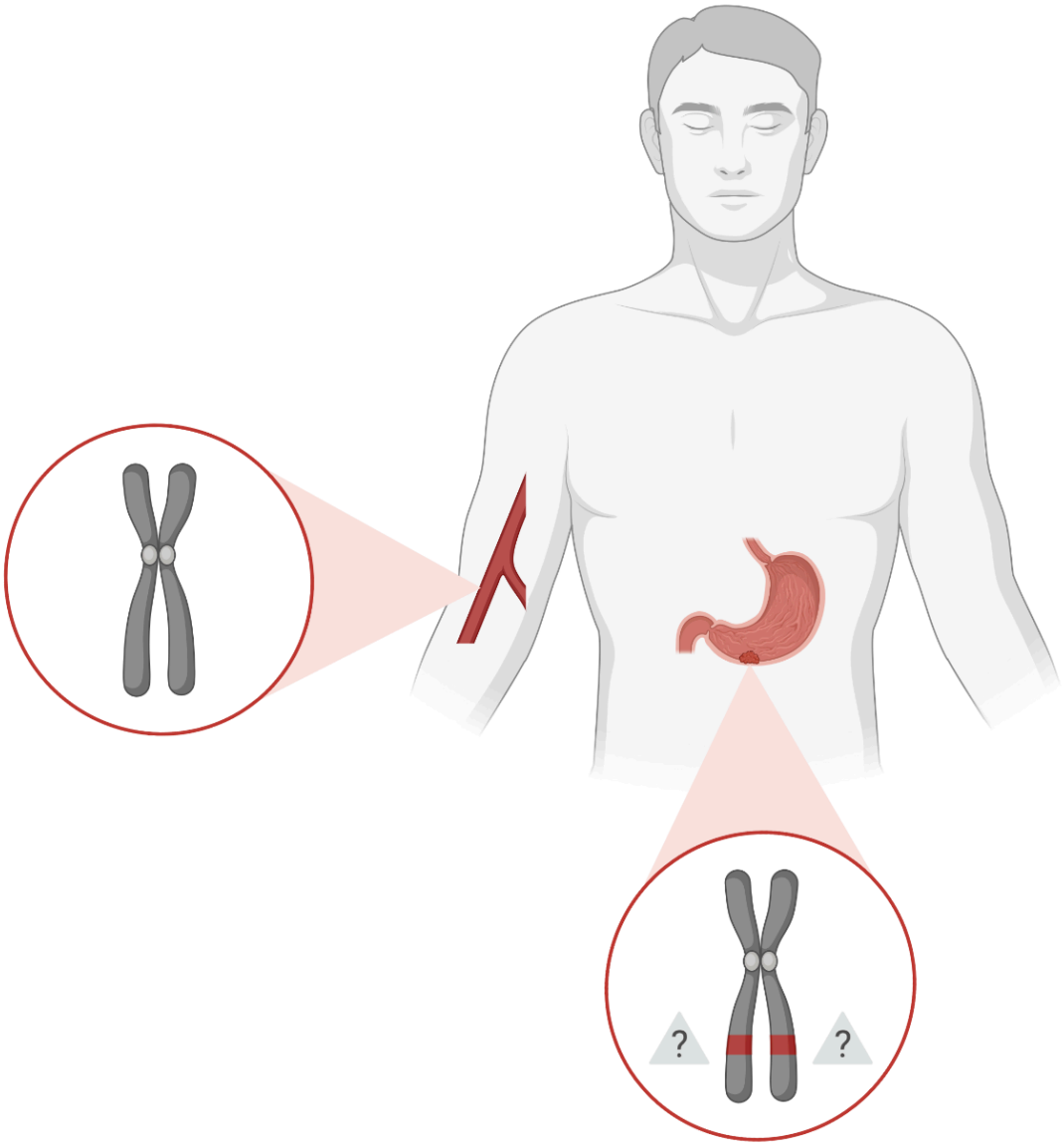
Primer	Sequence	Orientation	Used for
Cas9	GGGCCTATTTCCCATGATTCCTT	Forward	Colony PCR
Cas9	GACTCGGTGCCACTTTTTCAA	Reverse	Colony PCR
CDH1 del 5'	GGGCAGAATTGGATTAAGCA	Forward	Genotyping
CDH1 del 3'	TGGCCTGTACTCACTTGCTG	Reverse	Genotyping
CDH1-TANGO6 del 5'	GGGCAGAATTGGATTAAGCA	Forward	Genotyping
CDH1-TANGO6 del 3'	AACATGGTGCCCGTCTCTAC	Reverse	Genotyping



Supplementary Figure 1. *Evolution of familiar and individual clinical criteria for HDGC*

Chapter 4

Somatic regulatory elements



Original paper 3

Genetic and Epigenetic Alterations of *CDH1* Regulatory Regions in Hereditary and Sporadic Gastric Cancer

Gianluca Tedaldi[1]*, Chiara Molinari[1], **Celina São José**[2,3,4], Rita Barbosa-Matos[2,3,5], Ana André[2,3], Rita Danesi[6], Valentina Arcangeli[6], Mila Ravegnani[6], Luca Saragoni[7], Paolo Morgagni[8], Francesca Rebuzzi[1], Matteo Canale[1], Sara Pignatta[1], Elisa Ferracci[9], Giovanni Martinelli[10], Guglielmina Nadia Ranzani[11], Carla Oliveira[2,3,12]*, Daniele Calistri[1]†, Paola Ulivi[1]†

† These authors contributed equally to this work.

[1] Biosciences Laboratory, IRCCS Istituto Romagnolo per lo Studio dei Tumori (IRST) "Dino Amadori", 47014 Meldola, Italy

[2] Institute for Research and Innovation in Health, University of Porto, Porto, Portugal

[3] Institute of Molecular Pathology and Immunology, University of Porto, Porto, Portugal

[4] Doctoral Programme in Biomedicine, Faculty of Medicine, University of Porto, Porto, Portugal

[5] Doctoral Programme on Cellular and Molecular Biotechnology Applied to Health Sciences (Biotech Health), Institute of Biomedical Sciences Abel Salazar, University of Porto, Porto, Portugal

[6] Romagna Cancer Registry, IRCCS Istituto Romagnolo per lo Studio dei Tumori (IRST) "Dino Amadori", Meldola, Italy

[7] Pathology Unit, Morgagni-Pierantoni Hospital, Forlì, Italy

[8] Department of General Surgery, Morgagni-Pierantoni Hospital, Forlì, Italy

[9] Laboratory of Stem Cells and Cancer Genomics, University of Trento, Trento, Italy

[10] Department of Medical Oncology, IRCCS Istituto Romagnolo per lo Studio dei Tumori (IRST) "Dino Amadori", Meldola, Italy

[11] Department of Biology and Biotechnology, University of Pavia, Pavia, Italy

[12] Faculty of Medicine of the University of Porto, Porto, Portugal

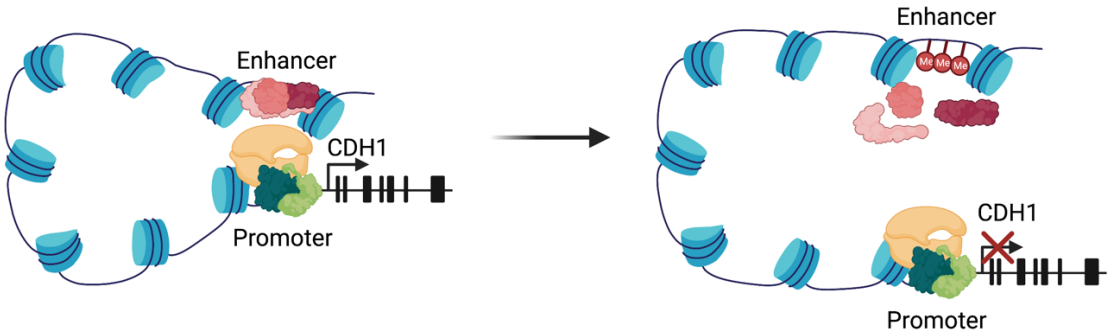
Pharmaceuticals (Basel, Switzerland), 14(5), 457 (appendix C).

Author statement

Celina São José: Investigation, methodology, visualization, review and editing.

Investigate the role of genetic variants and DNA methylation in *CDH1* promoter and enhancers for E-cadherin expression

E-cadherin is a major player in epithelial cancers and invasion. Herein, we explored the role of genetic variants and DNA methylation in *CDH1* promoter and its enhancers. Thus, we selected potential *CDH1* intronic enhancers and sequenced them alongside the promoter in several gastric cancer (GC) cell lines for genetic variants and *CDH1* mRNA expression. We further, bisulfite sequenced the methylation status of several gastric cancer cell lines, 13 sporadic GCs, 20 HDGC *CDH1*-negative patients and 6 healthy controls. *CDH1* regulatory elements display differential methylation patterns in GC cell lines expressing and non-expressing *CDH1* and in patients and controls. Thus, genetic variants and DNA methylation patterns in *CDH1* promoter and enhancers correlate with *CDH1* mRNA expression, pinpointing 7 potential *CDH1* enhancers.



Abstract

E-cadherin is a key player in gastric cancer (GC) and germline alterations of *CDH1*, its encoding gene, are responsible for Hereditary Diffuse Gastric Cancer (HDGC) syndrome. This study aimed at elucidating the role of genetic variants and DNA methylation of *CDH1* promoter and enhancers in the regulation of gene expression. For this purpose, we analyzed genetic variants of the *CDH1* gene through Next-Generation Sequencing (NGS) in a series of GC cell lines (NCI-N87, KATO-III, SNU-1, SNU-5, GK2, AKG, KKP) and the corresponding *CDH1* expression levels. By bisulfite genomic sequencing, we analyzed the methylation status of *CDH1* regulatory regions in 8 GC cell lines, in a series of 13 sporadic GC tissues and in a group of 20 HDGC *CDH1*-negative patients and 6 healthy controls. The NGS analysis on *CDH1* coding and regulatory regions detected genetic alterations in 3 out of 5 GC cell lines lacking functional E-cadherin. *CDH1* regulatory regions showed different methylation patterns in patients and controls, GC cell lines and GC tissues, expressing different E-cadherin levels. Our results showed that alterations in terms of genetic variants and DNA methylation patterns of both promoter and enhancers are associated with *CDH1* expression levels and have a role in its regulation.

Introduction

In 2020, gastric cancer (GC) ranked fifth for incidence and fourth for mortality, worldwide [1]. Although GC is commonly sporadic, about 10–20% of cases show familial clustering and 1–3% of cases can be considered hereditary [2]. Sporadic GC occurs at an older age, is mainly associated with intestinal histotype (IGC) [3] and is related to environmental risk factors, such as microbial infection and diet [4]. Hereditary GC is characterized by early-onset, is mainly of the diffuse histotype (DGC) and is associated with alterations in different predisposition genes [5]. Nevertheless, with the exception of the inherited familial syndromes, the observed familial clustering of cancer is not well explained and may occur due to an inherited genetic susceptibility, together with shared lifestyle and environmental factors [6].

The main gene involved in hereditary GC is *CDH1*, encoding the E-cadherin protein [7], whose germline mutations are responsible for Hereditary Diffuse Gastric Cancer (HDGC) [8], a genetic syndrome linked to a lifetime GC risk of 42–70% in male carriers and 33–56% in female carriers, who also have a 39–55% risk of lobular breast cancer (LBC) [9,10]. The oncogenesis in HDGC syndrome follows the “two-hit model” [11], in which the presence of a germline mutation of the *CDH1* gene (first hit) is followed by the inactivation of the other allele (second hit) by mutations or deletions (loss of heterozygosity-LOH). However, it has been shown that *CDH1* promoter hypermethylation represents the main cause of inactivation of the wild-type allele in HDGC tumors [12–14].

The *CDH1* gene is a driver of carcinogenesis also in sporadic GC of both diffuse and intestinal histotypes [15,16] and *CDH1* epigenetic alterations, such as promoter hypermethylation and histone modifications, have also been detected in most sporadic GCs as a cause of E-cadherin impairment [15,17–20].

Given the importance of *CDH1* in the oncogenesis of hereditary and sporadic GC, different studies were carried out to understand the regulatory mechanisms of *CDH1* gene expression. In particular, the studies focused on the 65kb-region of *CDH1* intron 2, which is characterized by the presence of cis-regulatory sequences necessary for transcription and, therefore, for the regulation of gene expression [21–23].

Moreover, not only a large body of data supports the notion that DNA methylation is crucial in regulating gene expression, but recent findings also point to a link between this mechanism and cancer predisposition. Accordingly, studies aimed at linking genomic regions associated with cancer susceptibility and expression of genes involved in cancer development have already shown that promoters’ and enhancers’ sequence polymorphisms and methylation status can contribute to cancer risk [24,25].

The present study aimed to perform a molecular characterization of genetic and epigenetic alterations of *CDH1* promoter and enhancer sequences to deeply

understand the mechanisms that regulate gene expression and to provide the rationale for investigating *CDH1* methylation in *CDH1*-negative patients with suspected genetic predisposition to GC.

Materials and Methods

Cell Lines

The study was performed on four commercial GC cell lines (SNU-1 ATCC Cat# CRL-5971, RRID:CVCL_0099, SNU-5 ATCC Cat# CRL-5973, RRID:CVCL_0078, KATO-III ATCC Cat# HTB-103, RRID:CVCL_0371, NCI-N87 ATCC Cat# CRL-5822, RRID:CVCL_1603) purchased from ATCC (American Type Culture Collection, Manassas, VA, USA), three GC cell lines (GK2, AKG, KKP) established and characterized at IRST laboratory, and one GC cell line (MKN-74) provided by the team of C. Oliveira, a co-author from the Institute of Molecular Pathology and Immunology of University of Porto (IPATIMUP, Portugal). The characteristics of the GC cell lines used in the present work are summarized in the Supplementary Table S4.

All the cell lines were maintained as a monolayer, incubated at 37 ◦C with 5% CO2 and subcultured weekly. For SNU-1, NCI-N87, and MKN-74, the culture medium was composed of RPMI-1640 Medium (ATCC) supplemented with 10% fetal bovine serum (FBS) (Euroclone, Milan, Italy). For SNU-5 and KATO-III, the culture medium was Iscove's Modified Dulbecco's Medium (ATCC) supplemented with 20% FBS. For AKG, GK2 and KKP, the culture medium was composed of DMEM/Ham's F12 (1:1) supplemented with FBS (10%), glutamine (2 mM), non-essential amino acids (1%) and insulin (10 mg/mL) (Sigma-Aldrich, St. Louis, MO, USA). Cell lines were routinely tested for Mycoplasma using MycoAlert Mycoplasma Detection Kit (Lonza, Basel, Switzerland).

Blood Samples

The patients included in the study were selected by the Genetic Counseling service of the Area Vasta Romagna from 2010 to 2018 because they had a family and personal history of DGC and/or LBC.

In particular, the selection of patients was performed according to the first four criteria established by the International Gastric Cancer Linkage Consortium (IGCLC) in the 2015 HDGC guidelines, which must be valid among the first- and second-degree relatives: 2 GC cases regardless of age with at least 1 confirmed DGC (I); 1 case of DGC < 40 years (II); personal or family history of DGC and LBC with 1 diagnosed < 50 years (III); bilateral LBC or family history of 2 or more cases of LBC < 50 years (IV) [36].

All 20 patients underwent molecular testing for point mutations and copy number variations (CNVs) of the *CDH1* gene using Next-Generation Sequencing (NGS) and Multiplex Ligation-dependent Probe Amplification (MLPA), as reported in our previous work [37]. Of the 20 patients, 15 had developed DGC (range 22–39 years) and 5 had LBC (range 35–48 years). Patients' characteristics are reported in Supplementary Table S5.

After the subscription of the informed consent, we collected the peripheral blood of these 20 patients and of 6 healthy controls.

Gastric Tissue Samples

Thirteen patients with sporadic GC submitted to surgical resection and classified by an expert pathologist, according to Lauren's classification [3], were included in this study. Surgical samples were immediately cryopreserved after resection and fresh-frozen (FF) tumor tissue and matched normal gastric epithelium samples were stored as previously described [33]. The patients' characteristics are summarized in Supplementary Table S6.

Nucleic Acid Extraction

The DNA was extracted from GC cell lines, peripheral blood samples and FF tumor and normal tissues using the QIAamp DNA Mini Kit (Qiagen, Hilden, Germany), according to manufacturer's instructions. DNA was quantified by Qubit 4 Fluorometer (Thermo Fisher Scientific, Waltham, MA, USA).

Total RNA extraction was performed from GC cell lines and from FF tumor and normal tissues using TRIzol (Invitrogen, Carlsbad, CA, USA) in accordance with manufacturer's instructions. Four micrograms of extracted RNA were treated with DNase and purified by using RNeasy MinElute Cleanup kit (Qiagen, Hilden, Germany), RNA was quantified by spectrophotometer Nanodrop-ND-1000 (Thermo Fisher Scientific).

Next-Generation Sequencing

The DNA libraries for the NGS analysis were prepared with Nextera Flex for Enrichment (Illumina, San Diego, CA, USA) and a custom probe panel (Integrated DNA Technologies, Coralville, IA, USA), and run on the Miseq sequencer (Illumina). Results were analyzed with Miseq Reporter software (Illumina) and further processed with a customized bioinformatics pipeline.

Identification of CDH1 Regulatory Regions

The promoter and the enhancers of the *CDH1* gene were identified through the UCSC Genome Browser [38] with the FANTOM5 tool [39,40] in the 2018 version. The enhancers were named from A to G starting from the most upstream one. The coordinates of *CDH1* promoter and enhancers are reported in the Supplementary Table S2.

DNA Methylation Analysis

DNA samples were treated with bisulfite using the Epitect Bisulfite kit (Qiagen, Hilden, Germany). *CDH1* promoter and enhancers were amplified by standard PCR with the Multiplex PCR kit (Qiagen, Hilden, Germany), and the PCR products were sequenced by the BigDye Terminator v.3.1 cycle sequencing kit (Thermo Fisher Scientific) on an ABI-3130 Genetic analyzer (Applied Biosystems, Foster City, CA, USA), with the exception of enhancer C that was excluded from the analysis because of the absence of CpG sites. The primers used for amplification and sequencing were designed on the in silico bisulfite converted sequence and are reported in Supplementary Table S7.

Expression Analysis

The retrotranscription of RNA to cDNA was performed on 1000 ng of total RNA using the SuperScript II Reverse Transcriptase, random primers, dNTP mix and RNaseOUT (Invitrogen, Carlsbad, CA, USA) according to manufacturer's instructions.

The *CDH1* expression analysis was performed by real-time PCR with TaqMan probes for *CDH1* gene as target (Hs01023894_m1, Thermo Fisher Scientific) and B2M gene as endogenous control for normalization (Hs00984230_m1, Thermo Fisher Scientific). The runs were performed on the 7500 Real-Time PCR System (Applied Biosystems).

Expression levels of the target genes were obtained by normalizing the results using the endogenous control B2M. For GC cell lines, the relative expression was quantified using the comparative $2^{-\Delta Ct}$ method. For the gastric tissues, the Fold Change (FC) values in gene expression were calculated using the $2^{-\Delta\Delta Ct}$ method [41,42]. An FC ≤ 0.50 and ≥1.5 were used as cut-off for downregulation and upregulation, respectively.

Results

CDH1 Variant Analysis on GC Cell Lines

The *CDH1* variant analysis was performed by Next-Generation Sequencing (NGS) on seven GC cell lines (NCI-N87, KATO-III, SNU-5, SNU-1, AKG, GK2, KKP).

Results showed the presence of a *CDH1* pathogenic variant in three cell lines (KATO-III, SNU-5 and KKP), whereas in the other four cell lines no *CDH1* variants with relevant significance were detected (**Supplementary Table S1**).

Identification of CDH1 Regulatory Regions

The promoter and the enhancers of the *CDH1* gene were identified through the UCSC Genome Browser. The *CDH1* promoter is located in a CpG island of 1310 bp covering exons 1 and 2 (**Figure 1**). The FANTOM5 tool of the UCSC Genome Browser identified seven enhancers in the *CDH1* locus, one located 4724 bp upstream from the *CDH1* transcription start site (TSS), and six located within intron 2 (**Figure 1**).

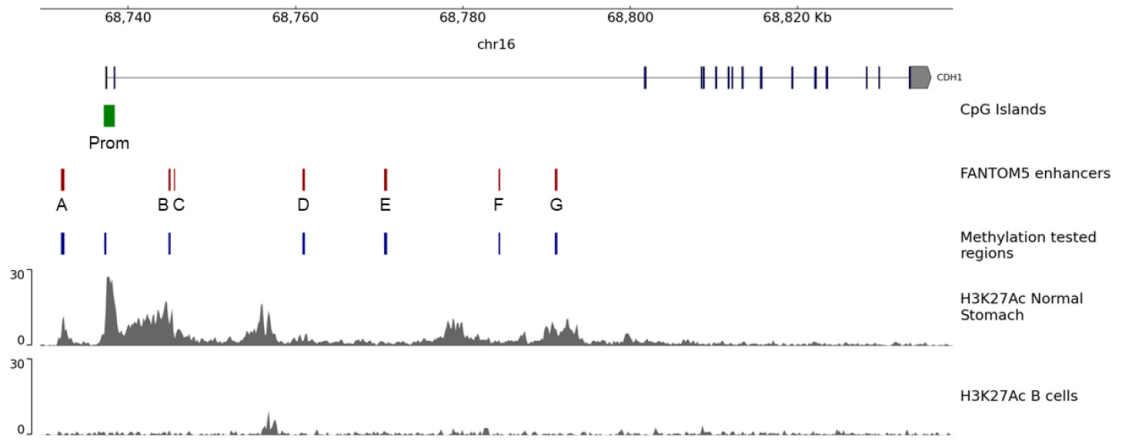


Figure 1. *Scheme of the CDH1 gene (chr16) with the CpG island represented by the green bar and enhancers by red bars. Blue bars represent the regions tested for methylation in the present work. The lower panels represent H3K27Ac, a chromatin mark associated with transcription activation [26], in normal stomach and B cells, respectively. The figure was generated with pyGenomeTracks [27].*

CDH1 Methylation Analysis on DNA from GC Cell Lines, Peripheral Blood and Gastric Tissues

The methylation analysis of *CDH1* promoter and enhancers was performed on eight GC cell lines (NCI-N87, KATO-III, SNU-5, SNU-1, AKG, GK2, KKP and MKN-74), on germline DNA from peripheral blood of 6 healthy individuals and 20 HDGC *CDH1*-negative patients (15 with DGC and 5 with LBC) and on gastric

normal and tumor tissues from 13 sporadic GC patients (10 with IGC, 1 with DGC, 1 with mucinous GC and 1 with mixed GC). We analyzed the methylation status of CpG sites located in the core of the *CDH1* promoter and in six enhancers of *CDH1* gene (**Figure 1 and Supplementary Table S2**).

The methylation analysis revealed different methylation patterns in the eight GC cell lines, with different CpG sites being methylated along promoter and enhancer sequences, in particular enhancers A, B and D (**Figure 2**).

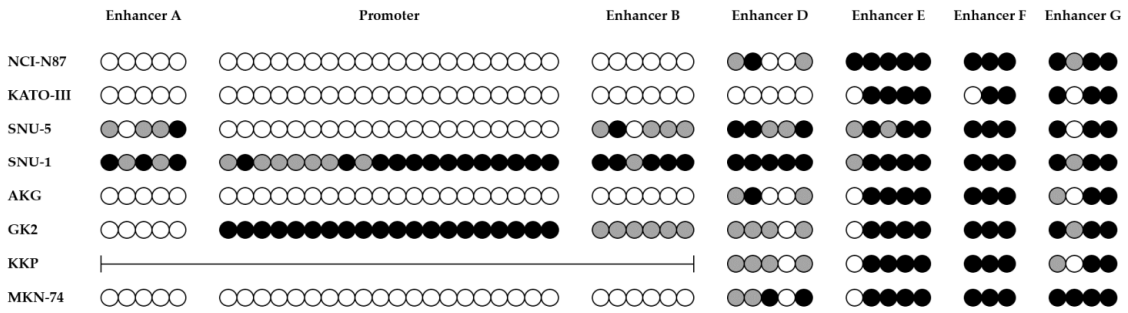


Figure 2. *Methylation status of CDH1 promoter and enhancers of eight GC cell lines expressing different levels of CDH1.*

The methylation analysis performed on the germline DNA derived from blood of 6 healthy controls and of 20 HDGC *CDH1*-negative patients (15 with DGC and 5 with LBC) revealed a strongly conserved methylation pattern along promoter and enhancer sequences, with absence of methylation in the promoter and almost full methylation in enhancer B in all patients and controls (**Figure 3 and Supplementary Figures S1 and S2**).

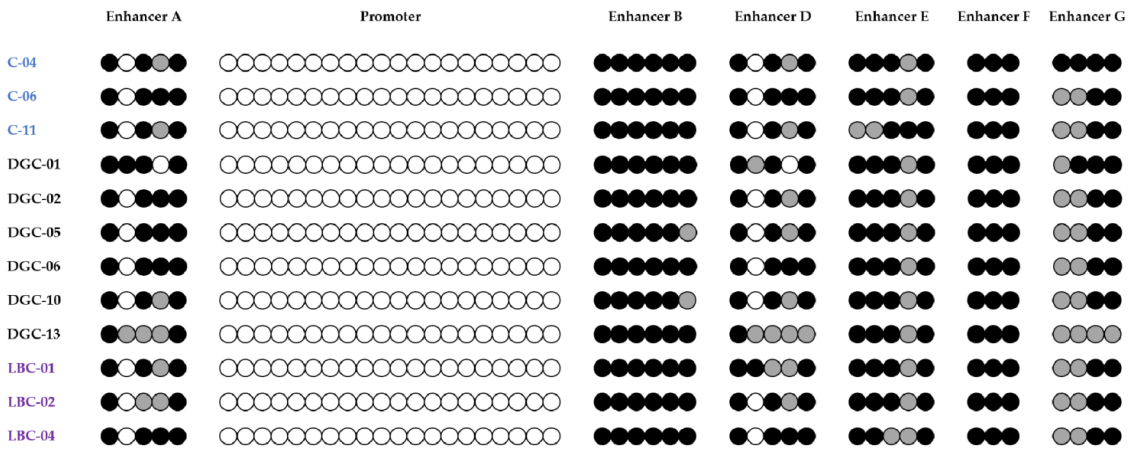


Figure 3. *Methylation analysis of CDH1 promoter and enhancers on a selection of DNA samples from peripheral blood of three healthy individuals (in blue) and nine HDGC CDH1-negative patients (six DGC patients in black and three LBC patients in violet). The complete case series is reported in Supplementary Figures S1 and S2.*

The methylation analysis performed on DNA from gastric normal and tumor tissues of 13 patients with sporadic GC revealed differences in the methylation pattern of normal vs. tumor tissues, with an increased methylation in particular in enhancers A, B and E of tumor tissues (**Figure 4 and Supplementary Figure S3**).

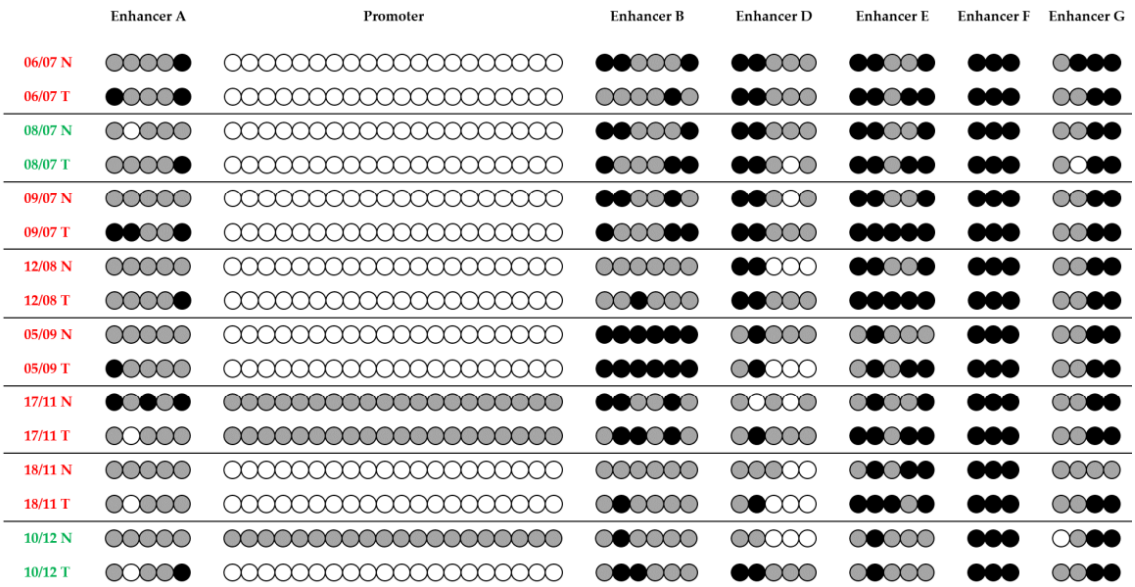


Figure 4. *Methylation analysis of CDH1 promoter and enhancers on DNA samples from gastric normal (N) and tumor (T)*

CDH1 Expression Analysis in GC Cell Lines and Gastric Tissues

The *CDH1* expression analysis was performed by quantitative PCR (qPCR) on seven GC cell lines (NCI-N87, KATO-III, SNU-5, SNU-1, AKG, GK2, KKP) and on gastric normal and tumor tissues of eight sporadic IGC patients.

Results from GC cells revealed a reduced *CDH1* expression in four of them (SNU-1, SNU-5, GK2 and KKP) and a "normal" *CDH1* expression in the other three cell lines (NCI-N87, KATO-III and AKG) (**Supplementary Table S3**).

Results from gastric normal vs. tumor tissues revealed *CDH1* downregulation in six IGC tissues (06/07, 09/07, 12/08, 05/09, 17/11 and 18/11) and *CDH1* upregulation in two IGC tissues (08/07 and 10/12) (**Table 1**).

Table 1. *Results of CDH1 expression on eight gastric normal (N) and tumor (T) tissues.*

Gastric Tissue	$2^{-\Delta Ct}$	Fold Change *CDH1*
06/07 N	2.412	0.056
06/07 T	0.135	
08/07 N	0.157	2.085
08/07 T	0.328	
09/07 N	0.210	0.279
09/07 T	0.059	
12/08 N	2.603	0.053
12/08 T	0.138	
05/09 N	1.338	0.028
05/09 T	0.037	
17/11 N	1.591	0.055
17/11 T	0.087	
18/11 N	2.297	0.042
18/11 T	0.097	
10/12 N	0.056	14.026
10/12 T	0.779	

Discussion

GC is one of the most aggressive cancers, with high mortality rates worldwide. For this reason, the prevention and treatment of GC represent the main topics of interest in the research on this disease. One of the most important players in carcinogenesis of both hereditary and sporadic GC is the *CDH1* gene, encoding E-cadherin, a component of the adherent junctions between epithelial cells with a tumor suppressor activity [28].

In the present study, we evaluated the role of genetic and epigenetic alterations of the *CDH1* gene, in particular the methylation of promoter and enhancers, in the regulation of gene expression and in the predisposition to the development of GC.

Indeed, studies aimed at linking genomic regions associated with cancer susceptibility and expression of genes involved in cancer development have already shown that promoters' and enhancers' sequence polymorphisms and methylation status can contribute to cancer risk [24,25].

This evidence prompted us to test this model in the *CDH1* gene on GC cell lines and different tissues from patients with hereditary and sporadic GC.

Initially, we identified regulatory regions of the *CDH1* gene using bioinformatics tools for the identification of genomic regions with enhancer properties. This analysis revealed that, besides the promoter, there are seven regions in the *CDH1* gene with regulatory functions and six of them are located in intron 2, which has been previously identified as a cis-regulatory element with a pivotal role in controlling the gene expression [22].

Consequently, we explored the landscape of the *CDH1* gene in terms of DNA mutation and methylation on a series of eight GC cell lines, correlating these results to the gene expression. In particular, we identified *CDH1* pathogenic variants in three GC cell lines (KATO-III, SNU-5, KKP). Interestingly, SNU-5 and KKP cell lines also had a significant downregulation of *CDH1* expression. KKP cell lines showed a homozygous deletion of a large region including the promoter (with exons 1 and 2) and enhancers A and B, located upstream and downstream, respectively, while the methylation status of the other enhancers (D–G) did not show significant differences compared to GC cell lines expressing *CDH1*. The SNU-5 cell line showed a *CDH1* variant immediately after exon 5 affecting splicing and partial methylation of enhancers A, B and D (absent in the promoter). Of note, these two cell lines are derived from ascitic effusion of patients with a poorly differentiated GC. On the other hand, the third *CDH1*-mutated GC cell line, KATO-III, despite the presence of a variant in exon 7 with a deleterious effect on splicing [29], showed neither decreased *CDH1* expression nor increased promoter/enhancer methylation, suggesting that this variant does not alter the transcription, but rather the protein function. Indeed, KATO-III cell line is derived

from a DGC with signet ring cells, a typical characteristic of GCs of HDGC patients.

The methylation analysis on the *CDH1* promoter and enhancers revealed increased methylation levels also in SNU-1 and GK2 cell lines, expressing low *CDH1* levels but lacking *CDH1* variants. Intriguingly, GK2 cells show full methylation of the *CDH1* promoter and partial methylation of enhancers B and D (but absent in enhancer A). On the other hand, SNU-1 cells show medium-high levels of methylation in promoter and enhancers A, B and D, compared to GC cell lines expressing *CDH1*. Of note, these two cell lines are both derived from poorly differentiated GCs.

The last three cell lines (NCI-N87, AKG and MKN-74) showed methylation neither in *CDH1* promoter nor in enhancers A and B. Indeed, these GC cell lines have a "normal" expression of *CDH1* gene, as demonstrated in the present work (for NCI-N87 and AKG) or previously reported (for MKN-74) [30,31]. Of note, these three GC cell lines are all derived from IGCs well or moderately differentiated.

Overall, the methylation results obtained in GC cell lines suggest that, besides the promoter, the methylation of CpG sites located in the *CDH1* enhancers correlates with the gene expression (**Table 2**).

Table 2. *Summary of the results obtained on GC cell lines from CDH1 sequencing, methylation and expression analyses.*

Cell Line	*CDH1* Status [1]	*CDH1* Promoter Methylation [2]	*CDH1* Enhancer Methylation [2]				*CDH1* Expression
			A	B	D	E/F/G	
NCI-N87	wt	–	–	–	+	+	normal
KATO-III	mut	–	–	–	–	+	normal
SNU-5	mut	–	+	+	+	+	low
SNU-1	wt	+	+	+	+	+	low
AKG	wt	–	–	–	+	+	normal
GK2	wt	+	–	+	+	+	low
KKP [3]	mut	ND	ND	ND	+	+	low
MKN-74 [4]	wt	–	–	–	+	+	normal

[1]wt: *CDH1* wild-type; mut: *CDH1* pathogenic variant. [2]–: absence of methylation; +: presence of partial or full methylation. [3]The *CDH1* promoter and enhancers A/B are deleted in KKP cell line consequently the methylation analysis for these regions was not detectable (ND). [4]*CDH1* sequencing and expression analyses for MKN-74 cell line have been previously reported [30,31].

These results prompted us to analyze the *CDH1* promoter/enhancer methylation in the germline DNA of HDGC *CDH1*-negative patients, in order to identify a signature predisposing GC development in these patients. Indeed, an increased methylation of *CDH1* promoter was previously identified in an HDGC *CDH1*-negative family [32].

To study the methylation of *CDH1* promoter and enhancers in the germline DNA, we analyzed the methylation on DNA extracted from peripheral blood on a series of 6 healthy individuals and 20 HDGC *CDH1*-negative patients (15 with DGC and 5 with LBC). Unexpectedly, this analysis showed a strongly conserved

methylation pattern in *CDH1* promoter/enhancers both in the healthy controls and the HDGC patients. In particular, enhancer B showed a strong methylation pattern, together with enhancers A and D that were characterized by the presence of CpG sites with partial or full methylation.

E-cadherin, being a component of cell junctions, is not expressed in blood cells but the absence of methylation in the promoter suggests that another mechanism is responsible for *CDH1* downregulation. Our results highlight that, in blood cells, the methylation of *CDH1* enhancers could be responsible for the lack of expression of the gene. On the other hand, the methylation analysis of promoter/enhancers did not prove to be a valid method to identify individuals with a genetic predisposition to GC. Although DNA methylation is generally not heritable, alterations of the mechanisms underlying DNA methylation can, instead, be inherited, for example, in the form of germline mutations in genes involved in the DNA methylation/demethylation processes. However, considering our results, this event seems not to be a frequent cause of HDGC syndrome.

Finally, we tested the methylation levels of *CDH1* promoter/enhancers on a series of sporadic GCs. Results obtained from the analysis of eight IGCs expressing different levels of *CDH1* gene showed the presence of some differences in the methylation of *CDH1* promoter/enhancers between normal and tumor tissues. One of the two tumors (10/12) with upregulation of *CDH1* gene showed a decrease in the promoter methylation compared to the normal tissue, whereas the other (08/07) was characterized by a methylation status substantially comparable with the normal tissue, with some CpG sites of enhancers D and G less methylated in the tumor tissue.

Regarding the six tumors with downregulation of *CDH1* gene, they were characterized by the absence of promoter methylation in all normal and tumor tissues, with the exception of one case (17/11) that showed partial methylation of all the CpG sites analyzed in both normal and tumor tissues. On the other hand, the methylation status of *CDH1* enhancers showed slight differences between normal and tumor tissues, with an increased methylation especially in the CpG sites of enhancers A, B and E in almost all the cases. An additional series of five patients with sporadic GC was studied for *CDH1* promoter, enhancer A and B methylation (Supplementary Figure S3) showing a similar pattern.

This result suggests that *CDH1* enhancer methylation could have a role in gene expression regulation also in GC of the intestinal histotype, in which E-cadherin impairment is a less frequent event [15]. Moreover, in IGC, the *CDH1* gene has been demonstrated to be regulated by other epigenetic mechanisms such as microRNA expression [16,33].

Taking into account all these results, we can conclude that the methylation status of *CDH1* enhancers correlates with the gene expression levels and seems to have an important role in the gene regulation. Particularly in blood cells, where promoter methylation is absent, the presence of a specific pattern of methylation

in the enhancers suggests that it could be a repression mechanism stronger than promoter methylation. This provides the rationale to investigate *CDH1* promoter/enhancer methylation in the gastric tissue of patients with suspected genetic predisposition to GC, since it could be an early event of the carcinogenesis process in HDGC *CDH1*-negative patients.

Moreover, *CDH1* hypermethylation could be a potential novel drug target for developing personalized therapies [34,35]. However, further studies on a larger number of patients with GC of different histotypes are necessary to confirm our results and to deeply understand the role of enhancer methylation in the regulation of *CDH1* gene.

Conclusions

In conclusion, the results we obtained from the methylation analysis of *CDH1* promoter and enhancers demonstrate the existence of different methylation patterns within the regulatory regions of the gene and highlight that the methylation in *CDH1* enhancers could have an important role in the regulation of gene expression.

Supplementary Materials

The following are available online at https://www.mdpi.com/article/10.3390/ph14050457/s1, Figure S1: Methylation analysis of *CDH1* promoter and enhancers from peripheral blood of 6 healthy individuals, Figure S2: Methylation analysis of *CDH1* promoter and enhancers from peripheral blood of 20 HDGC *CDH1*-negative patients, Figure S3: Methylation analysis of *CDH1* promoter and enhancers A and B on DNA samples from gastric normal (N) and tumor (T) tissues from 5 patients with sporadic GC, Table S1: Results of the *CDH1* variant analysis on 7 GC cell lines, Table S2: Genomic coordinates of *CDH1* regulatory elements identified by FANTOM5, Table S3: Results of *CDH1* expression on 7 different GC cell lines, Table S4: Characteristics of the 8 GC cell lines used in the study, Table S5: Clinical characteristics of the 20 HDGC *CDH1*-negative patients whose peripheral blood was analyzed, Table S6: Clinical characteristics of the 13 patients from whose gastric normal and tumor tissues were analyzed, Table S7: Sequences of the primers used for amplification and sequencing in the methylation analysis of *CDH1* regulatory regions on bisulfite-converted DNA.

Acknowledgments

C.S.J. acknowledges the support of the Faculty of Medicine from the University of Porto, specifically by the Doctoral Programme in Biomedicine and the FCT–Fundação para a Ciência e a Tecnologia for her Ph.D. fellowship (ref. SFRH/BD/140796/2018). R.B.-M. acknowledges the support of the Institute of Biomedical Sciences Abel Salazar from the University of Porto, specifically by the Doctoral Programme BiotechHealth and the FCT–Fundação para a Ciência e a Tecnologia for her Ph.D. fellowship (ref. SFRH/BD/145132/2019).

Funding

This research and its authors were funded by IRCCS IRST (G.T., C.M., R.D. V.A., M.R., F.R., M.C., S.P., G.M., D.C., P.U.) and by FEDER-Fundo Europeu de Desenvolvimento Regional funds through the COMPETE 2020–Operacional Programme for Competitiveness and Internationalization (POCI), Portugal 2020, and by Portuguese funds through FCT–Fundação para a Ciência e a Tecnologia/Ministério da Ciência, Tecnologia e Inovação in the framework of the project "Institute for Research and Innovation in Health Sciences" (POCI-01-0145-FEDER-007274) (C.S.J., R.B.-M., A.A., C.O.). This work was also financed by the project NORTE-01-0145-FEDER-000029 (CANCER)-supported by Norte Portugal Regional Programme (NORTE 2020), under the PORTUGAL 2020 Partnership Agreement, through the European Regional Development Fund (ERDF)–project POCI-01-0145-FEDER-016390 (CancelStem) and PTDC/BTM-TEC/30164/2017 (3DChroMe), funded by ERDF, POCI and FCT.

Institutional Review Board Statement: The study was conducted according to the guidelines of the Declaration of Helsinki and approved by the Institutional Ethics Committee of IRST IRCCS (IRSTB069, approved on 20 September 2017).

Informed Consent Statement: Informed consent was obtained from all subjects involved in the study.

Data Availability Statement: The data presented in this study are available on request from the corresponding authors.

Conflict of interest: The authors disclosure no conflicts of interest.

References

1. Sung, H.; Ferlay, J.; Siegel, R.L.; Laversanne, M.; Soerjomataram, I.; Jemal, A.; Bray, F. Global cancer statistics 2020: Globocan estimates of incidence and mortality worldwide for 36 cancers in 185 countries. CA. Cancer J. Clin. 2021. [CrossRef]
2. Oliveira, C.; Pinheiro, H.; Figueiredo, J.; Seruca, R.; Carneiro, F. Familial gastric cancer: Genetic susceptibility, pathology, and implications for management. Lancet Oncol. 2015, 16, e60–e70. [CrossRef]
3. Lauren, P. The two histological main types of gastric carcinoma: Diffuse and so-called intestinal-type carcinoma. An attempt at a histo-clinical classification. Acta Pathol. Microbiol. Scand. 1965, 64, 31–49. [CrossRef]
4. Van Cutsem, E.; Sagaert, X.; Topal, B.; Haustermans, K.; Prenen, H. Gastric cancer. Lancet 2016, 388, 2654–2664. [CrossRef]
5. van der Post, R.S.; Oliveira, C.; Guilford, P.; Carneiro, F. Hereditary gastric cancer: What's new? Update 2013–2018. Fam. Cancer 2019, 18, 363–367. [CrossRef] [PubMed]
6. Choi, Y.J.; Kim, N.; Jang, W.; Seo, B.; Oh, S.; Shin, C.M.; Lee, D.H.; Jung, H.C. Familial Clustering of Gastric Cancer: A Retrospective Study Based on the Number of First-Degree Relatives. Medicine 2016, 95, e3606. [CrossRef]
7. Berx, G.; Staes, K.; van Hengel, J.; Molemans, F.; Bussemakers, M.J.; van Bokhoven, A.; van Roy, F. Cloning and characterization of the human invasion suppressor gene E-cadherin (*CDH1*). Genomics 1995, 26, 281–289. [CrossRef]
8. Blair, V.R.; McLeod, M.; Carneiro, F.; Coit, D.G.; D'Addario, J.L.; van Dieren, J.M.; Harris, K.L.; Hoogerbrugge, N.; Oliveira, C.; van der Post, R.S.; et al. Hereditary diffuse gastric cancer: Updated clinical practice guidelines. Lancet Oncol. 2020, 21, e386–e397. [CrossRef]
9. Hansford, S.; Kaurah, P.; Li-Chang, H.; Woo, M.; Senz, J.; Pinheiro, H.; Schrader, K.A.; Schaeffer, D.F.; Shumansky, K.; Zogopoulos, G.; et al. Hereditary Diffuse Gastric Cancer Syndrome: *CDH1* Mutations and Beyond. JAMA Oncol. 2015, 1, 23–32. [CrossRef]
10. Roberts, M.E.; Ranola, J.M.O.; Marshall, M.L.; Susswein, L.R.; Graceffo, S.; Bohnert, K.; Tsai, G.; Klein, R.T.; Hruska, K.S.; Shirts, B.H. Comparison of *CDH1* Penetrance Estimates in Clinically Ascertained Families vs Families Ascertained for Multiple Gastric Cancers. JAMA Oncol. 2019, 5, 1325–1331. [CrossRef]
11. Guilford, P.; Hopkins, J.; Harraway, J.; McLeod, M.; McLeod, N.; Harawira, P.; Taite, H.; Scoular, R.; Miller, A.; Reeve, A.E. E-cadherin germline mutations in familial gastric cancer. Nature 1998, 392, 402–405. [CrossRef] [PubMed]
12. Grady, W.M.; Willis, J.; Guilford, P.J.; Dunbier, A.K.; Toro, T.T.; Lynch, H.; Wiesner, G.; Ferguson, K.; Eng, C.; Park, J.G.; et al. Methylation of the *CDH1* promoter as the second genetic hit in hereditary diffuse gastric cancer. Nat. Genet. 2000, 26, 16–17. [CrossRef] [PubMed]
13. Machado, J.C.; Oliveira, C.; Carvalho, R.; Soares, P.; Berx, G.; Caldas, C.; Seruca, R.; Carneiro, F.; Sobrinho-Simões, M. E-cadherin gene (*CDH1*) promoter methylation as the second hit in sporadic diffuse gastric carcinoma. Oncogene 2001, 20, 1525–1528. [CrossRef] [PubMed]
14. Graziano, F.; Humar, B.; Guilford, P. The role of the E-cadherin gene (*CDH1*) in diffuse gastric cancer susceptibility: From the laboratory to clinical practice. Ann. Oncol. Off. J. Eur. Soc. Med. Oncol. 2003, 14, 1705–1713. [CrossRef]
15. Corso, G.; Carvalho, J.; Marrelli, D.; Vindigni, C.; Carvalho, B.; Seruca, R.; Roviello, F.; Oliveira, C. Somatic mutations and deletions of the E-cadherin gene predict poor survival of patients with gastric cancer. J. Clin. Oncol. 2013, 31, 868–875. [CrossRef]
16. Carvalho, J.; van Grieken, N.C.; Pereira, P.M.; Sousa, S.; Tijssen, M.; Buffart, T.E.; Diosdado, B.; Grabsch, H.; Santos, M.A.S.; Meijer, G.; et al. Lack of microRNA-101 causes E-cadherin functional deregulation through EZH2 up-regulation in intestinal gastric cancer. J. Pathol. 2012, 228, 31–44. [CrossRef]
17. Tamura, G.; Yin, J.; Wang, S.; Fleisher, A.S.; Zou, T.; Abraham, J.M.; Kong, D.; Smolinski, K.N.; Wilson, K.T.; James, S.P.; et al. E-Cadherin gene promoter hypermethylation in primary human gastric carcinomas. J. Natl. Cancer Inst. 2000, 92, 569–573. [CrossRef]
18. To, K.-F.; Leung, W.K.; Lee, T.-L.; Yu, J.; Tong, J.H.M.; Chan, M.W.Y.; Ng, E.K.W.; Chung, S.C.S.; Sung, J.J.Y. Promoter hypermethylation of tumor-related genes in gastric intestinal metaplasia of patients with and without gastric cancer. Int. J. Cancer 2002, 102, 623–628. [CrossRef]
19. Shokraii, F.; Moharrami, M.; Motamed, N.; Shahhoseini, M.; Totonchi, M.; Ezzatizadeh, V.; Firouzi, J.; Khosravani, P.; Ebrahimi, M.Histone Modification Marks Strongly Regulate *CDH1* Promoter in Prostospheres as A Model of Prostate Cancer Stem Like Cells. Cell J. 2019, 21, 124–134.
20. Oikawa, T.; Otsuka, Y.; Onodera, Y.; Horikawa, M.; Handa, H.; Hashimoto, S.; Suzuki, Y.; Sabe, H. Necessity of p53-binding to the *CDH1* locus for its expression defines two epithelial cell types differing in their integrity. Sci. Rep. 2018, 8, 1595. [CrossRef]
21. Stemmler, M.P.; Hecht, A.; Kinzel, B.; Kemler, R. Analysis of regulatory elements of E-cadherin with reporter gene constructs in transgenic mouse embryos. Dev. Dyn. 2003, 227, 238–245. [CrossRef] [PubMed]
22. Stemmler, M.P.; Hecht, A.; Kemler, R. E-cadherin intron 2 contains cis-regulatory elements essential for gene expression. Development 2005, 132, 965–976. [CrossRef]
23. Oliveira, P.; Sanges, R.; Huntsman, D.; Stupka, E.; Oliveira, C. Characterization of the intronic portion of cadherin superfamily members, common cancer orchestrators. Eur. J. Hum. Genet. 2012, 20, 878–883. [CrossRef]
24. Aran, D.; Hellman, A. DNA methylation of transcriptional enhancers and cancer predisposition. Cell 2013, 154, 11–13. [CrossRef] [PubMed]
25. Aran, D.; Sabato, S.; Hellman, A. DNA methylation of distal regulatory sites characterizes dysregulation of cancer genes. Genome Biol. 2013, 14, R21. [CrossRef] [PubMed]
26. Heintzman, N.D.; Hon, G.C.; Hawkins, R.D.; Kheradpour, P.; Stark, A.; Harp, L.F.; Ye, Z.; Lee, L.K.; Stuart, R.K.; Ching, C.W.; et al. Histone modifications at human enhancers reflect global cell-type-specific gene expression. Nature 2009, 459, 108–112. [CrossRef]
27. Lopez-Delisle, L.; Rabbani, L.; Wolff, J.; Bhardwaj, V.; Backofen, R.; Grüning, B.; Ramírez, F.; Manke, T. pyGenomeTracks: Reproducible plots for multivariate genomic data sets. Bioinformatics 2020, 37, 422–423. [CrossRef]
28. Christofori, G.; Semb, H. The role of the cell-adhesion molecule E-cadherin as a tumour-suppressor gene. Trends Biochem. Sci.1999, 24, 73–76. [CrossRef]
29. Oda, T.; Kanai, Y.; Oyama, T.; Yoshiura, K.; Shimoyama, Y.; Birchmeier, W.; Sugimura, T.; Hirohashi, S. E-cadherin gene mutations in human gastric carcinoma cell lines. Proc. Natl. Acad. Sci. USA 1994, 91, 1858–1862. [CrossRef] [PubMed]
30. Shimada, S.; Akiyama, Y.; Mogushi, K.; Ishigami-Yuasa, M.; Kagechika, H.; Nagasaki, H.; Fukamachi, H.; Yuasa, Y.; Tanaka, S. Identification of selective inhibitors for diffuse-type gastric cancer cells by screening of annotated compounds in preclinical models. Br. J. Cancer 2018, 118, 972–984. [CrossRef]
31. Carneiro, P.; Moreira, A.M.; Figueiredo, J.; Barros, R.; Oliveira, P.; Fernandes, M.S.; Ferro, A.; Almeida, R.; Oliveira, C.; Carneiro, F.; et al. S100P is a molecular determinant of E-cadherin function in gastric cancer. Cell Commun. Signal. 2019, 17, 155. [CrossRef] [PubMed]
32. Pinheiro, H.; Bordeira-Carriço, R.; Seixas, S.; Carvalho, J.; Senz, J.; Oliveira, P.; Inácio, P.; Gusmão, L.; Rocha, J.; Huntsman, D.; et al. Allele-specific *CDH1* downregulation and hereditary diffuse gastric cancer. Hum. Mol. Genet. 2010, 19, 943–952. [CrossRef] [PubMed]
33. Rossi, T.; Tedaldi, G.; Petracci, E.; Abou Khouzam, R.; Ranzani, G.N.; Morgagni, P.; Saragoni, L.; Monti, M.; Calistri, D.; Ulivi, P.; et al. E-cadherin Downregulation and microRNAs in Sporadic Intestinal-Type Gastric Cancer. Int. J. Mol. Sci. 2019, 20, 4452. [CrossRef]
34. Huang, R.; Ding, P.; Yang, F. Clinicopathological significance and potential drug target of *CDH1* in breast cancer: A meta-analysisand literature review. Drug Des. Devel. Ther. 2015, 9, 5277–5285. [PubMed]
35. Canale, M.; Casadei-Gardini, A.; Ulivi, P.; Arechederra, M.; Berasain, C.; Lollini, P.-L.; Fernández-Barrena, M.G.; Avila, M.A. Epigenetic Mechanisms in Gastric Cancer: Potential New Therapeutic Opportunities. Int. J. Mol. Sci. 2020, 21, 5500. [CrossRef] [PubMed]
36. van der Post, R.S.; Vogelaar, I.P.; Carneiro, F.; Guilford, P.; Huntsman, D.; Hoogerbrugge, N.; Caldas, C.; Schreiber, K.E.C.; Hardwick, R.H.; Ausems, M.G.E.M.; et al. Hereditary diffuse gastric cancer: Updated clinical guidelines with an emphasis on germline *CDH1* mutation carriers. J. Med. Genet. 2015, 52, 361–374. [CrossRef]

37. Tedaldi, G.; Pirini, F.; Tebaldi, M.; Zampiga, V.; Cangini, I.; Danesi, R.; Arcangeli, V.; Ravegnani, M.; Abou Khouzam, R.; Molinari, C.; et al. Multigene Panel Testing Increases the Number of Loci Associated with Gastric Cancer Predisposition. Cancers 2019, 11, 1340. [CrossRef]
38. UCSC Genome Browser. Available online: https://genome.ucsc.edu/ (accessed on 8 May 2021).
39. FANTOM Consortium and the RIKEN PMI and CLST (DGT); Forrest, A.R.R.; Kawaji, H.; Rehli, M.; Baillie, J.K.; de Hoon, M.J.L.; Haberle, V.; Lassmann, T.; Kulakovskiy, I.V.; Lizio, M.; et al. A promoter-level mammalian expression atlas. Nature 2014, 507, 462–470.
40. Fishilevich, S.; Nudel, R.; Rappaport, N.; Hadar, R.; Plaschkes, I.; Iny Stein, T.; Rosen, N.; Kohn, A.; Twik, M.; Safran, M.; et al. GeneHancer: Genome-wide integration of enhancers and target genes in GeneCards. Database 2017, 2017. [CrossRef]
41. Livak, K.J.; Schmittgen, T.D. Analysis of relative gene expression data using real-time quantitative PCR and the 2(-Delta Delta C(T)) Method. Methods 2001, 25, 402–408. [CrossRef]
42. Schmittgen, T.D.; Livak, K.J. Analyzing real-time PCR data by the comparative C(T) method. Nat. Protoc. 2008, 3, 1101–1108. [CrossRef] [PubMed]

Original paper 4

Combined disruption of two *CDH1* intronic enhancers outperform the effect of loss of function *CDH1* coding mutations and modulate lung metastases in mice

Celina São José[1,2,3]†, Rita Barbosa-Matos[1,2,4]†, Silvana Lobo[1,2,4], Marta Ferreira[1,2], Ana André[1,2], Nelson Martins[1,2], Sofia Valente[1,2], Hugo Pinheiro[1,2], Renata Carriço[1], Rocío Acuña-Hidalgo[7], Lila Allou[7], José Bessa[1], Stefan Mundlos[7,8,9], Carla Oliveira[1,2,12]*

† These authors contributed equally to this work.

[1] Institute for Research and Innovation in Health, University of Porto, Porto, Portugal;

[2] Institute of Molecular Pathology and Immunology, University of Porto, Porto, Portugal

[3] Doctoral Programme in Biomedicine, Faculty of Medicine, University of Porto, Porto, Portugal

[4] Doctoral Programme on Cellular and Molecular Biotechnology Applied to Health Sciences, Institute of Biomedical Sciences Abel Salazar, University of Porto, Porto, Portugal

[7] Max Planck Institute for Molecular Genetics, RG Development & Disease, Berlin, Germany

[8] Institute of Medical and Human Genetics, Charité Universitätsmedizin Berlin, Berlin, Germany

[9] Berlin-Brandenburg Center for Regenerative Therapies (BCRT), Charité Universitätsmedizin Berlin, Berlin, Germany

[10] Faculty of Medicine of the University of Porto, Porto, Portugal

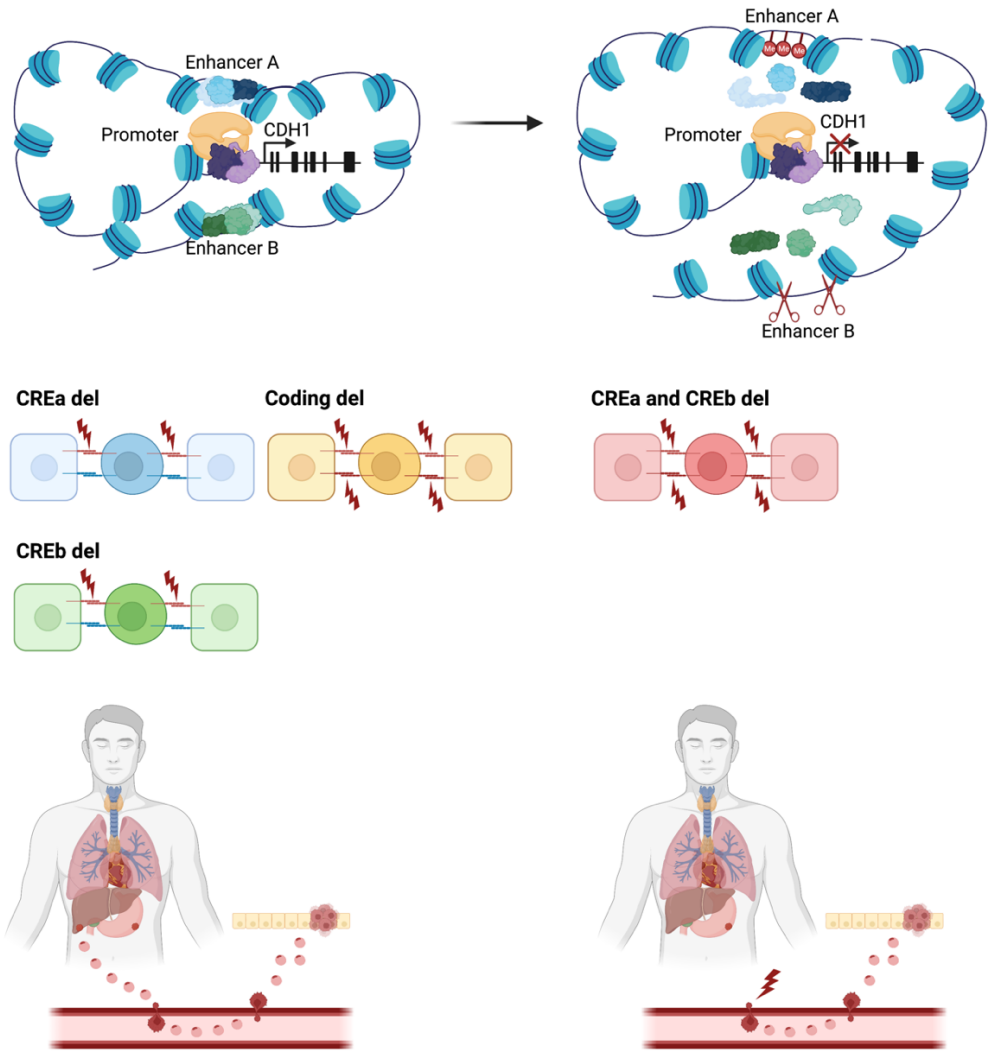
Manuscript *in preparation*.

Author statement

Celina São José: Investigation, methodology, conceptualization, formal analysis, *in vitro* experimental work, bioinformatic analysis, *in vivo* experimental work, writing – original draft, writing – review and editing.

Explore the impact of exonic and hypomorphic *CDH1* intronic enhancers in E-cadherin expression

In an attempt to identify novel causes of E-cadherin loss of function in gastric cancer (GC), we explored the *CDH1* regulatory network. Working under the hypothesis that deletion of a regulatory element has a similar impact as a coding deletion, we selected and CRISPR-Cas9 edited two potential intronic regulatory elements by comparison with a common exonic deletion. At the *CDH1 locus,* intronic regulatory elements (CRE) act in a hypomorphic and redundant manner, given that impacting one CRE results in 50% *CDH1* expression loss, while editing both CREs completely downregulates *CDH1* expression, similarly to an exonic deletion. At the genome-wide level, exonic deletion impairs interferon-associated pathways. While edition of both CREs has a stronger impact in adhesion-associated pathways, that translates in reduced metastasis in mice and increased survival in breast cancer patients. Altogether, our results show two hypomorphic regulatory elements that have a similar impact as a coding deletion in *cis*, yet in *trans* act as protectors against metastasis formation, in opposition to the coding deletion.



Abstract

Regulatory elements are instrumental for proper gene expression in time and space, for which enhancers play an important role in a dose-dependent manner. Given that E-cadherin plays a major role in epithelial tissue architecture and polarity, and may trigger gastric cancer (GC) development if disrupted. Nevertheless, genetic variants, methylation and post-translational modifications fail to explain most *CDH1* loss of function phenotypes in sporadic GC and their hereditary counterpart. Working under the hypothesis that intronic *cis*-regulatory elements (CREs) control *CDH1* expression, similarly to exonic regions, we selected two CREs and evaluated their impact in *CDH1* expression, genome-wide transcriptome and metastasis formation. Edition of two hypomorphic enhancers completely downregulates *CDH1* expression, similarly to an exonic deletion, that separately contribute with 50% expression loss. At the genome-wide level, we found increased impairment of adhesion-associated pathways upon both CREs edition, as opposed to the coding deletion that displayed impairment of interferon-associated pathways alone. The combination of *cis* and *trans* impairments translates in reduced metastasis in mice and increased survival in breast cancer patients. Altogether, our results display CREs controlling *CDH1* expression and suggest a role beyond the *CDH1 locus*, pinpointing CREa and CREb as elements with a major role in metastasis establishment, which when amplified may lead to metastasis development.

Introduction

The exponential increase in next-generation sequencing (NGS) technologies in recent years originated large amounts of data, particularly for whole genome sequencing (WGS). Assigning a functional impact to non-coding copy number variants (CNVs) and single nucleotide variants (SNVs) pose a particularly arduous task for WGS since it is complex to differentiate all the functionally significant variants and validate their protein impact. Another critical challenge for the post-genome era is the validation of *cis*-regulatory elements (CREs) as functional enhancers, capable of regulating gene expression. Even when non-coding mutations corroborate a disease phenotype, not always a causality can be inferred. Despite the support of the encyclopedia of DNA elements (ENCODE) consortium (Feingold et al. 2004) to assign function to non-coding variants, regions lacking biochemical annotations and, thus, multiple non-coding CNVs and SNVs remain functionally unexplained (Abascal et al. 2020).

Accordingly, most enhancers are selected through biochemical annotations, encompassing interaction with gene promoter, open chromatin, histone modifications, and transcription factors binding sites, and CRISPR-Cas9 editing to perturb their functions (Will et al. 2017). A second approach to test functional enhancer properties is enhancer reporter assay (Banerji et al. 1981) or a massive parallel enhancer assay (Inoue and Ahituv 2015). In such cases, the enhancer may drive a reporter gene expression via a minimal promoter. Enhancers can act individually or synergistically to ensure proper gene expression of a target gene or multiple genes (Will et al. 2017; Gasperini et al. 2019). Previous studies reported that combinations of enhancers can lead to a sharper and more robust pattern of gene expression (Perry et al. 2011; Panigrahi and O'Malley 2021). Likewise, due to the redundancy of certain enhancers they can compensate for loss of other enhancers ((Panigrahi and O'Malley 2021). Non-coding CREs frequently act in a synergic and dose-dependent manner that may result in hypomorphic alleles (Gasperini et al. 2019). Hence, non-coding homozygous variants may contribute with different degrees of loss of function and multiple variants may be required in similar CREs to produce a loss of function phenotype similar to a coding variant (Goto et al. 2016; Minnerop et al. 2017). Promoters compose another class of regulatory elements with similar biochemical annotations that can also be simultaneously associated with several enhancers (Barski et al. 2007). Hence, many regulatory elements are yet to be identified, as well as their action mechanism. Their interplay can elucidate novel DNA regions pivotal for an accurate gene expression which can be more impact-full than the coding region itself. The complete functionality of a genetic *locus* exceeds its expressed protein product, being its autonomy and potency a result from the combination of its coding and regulatory regions (Grosveld et al. 2021). To further explore this concept, we studied a gene that is essential for embryonic development, whose expression is dynamically controlled during tumour initiation and progress of epithelial cancers. The *CDH1* gene encodes for the E-cadherin

protein, and the basis of its dynamic expression and dosage effect is far from being understood.

Despite E-cadherin is a major player in EMT (epithelial-mesenchymal transition), hereditary diffuse gastric cancer (HDGC) syndrome caused by E-cadherin loss of function due to coding mutations does not lead to a tumour with a mesenchymal phenotype. Instead, it originates diffuse-type gastric cancer (DGC) and lobular breast cancer (LBC) with a unique carcinogenic behaviour. It is characterized by poorly cohesive round-shaped single cells with large vacuoles filled with mucins known as signet-ring cells. One of its main oncogenic processes is the disruption of cellular adhesion, where E-cadherin loss occurs in the first step of tumour formation, unlike what happens in other types of gastric cancers (GC).

This raises the hypothesis that pivotal E-cadherin functions are determined by regulatory elements embedded in its non-coding regions rather than in its coding regions. Our idea was supported by previous work from our group and others, in which *CDH1* intron 2 stands out as an intronic sequence that regulates gene expression in mice and originates a tissue-specific alternative transcript (*CDH1a*; ENST00000564745.1) (Stemmler et al. 2003, 2005). Moreover, *CDH1* exonic mutations cannot explain all *CDH1*-associated disease phenotypes. For example, families carrying germline *CDH1* mutations diagnosed with HDGC, despite not having coding mutations in *CDH1* or in the remaining exome, still present loss of RNA expression of a single *CDH1* allele.

Thus, we investigated coding and non-coding regions of the *CDH1* locus, to understand their contribution to E-cadherin loss of function and unveil its regulatory network. We performed circular chromosome conformation capture and sequencing (4C-seq) and Assay for Transposase-Accessible Chromatin and sequencing (ATAC-seq) to dissect the *CDH1* regulatory network and define putative CREs. Further, we CRISPR-Cas9 disrupted the selected CREs and evaluated its impact in metastasis development *in vivo* and at the genome-wide transcriptomic level. Lastly, we evaluated the clinical impact of the selected CREs in breast cancer (BC) patients. Our data show that intronic hypomorphic alleles mediate E-cadherin expression. Moreover, the combined disruption of *CDH1* enhancer regions caused a synergistic effect leading to complete loss of function. Beyond the *CDH1* gene, CREs display a decreased metastatic potential *in vivo* associated with a genome-wide transcriptomic downregulation of adhesion-associated pathways. In patients, the lack of CREs leads to increased overall survival, while their amplification has the opposite effect.

Materials and Methods

Cells and cell culture

MKN74, MKN45, respective clones, GES-1 and AGS were cultured in 1x RPMI 1640 medium (Gibco), supplemented with 10% of Foetal Bovine Serum (FBS, Gibco), and 1% penicillin-streptomycin (Gibco). Hek293T cell line was grown in DMEM supplemented with 10% FBS (Gibco) (FBS, Gibco), and 1% penicillin-streptomycin (Gibco). Embryonic Stem Cells C2 (ESC) were cultured in gelatinized dishes on a layer of mitosis inactivated murine embryonic fibroblasts in KO-DMEM, supplemented with 4,500 mg/ml glucose, with sodium pyruvate (Gibco) containing 15 % FCS (PANSera ES), 1x glutamine (Lonza), 1x penicillin/streptomycin (Lonza), 1x non-essential amino acids (Gibco), 1x nucleosides (Chemicon), 0.1 mM β-mercaptoethanol (Gibco) and 1000 U/ml LIF (Murine Leukemia Inhibitory Factor ESGRO™ (Chemicon). All cell lines were cultured in a 5% CO_2 atmosphere at 37°C. All experiments were performed in triplicates, unless stated otherwise.

4C-seq

4C-seq libraries from the MKN74 cell line and clones were generated, as previously described (Van De Werken et al. 2012; Krijger et al. 2020). 4C-seq libraries containing $1x10^7$ cells were cross-linked in 2% paraformaldehyde (PFA) and lysed to form a nuclei suspension. The strategy was designed to include 4-bp cutters, DnpII (New England BioLabs) as a primary and Csp6I (New England BioLabs) as a secondary restriction enzymes and T4 DNA Ligase (Thermofisher) for re-circulation. 4C-seq libraries were purified using Amicon Ultra-15 10 kDa (MWCO) (Millipore) 50 mL columns and amplified by PCR with a total of 3.2µg per reaction (**Supplementary Table 1**). Samples were sequenced on Illumina HiSeqX technology according to standard protocols.

Following high-throughput sequencing of 4C libraries, we mapped *CDH1* interactions on a genome scale, using a bioinformatics pipeline, based on Krijger PHL, et al for demultiplexing, alignment, trimming of adaptors and normalization, and on PeakC and fourSig to identify both *cis*- and *trans*- interactions, respectively. For PeakC we used window size 2, alpha fdr 0.1 and minimal distance 500 (Geeven et al. 2018)and for fourSig we used window size 5, 1000 iterations, 0.001 fdr and 0.01 fdr probability and considered interactions in categories 1 and 2.

ATAC-seq

ATAC-seq libraries were generated from the MKN74 cell line and respective clones, as described previously (Buenrostro et al. 2013, 2015). The starting material was $5x10^4$ cells for all ATAC-seq libraries. Cells were lysed and

chromatin was tagmented, using Tn5 transposase and purified with the MiniElute PCR purification kit (Qiagen), following the manufacturer's protocol. The appropriate number of PCR cycles was determined using SYBR Green qPCR. Transposed DNA fragments were amplified by PCR and purified using QIAquick PCR purification kit (Qiagen), following the manufacturer's protocol. Quality control of ATAC-seq libraries were accessed by tapestation and samples were sequenced on Illumina HiSeqX technology according to standard protocols.

We mapped ATAC-seq peaks, resourcing to an ENCODE bioinformatics pipeline with default parameters for adaptors trimming (cutadapt: -e 0.1 -m 5), alignment (bowtie), quality filtering for mapq < 30, removal of duplicates and mitochondrial reads (piccard), peak calling (macs2), replicate statistical analysis (idr, p-value < 0.05) and differential peaks analysis (homer).

Luciferase activity assays

Exon2, CREa and CREb were PCR amplified with QIAGEN Multiplex PCR Master Mix, following the manufacturer's instructions, using genomic DNA (gDNA) from MKN74 gastric cancer cells and specific primers with 5' *Sac*I and 3' *Nhe*I overhangs (**Supplementary Table 2**). Next, each PCR amplified sequence, pGL3-Basic and pGL3-Promoter vectors (Promega) were digested with *SacI* - High Fidelity (New England BioLabs) and *NheI* (New England BioLabs), according to the manufacturer's instructions. The target sequences were cloned into both vectors using T4 DNA Ligase (Invitrogen, Life Technologies) following the manufacturer's instructions. GES-1 and Hek293T cells were used for the luciferase reporter assays. For the promoter reporter assay, cells were transfected with pGL3-Basic and pNL1.1.PGK[Nluc/PGK] (Promega), the internal transfection efficiency control, using Lipofectamine® 3000 (ThermoFisher Scientific) accordingly to manufacturers' instruction. For the enhancer reporter assay, cells were transfected with pGL3-Promoter and pNL1.1.PGK[Nluc/PGK]. Data was acquired using the Nano-Glo® Dual-Luciferase® Reporter Assay System in a Synergy™ 2 Multi-Mode Microplate Reader (BioTek). The promoter and enhancer activity results were normalized to the activity of the transfection control, and are presented as a relative response ratio.

Animal care

Experiments were performed using wild-type AB, Tübingen (TU) and Casper Zebrafish (*Danio rerio*) strains. Zebrafish and mouse were maintained and manipulated in the animal facility at i3S and MPIMG, respectively. All zebrafish and mouse experiments complied with standard animal care guidelines and national legislation for animal experimentation, and only certified persons carried out the experimentation.

Zebrafish Enhancer Assay

Exon2 and CREb regions were PCR amplified with QIAGEN Multiplex PCR Master Mix (Qiagen), following the manufacturer's instructions. PCR amplified products were cloned into pCR™8/GW/TOPO™ TA vector (Invitrogen) following manufacturer's recommendations and then cloned into the ZED vector (Bessa et al. 2009) via Gateway cloning using Gateway™ LR Clonase™ II Enzyme mix kit (Invitrogen) according to standard protocols. Vectors were purified using phenol:chloroform and injected into one-cell-stage zebrafish embryos (AB strain), together with mRNA encoding Tol2 protein, in a 1:1 ratio and a final concentration of 50 ng/ul each and 0.05% phenol red. For mRNA Tol2 preparation, transposase cDNA is cloned into the pCS2FA-Tol2 plasmid and linearized by *Not*I restriction digestion (New England BioLabs) followed by purification with phenol/chloroform according to the manufacturer's protocol. Tol2 mRNA was synthesized in vitro by using Sp6 RNA polymerase kit and purified with a Sephadex column followed by phenol/chloroform purification. A minimum of 1000 embryos were injected per each ZED construct. Embryos were kept in E3 medium with 0.003% 1-phenyl-2-thiourea (Sigma-Aldrich) to prevent pigmentation and were staged as reported in Kimmel et al. 1995. At day 3 post fertilization (dpf), injected embryos showing reporter expression were raised to sexual maturity and germinal line transmission was performed by outcrossing injected fish with wild-type animals (AB, TU or Casper strains) and isolating F1 progeny with GFP and RFP expression from 24 hours post fertilization to 4 dpf. Embryos showing GFP expression were anesthetized with 0.01% of tricaine methanesulfonate (Sigma-Aldrich) and 0.003% 1-phenyl-2-thiourea (Sigma-Aldrich) in E3 medium. Expression patterns were observed using a stereomicroscope Leica M80 (Leica Microsystems), and documented using a stereomicroscope Leica M205 with epifluorescence filters with a digital camera attached (Leica LAS BRAND+TYPE/ Leica Microsystems). Post-imaging analysis was done in Fiji-ImageJ.

Mouse LacZ Enhancer Assay

The enhancer activity was studied by pPGK-FRT-LacZ reporting vector. Briefly, the selected intronic regions were PCR amplified from human gDNA (fibroblasts), using Phusion High-Fidelity DNA polymerase (New England Biolabs) with PacI overhang primers (**Supplementary Table 3**). Enhancer candidate regions were cloned into a pPGK-FRT-LacZ vector and transfected into C2 Mouse embryonic stem (C2 ES) cells targeted by flippase recognition target system. Successfully transfected cells express hygromycin resistance for selection. Transgenic embryos were generated from positive genetically modified C2 ES cells by diploid or tetraploid aggregation (Artus and Hadjantonakis 2011) and tested for LacZ reporter activity at E12.5 and E14.5. CD1 female mice were used as foster mothers. The chimeric state of embryos was evaluated based on the eye colour of the mouse (black for high chimeric state and white for low chimeric

state).Tissues from E12.5 and E14.5 mouse embryos were dissected, fixed in 4% PFA for 10 minutes and stained in X-gal solution for a few hours to overnight, as described previously (Lobe et al. 1999). Dissected organs were imaged using a ZEISS SteREO Discovery.V12 with cold light source CL9000 microscope and Leica DFC420 digital camera (Leica Microsystems). Embryos were stored at 4 °C in 4% PFA in 1x PBS.

CRISPR-mediated genome editing

CRISPR single-guide RNAs (sgRNAs) (**Supplementary Table 4**) were designed using Zhang lab web platform and benchling online platform, and cloned into the pSpCas9(BB)-2A-Puro (PX459; 48139) vector as previously described (Ran et al. 2013). Plasmids were sequenced to confirm correct Cas9 and sgRNA's sequences integration and further grown for DNA purification using the Qiagen maxiprep kit (Qiagen), following the manufacturer's protocol.

MKN74 and MKN45 cell lines were seeded in twelve-well plates and transfected with 1µg of each PX459-sgRNA vector targeting each breakpoint, and transiently transfected using Lipofectamine P3000 reagent® following manufacturer's protocol. After 48h, CRISPR-edited cells were selected with 5 µg/mL of puromycin (Merck) added to RPMI medium, and single-cell sorted to generate phenotypically homogeneous cell populations. Cell clones containing the desired editing were expanded, cryopreserved and used for gDNA, total RNA and total protein extraction and analysis.

CRISPR/Cas9 clones genotyping and expression analysis

gDNA was extracted from cells using the NZY Tissue gDNA Isolation kit (Nzytech), according to the manufacturer's instructions. CRISPR-Cas9 engineered clones were genotyped by PCR using Multiplex PCR kit (Qiagen) with primers flanking exon2, CREa and CREb (**Supplementary Table 5**), and the resulting PCR products were runned in 2% agarose gel, purified by ExoSAP-IT™ PCR Product Cleanup Reagent ™ (Applied Biosystems™) or Gel Band Purification Kit (GE Healthcare), according to the manufacturer's instructions, and sanger sequenced through capillary sequencing with the Abi Prism 310 Genomic Sequencer, using the BigDye™ Terminator v3.1 Cycle Sequencing Kit (Thermo Fisher Scientific Inc.©). The only exception was Clone CRE(A)-74.1 which was genotyped with the SALSA®MLPA® probe mix P083-D1 *CDH1* (MCR Holland) and compared with the MKN74 genomic pattern obtained, used as a reference sample. For *CDH1* expression analysis, total RNA was extracted following instructions of the mirVana miRNA Isolation Kit (Thermo Fisher Scientific). Reversed transcription to obtain cDNA was performed using Superscript II Reverse Transcriptase and random hexamer primers (Thermo Fisher Scientific). Quantitative real time-PCRs (qRT-PCR) were carried out using PrimeTime-qPCR

assays on a 7500 Fast Real-Time PCR system (Thermo Fisher Scientific).For *CDH1* targeted regions probes spanned from exon 1 to exon 2 (Hs0103959_m1 TaqMan probe, Thermo Fisher Scientific) and from exon 4 to exon 5 (Hs.PT.58.14402586.g, Integrated DNA Technologies). As an endogenous control human 18S ribosomal RNA was used (Custom made probe, Integrated DNA Technologies). Data was analyzed by the comparative $2^{-\Delta\Delta CT}$ method (Livak and Schmittgen 2001).

Promoter and Enhancer B methylation analysis

For the methylation analysis, gDNA was converted using the EpiTect Bisulfite Kit (Qiagen®) following manufacturer's instructions. DNA of the SNU-1 cell line, which is constitutively methylated at the *CDH1* gene, was used as a positive methylation control. A 173-bp fragment in the promoter region (12 CpG sites) and a 262-bp fragment in CREb (6 CpG sites) were PCR amplified with specific primers for the converted DNA (**Supplementary Table 6**). PCR products were purified and sequenced in the ABI 3130 Automated Capillary DNA sequencer (Applied Biosystems™).

Flow cytometry

To determine E-cadherin expression at the cell surface of MKN74 and MKN45 CRISPR-edited cells, flow cytometry was performed in BD Accuri™ C6, from BD Biosciences, and FlowJo version 10 was used to analyze the acquired data. Briefly, cells were harvested from a six-well plate at 90% confluency using Versene (ThermoFisher Scientific) and blocked with 3% bovine serum albumin (BSA) in 1x phosphate buffer saline (PBS), for 30 min. Cells were incubated with HECD1 (ThermoFisher Scientific), the primary monoclonal antibody against E-cadherin at 1:100, during 1 hour, in the dark, at 4ºC. Subsequently, anti-mouse Alexa Fluor 647 (ThermoFisher Scientific) was the secondary antibody used at 1:500 for 30 minutes at 4ºC in the dark. Cells were fixed in 2% PFA in 1x PBS at 4ºC until being analyzed in the cytometer. The mean fluorescence intensity was measured for at least 20,000 gated events per sample.

Mouse intravenous injection

MKN74 cell line and respective clones were injected in the tail vein of nude CBA female mice with 6 to 8 weeks. Bodyweight loss, dehydration and tumour presence was evaluated once a week. Mice were euthanised at 2 and 4 weeks, at each timepoint, lungs were collected, fixed in formalin, paraffin-embedded and stained for hematoxylin/eosin for further analysis. Lung metastases were evaluated by an experienced pathologist in subpleural, peribronchial,

intraparenquimatous and intravenous. For metastatic potential analysis, intravenous lesions were not considered.

RNA-seq

Total RNA from cell lines was extracted using mirVana miRNA Isolation Kit (Invitrogen) following the manufacturer's instructions. For each cell line and clones, RNA was extracted from two independent biological replicates. Total RNA integrity/quality was assessed on an Agilent 2100 bioAnalyzer RNA Chip (GenCore, i3S). High-quality RNA eluates (RIN>9.0 and 18S/28S ratio>2.0) were sequenced in Illumina HiSeqX technology according to standard protocols. The global expression changes associated with specific CREs of *CDH1* was accessed by Differential Expression (DE) analysis using the DESeq2 package (p-value FDR <0.01 and a logarithmic fold change > 2 or <-2). Gene list was obtained by comparing each CRISPR/Cas9-engineered clone with the parental cell line.

Survival analysis

Data was collected from the metabric project (Curtis et al. 2012)to determine survival curves. Kaplan-Meier was used to plot survival curves and differences were considered significant if p-value≤0.05. R was used for statistics.

Statistical analysis

Statistical analysis was performed using GraphPad Prism version 7.00 software (GraphPad Software Inc.). A T-student test was used for comparison analysis, assuming equal variance between clones and WT samples. Differences were considered significant when p-value<0.05.

Results

Selection of putative cis-regulatory elements (CREs)

Working under the hypothesis that E-cadherin functions are determined by regulatory elements embedded in its non-coding regions, rather that the coding sequence, we selected a 5-prime exon (exon 2) and two intronic regions (CREa and CREb) to unveil their contribution to E-cadherin loss of function.

Exon 2 displays promoter-like features and accessible chromatin in normal stomach tissue (**Figure 1A**). Chromatin accessibility and enhancer-like features were found for CREa (**Figure 1A**). Indeed, CREa methylation has been linked with GC tumours and cell lines with *CDH1* knockdown, suggesting a potential role in controlling E-cadherin expression (Tedaldi et al. 2021). Although CREb lacks enhancer-associated chromatin marks (**Figure 1A**), is located within a FANTOM5 enhancer, the TSS of transcript ENST00000564745.1 (*CDH1a*) with a role in gastric cancer invasion (Pinheiro et al. 2012) and in an overall important intron for E-cadherin tissue-specific expression (Stemmler et al. 2005). We collected relevant information depicted in **Figure 1A** and focused on the stomach, as a hallmark tissue for E-cadherin expression.

Suggestive of a regulatory role, we found accessible chromatin in exon 2 and CREa in MKN74 gastric cancer cell line and validated in normal stomach tissue, and *CDH1* promoter interactions with all three selected regions (**Figure 1B**). Thus, we prioritized exon 2, CREa, and CREb for enhancer studies to understand whether these regions have enhancer activity.

We found tissue-specific enhancer activity in stomach-derived GES-1 cell line for all tested regions (**Figure 1C**). *In vivo* enhancer assays revealed exon 2 and CREb enhancer activity in zebrafish and CREa tissue-specific enhancer activity in mouse embryos (**Figure 1D,e**). These results suggest exon 2, CREa and CREb as potential regulatory elements, and pinpoint E-cadherin as their main target.

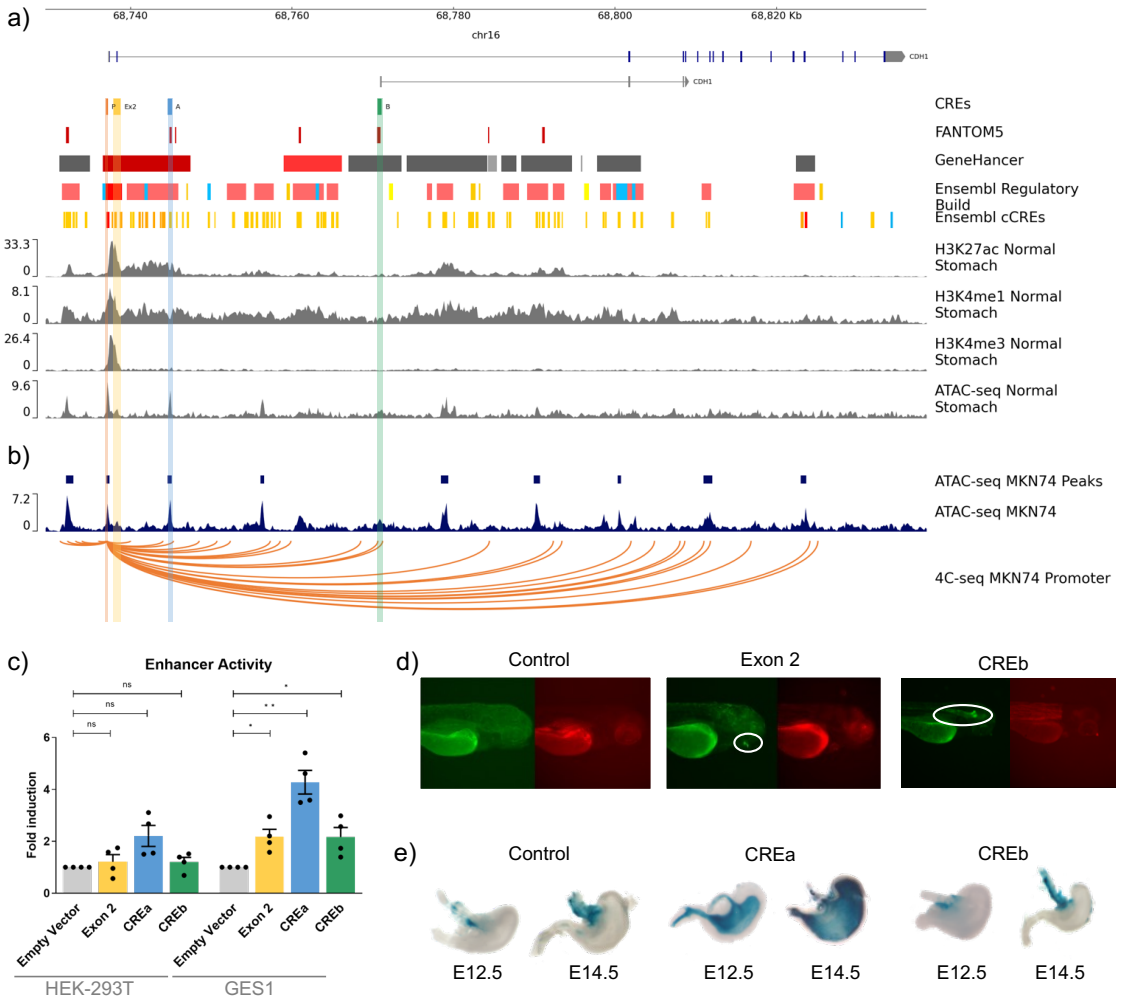


Figure 1. Selection of cis-regulatory elements (CREs). *a) Public available data on putative enhancers, chromatin marks and accessible chromatin in normal stomach tissue from ENCODE, b) Accessible chromatin and interactors of the CDH1 promoter in MKN74 cell line, c) Luciferase assay to test enhancer activity in HEK-293T and GES1 cell line, d) zebrafish enhancer assay for GFP expression, f) enhancer assay in mouse embryos for β-galactosidase expression.*

Hypomorphic enhancers lead to E-cadherin loss of function

A straightforward way to explore the functional impact of these CREs is to evaluate their impact on E-cadherin expression upon perturbation by CRISPR-editing. We edited exon 2, CREa and CREb in two gastric cancer cell lines and obtained six clones, depicted in **Figure 2A**. Homozygous deletion of a coding region, herein represented as exon 2, leads to *CDH1*/E-cadherin complete downregulation at the RNA and protein levels (**Figure 2B,c**). Homozygous edition of either CRE by deletion or inversion leads to an approximate 50% decrease in *CDH1*/E-cadherin expression, suggesting CREa and CREb as *CDH1* hypomorphic enhancers (**Figure 2B,c**). Edition of both CREs (CREa + CREb) leads to complete *CDH1*/E-cadherin loss, as observed for the coding deletion

(**Figure 2B,c**). These results hint the CREs as hypomorphic enhancers with co-operative mode of action, requiring co-inactivation of additional regulatory elements to produce the same effect as a coding variant in *CDH1* expression. Its mode of action may be the result of chromatin accessibility changes in the *CDH1 locus*.

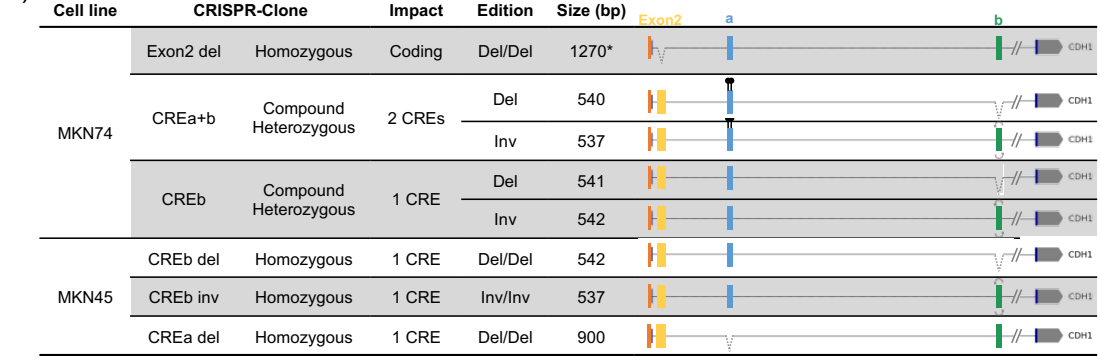
Zooming in on the *CDH1* interactors and accessible chromatin profiles, loss of chromatin accessibility and *CDH1*-promoter interactions, namely with CREa was observed in all MKN74 edited clones, when compared to the parental *CDH1*-expressing cell line (**Figure 2E**). Further, we found a significant reduction in *CDH1* promoter interactions and accessible chromatin within the *CDH1 locus* in all edited clones, particularly in exon 2 deletion (**Figure 2E**). These results may suggest a regulatory re-shuffling of the *CDH1 locus* upon both exonic and CREs edition.

Altogether, these results pinpoint CREa and CREb as hypomorphic enhancers that may act synergistically to drive *CDH1*/E-cadherin expression. To evaluate their potential synergistic effect, we explored the *CDH1*/CREs regulatory network. In a *CDH1*-expressing cell line, the *CDH1* promoter interacts with CREa bilaterally and CREb unilaterally (**Figure 2E**), and, supporting a synergistic effect, a bilateral interaction was identified between CREa and CREb (**Figure 2E**). Within the *CDH1 locus* in a *CDH1*-expressing cell line, while CREa has a similar number of interactions as the promoter, there is a significant decrease in exon 2 and an increase in CREb interactions (**Figure 2E**). These results suggest non-coding regions - CREa and CREb - as potential *CDH1*-regulators, rather than coding regions - exon 2. Given the E-cadherin role in development and adhesion, we investigated its role at the genome-wide transcriptomic level.

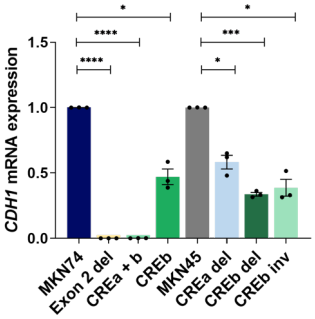
We performed an integrative analysis of the cells' transcriptome and analysed differentially expressed genes between each clone and wild-type cells. We found differentially expressed genes between exon 2 deletion and MKN74 (180 genes), CREa+b edition and MKN74 (2984 genes), CREb edition and MKN74 (660 genes), CREa deletion and MKN45 (556 genes), CREb deletion and MKN45 (1042 genes) and CREb inversion and MKN45 (333 genes). Edition of both CREs has a major impact in both *CDH1* mRNA expression and differentially expressed genes (**Figure 2D**). Despite a similar impact in *CDH1* mRNA expression, exon 2 deletion has a much-reduced number of differentially expressed genes at the genome-wide level (**Figure 2D**). Edition of one CRE has a mild impact in both *CDH1* mRNA expression and differentially expressed genes at the genome-wide level (**Figure 2D**). These results pinpoint CREa and CREb as regulators of *CDH1* expression, but also at the genome-wide level, with co-operative mode of action, in contrast with deletion of a coding region.

a)

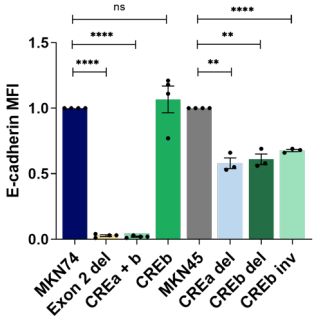
Cell line	CRISPR-Clone		Impact	Edition	Size (bp)
MKN74	Exon2 del	Homozygous	Coding	Del/Del	1270*
	CREa+b	Compound Heterozygous	2 CREs	Del	540
				Inv	537
	CREb	Compound Heterozygous	1 CRE	Del	541
				Inv	542
MKN45	CREb del	Homozygous	1 CRE	Del/Del	542
	CREb inv	Homozygous	1 CRE	Inv/Inv	537
	CREa del	Homozygous	1 CRE	Del/Del	900



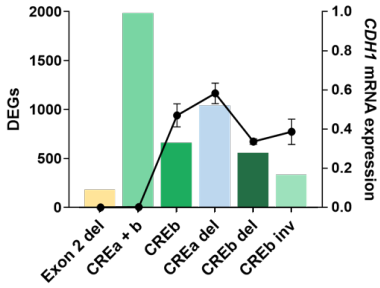
b) *CDH1* mRNA expression



c) E-cadherin protein expression



d) RNA-seq DEGs



e)

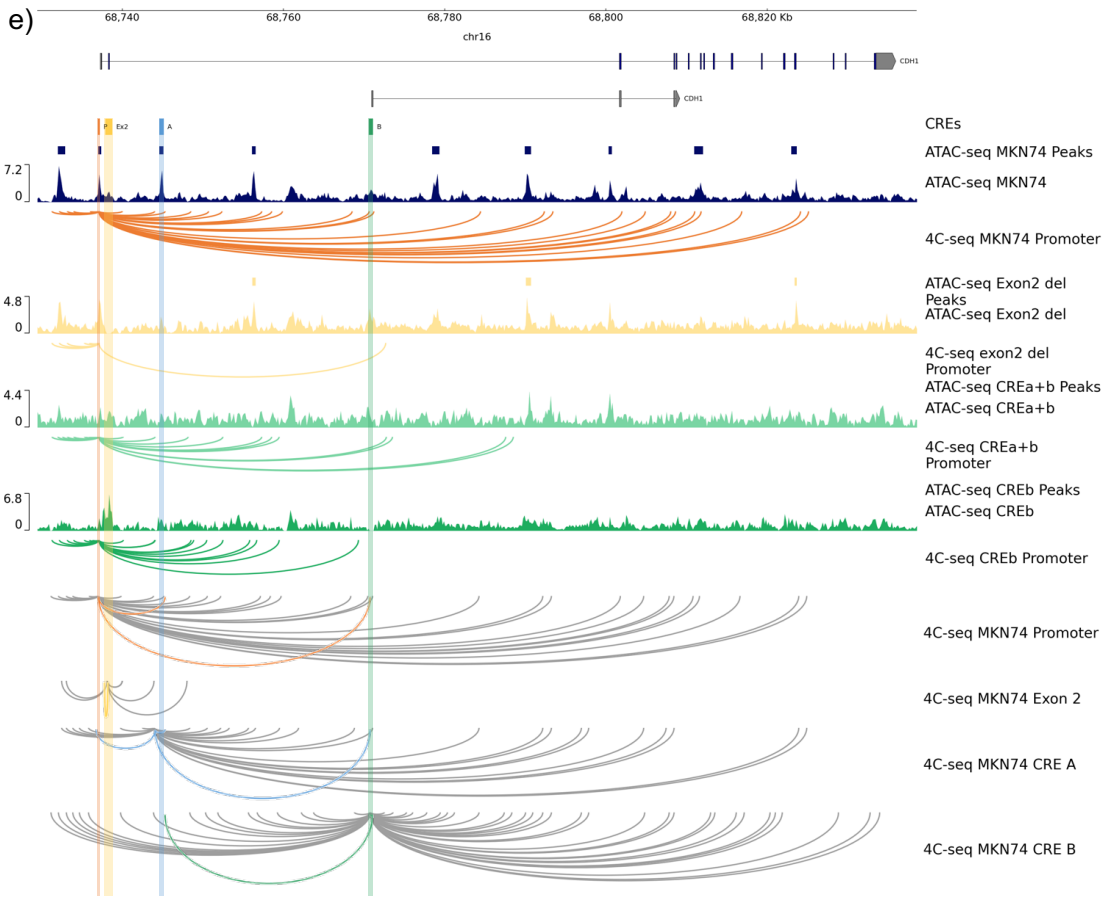


Figure 2. Characterization of hypomorphic alleles. *a) Genotyping of selected clones in MKN74 and MKN45 cell lines, b) CDH1 mRNA expression, c) E-cadherin protein expression, d) RNA-seq differentially expressed genes (DEGs) and normalized CDH1 mRNA expression e) Accessible chromatin and CDH1 promoter interactors map of MKN74 parental and isogenic clones.*

(Epi)genetic alterations at CREs reduce metastatic potential

Our earlier findings suggest that two CREs edition have the same impact on *CDH1*/E-cadherin expression as a coding deletion, but an increased impact in gene expression at the genome-wide level. E-cadherin is a tumour suppressor gene, and its expression loss has been associated with EMT and distant metastasis. Thus, we dissected the metastatic potential of a metastatic gastric cancer cell line (MKN74) carrying either/both CREs or coding deletions. To address this aim, we performed an intravenous injection of MKN74 and respective clones and evaluated lung metastasis at 2 and 4 weeks (**Figure 3A**). A small pilot study has shown that MKN74 intravenous injection produces extensive macroscopic lung metastasis at 80 days post-injection (Sara Rocha 2020).

MKN74 produced the higher amount of lung metastasis, similarly to exon 2 deletion (**Figure 3B**). While edition of one or two CREs produces the least amount of lung metastasis, namely with edition of two CREs with complete absent metastasis (**Figure 3B**). A pairwise comparison of inoculated clones reveals that MKN74 and exon 2 deletion lead to increased metastasis than edition of both CREs (**Figure 3C**). Given that coding deletion and edition of both CREs display a similar impact for *CDH1*/E-cadherin expression, but with distinct metastatic abilities, we dissected the genome-wide transcriptome that may hint on the biological mechanisms behind distant metastasis development.

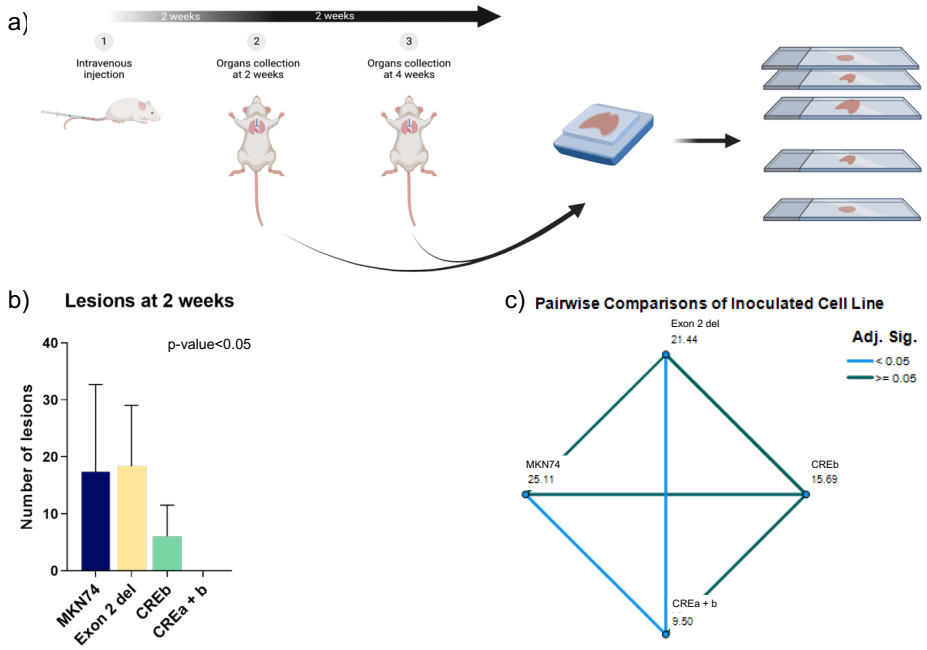


Figure 3. Metastatic potential in mice. *a) Schematic timeline of metastatic assay. b) Number of peribronchial, intraparenchymal and subpleural lesions identified at 2 weeks in each clone. p-value calculated with one-way ANOVA. c) Pairwise comparison of inoculated clones.*

CREs impairment impacts CDH1-associated pathways

In an attempt to unveil the mechanism inhibiting distant lung metastasis in mice, we mined the transcriptomic profile of edited clones. Edition of a coding sequence, either CRE or both CREs display a different transcriptional profile (**Figure 4A**). Enrichment analysis revealed downregulation of cell adhesion and polarity when editing one or both CREs, but not impacted in a coding deletion, herein represented as *CDH1* exon 2 deletion (**Figure 4B**). Yet, upregulation of 'adhesion mediated by integrins' was found in exon 2 deletion and CREa+b edition, and heterophilic/homophilic cell adhesion via plasma membrane adhesion molecules in CREa+b edition (**Figure 4B**). Events associated with actin cytoskeleton were found upregulated in CREa deletion and downregulated in CREb deletion, correlating well with neutral differential expressed genes when both CREs are edited (**Figure 4B**). Invasion-associated pathways, including 'wound healing' were found upregulated in either one or both CREs edition, feature absent in *CDH1* exon 2 deletion (**Figure 4B**). *CDH1* coding deletion, herein shown as exon 2, display upregulation of immune-associated pathways, encompassing 'antigen processing and presentation of peptide antigen via MHC class I', 'cellular response to type I interferon', 'positive regulation of cytokine production', 'regulation of innate immune response' and 'viral life cycle' (**Figure 4B**). These features have been reported for a *CDH1-TANGO6* deletion in a family with extremely early onset DGC and increased disease penetrance (São José et

al. 2023) and for germline coding CNVs and SNVs of unsolved HDGC patients (**original paper 1**). Yet, edition of both CREs downregulate most of these immune-associated pathways (**Figure 4B**). Besides an opposite effect on immune-related pathways, edition of both CREs displays increased proliferation and migration with upregulation of 'WNT signalling pathway', 'epithelial cell proliferation', 'ERK1 and ERK2 cascade', 'epithelial to mesenchymal transition' and 'epithelium migration', features associated with more aggressive tumours (**Figure 4B**). Despite showing features of more aggressive tumours, edition of both CREs still lack one of its hallmarks, metastatic capacity (**Figure 3B**). This is likely related with an impact in a metastatic process, such as 'cellular extravasation', herein downregulated (**Figure 4B**).

We next explored if this impact was related with genome-wide accessible chromatin rewiring. For this, we scanned accessible chromatin changes in promoters or transcription start sites of differentially expressed genes. *CDH1* exon 2 deletion does not share any differentially accessible chromatin in promoters or transcription start sites with differential expressed genes. CREb edited clone displays a mild positive correlation (**Figure 4C**), while both CREs edition displays a strong genome-wide chromatin accessibility with gene expression positive correlation (**Figure 4C**). Notably, these regions represent adhesion- or cadherin-associated pathways in CREa and CREb edition simultaneously, while absent in CREb edition (**Figure 4D**).

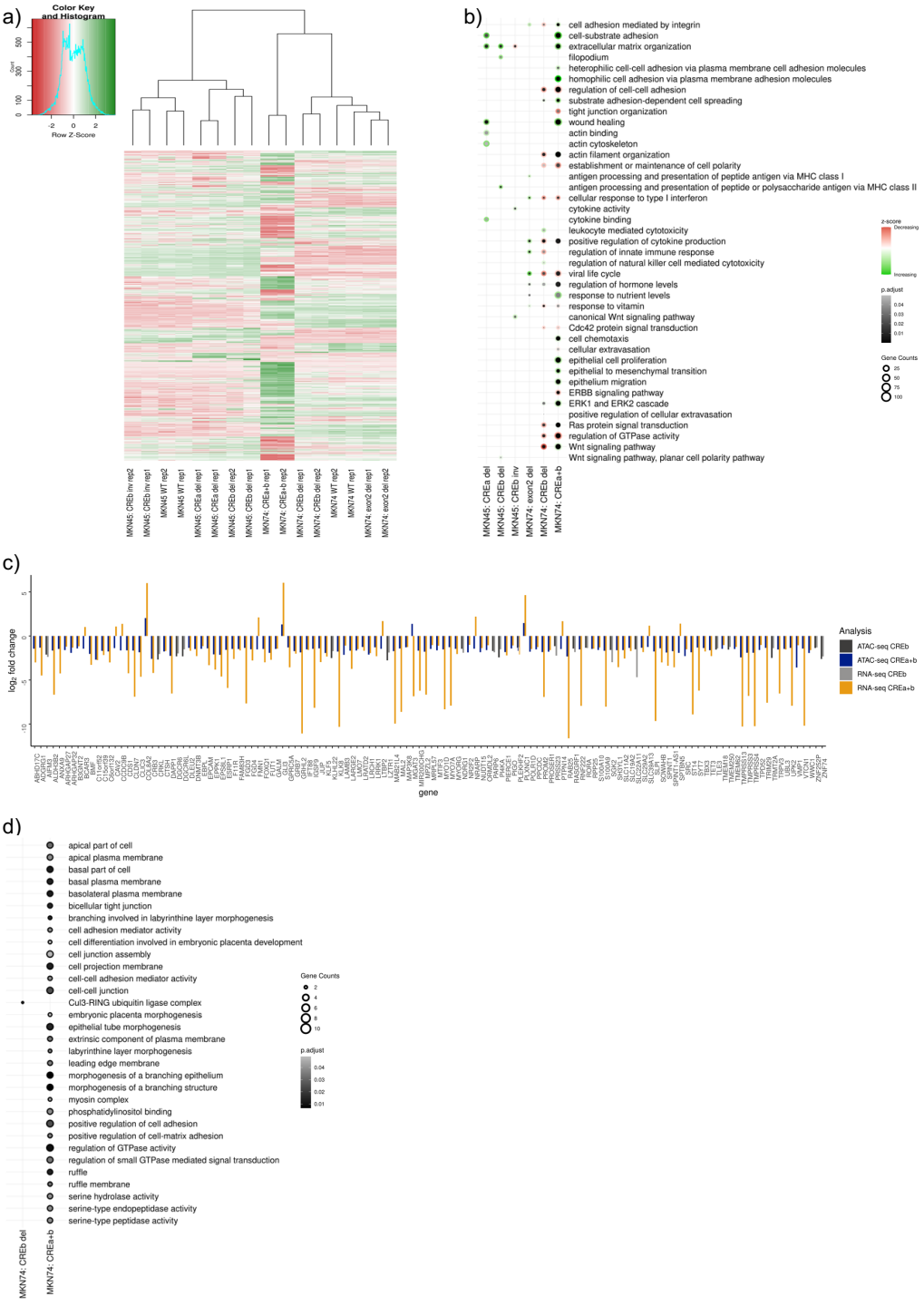


Figure 4. Whole genome transcriptomic impact of CREs. *a) Heatmap representing differentially expressed genes in MKN74 and MKN45 clones, b) Impaired pathways of differentially expressed genes in MKN74 and MKN45 clones, c) Differential accessible chromatin regions overlapping differential expressed genes in MKN74: CREb del and MKN74: CREa+b, d) Impaired pathways of differential accessible chromatin regions overlapping differential expressed genes, e) Gene-set enrichment analysis. Gradient of green represent upregulated genes or pathways, gradient of red represents upregulated genes or pathways, adj p-value in pathways represented in gradient of grey.*

CREs impairment reduces metastatic potential in breast cancer

To dissect this mechanism in tumours, we explored the METABRIC dataset (Curtis et al. 2012), with 2085 breast cancers, for CREs clinical impact. We categorized breast cancer tumours into 4 groups, 1) normal CREa and CREb copy-number, 2) heterozygous CREa and CREb deletion, 3) homozygous CREa and CREb deletion, and 4) CREa and CREb amplification, and evaluated the patients' overall survival. We found that patients with CREs amplification have reduced survival than patients with homozygous deletion of both CREs (p-value=0.055) (**Figure 5A,b**). CREs homozygous deletion is scarce in breast cancer tumours (**Figure 5A,b**), suggesting that this event likely doesn't confer advantage to the tumour. Altogether, these results indicate CREs may act as tumour/metastasis prone if amplified and protective when deleted, pinpoint their genome-wide role for cancer/metastasis development, besides their impact on *CDH1* expression alone.

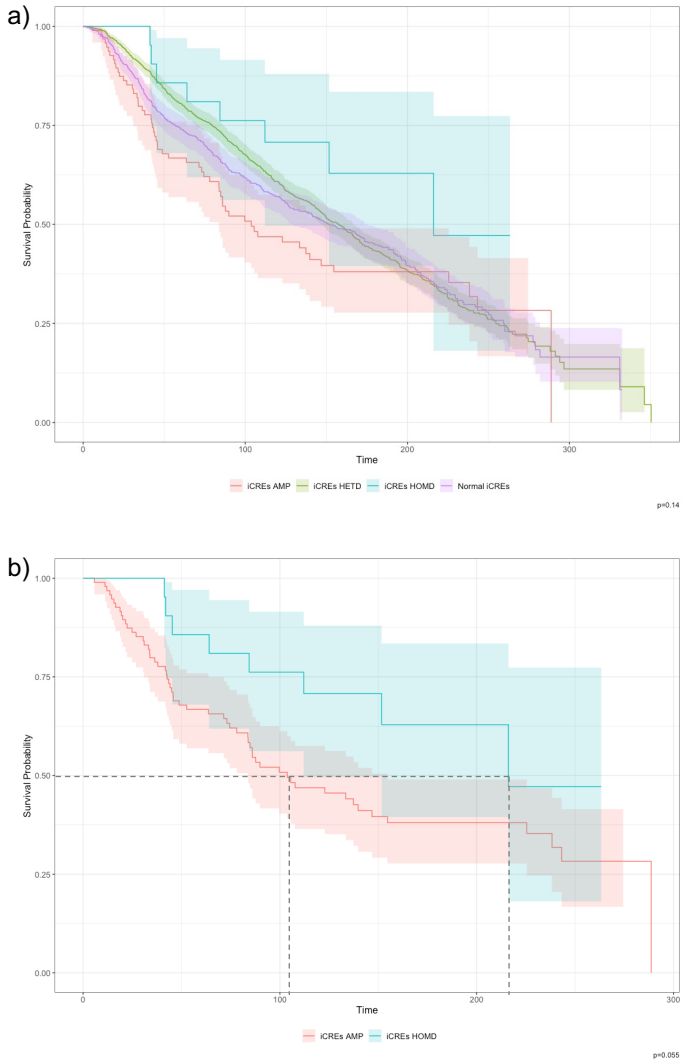


Figure 5. Overall survival of CREs CNVs in breast cancer patients. *a) Overall survival in tumours of the 4 testing groups, b) Overall survival in tumours with CREs homozygous deletion*

and amplification. Red represents CREs amplification (AMP), dark green represents heterozygous CREs (HETD), blue represents homozygous CREs deletion (HOMD), light green represents normal CREs, and violet represents tumours without CREs information.

Discussion

The 3D chromatin architecture grants enhancer-promoter communication, despite distant in the DNA sequence (Will et al. 2017). Their interplay can elucidate novel DNA regions pivotal for an accurate gene expression which can be more impact-full than the coding region itself.

Herein we selected a coding region and two intronic putative CREs to evaluate their impact on a tumour suppressor gene. Our main goal is to determine that non-coding regulatory elements, rather than exons, contribute to a tumour suppressor gene loss of function and genome-wide transcriptomic re-wiring, controlling tumour/metastasis formation.

We collected biochemical annotations and prioritized exon 2, CREa and CREb for enhancer reporter assays to assess their enhancer activity. Both putative CREs and exon 2 display enhancer biochemical/3D-chromatin annotations and tissue-specific enhancer activity in cells and zebrafish/mouse embryos. Besides interactions with the *CDH1* promoter and carrying the promoter for an alternative transcript, CREb lacks enhancer biochemical annotations (Benton et al. 2019). Yet, given its *in vitro* and *in vivo* enhancer activity, we proceeded the analysis with this enhancer as well. CRISPR-Cas9 edition is straightforward way to perturb enhancers and dissect their functional impact. While edition of either CRE impacts *CDH1*/E-cadherin expression by 50%, edition of both CREa and CREb completely knocks down the gene, producing a similar outcome as a coding deletion (exon 2). Evidence for enhancer co-operativity also relates with bilateral interactions between CREa and CREb in a *CDH1*-expressing context. These results pinpoint the CREs as hypomorphic enhancers with co-operative mode of action, requiring co-inactivation of additional regulatory elements to produce the same effect as a coding variant in *CDH1* expression. Indeed, most enhancers are co-operative and hypomorphic (Will et al. 2017). Its mode of action may be the result of chromatin accessibility changes and regulatory re-shuffling of the *CDH1 locus*, as observed in the *EPHA4 locus* (Lupianez et al. 2015).

Given that E-cadherin is a mark of epithelial to mesenchymal transition and epithelial cell migration (Elisha et al. 2018), we tested whether edition of one/both CREs could control migration and metastasis development, beyond their impact in E-cadherin expression. Metastasis assay revealed that both CREs edition lacks lung metastasis formation. Despite a similar impact on *CDH1*/E-cadherin expression, a CREa and CREb simultaneous deletion produces reduced metastasis than a coding deletion, suggesting a regulatory role beyond the *CDH1 locus*.

Thus, we explored the cells' transcriptome to dissect the role of *CDH1* CREs in metastasis formation. Despite a similar *CDH1* knock-down, edition of both CREs displays a much stronger impact on the transcriptome, carrying almost 17 times more misregulated genes, than a *CDH1* coding deletion. This result suggests CREs as major regulators beyond *CDH1* expression. Pathways associated with adhesion, polarity and actin cytoskeleton have been impaired in most tested clones, features commonly observed in E-cadherin deficient cells (Ayollo et al. 2009; Desai et al. 2009; Chen et al. 2014). A yet to prove compensatory adhesion mechanism may rescue adherens junctions in these clones, mediated by integrins in exon 2 deletion and CREa+b edition, or heterophilic/homophilic cell adhesion via plasma membrane adhesion molecules in CREa+b edition. Actin-related enriched terms were upregulated in CREa deletion and downregulated in CREb deletion and neutral in CREa and CREb edition, suggesting that both CREa and CREb regulate actin cytoskeleton, but with opposite regulating patterns. Events associated with immune system were upregulated in the *CDH1* coding deletion (exon 2). Indeed, overexpression of type I interferon genes have been associated with extremely early onset DGC and increased disease penetrance in a family with a *CDH1-TANGO6* large deletion (**original paper 2**). Further, the germline landscape of unsolved HDGC patients display impaired immune-related pathways, specifically in stomach regulatory regions of young-affected individuals (**original paper 1**). Interestingly, these features were absent upon edition of both CREs, indicating an alternative tumorigenic mechanism. Despite showing increased proliferation, migration, and invasion, features commonly observed in more aggressive tumours, edition of both CREs lacks metastatic capacity. Indeed, metastasis is a multi-step process, for which 'cellular extravasation', downregulated in both CREs edition, is essential. Among the cellular extravasation-associated genes herein downregulated, ADAM8, FUT4 and IL1R1 have been linked with promoting invasion, migration and epithelial to mesenchymal transition in breast cancer cells or lung metastasis (Das et al. 2016; Dagenais et al. 2017; Lu et al. 2020), while F11R has been associated with circulating tumour cells (Yadavalli et al. 2017). Evidence from breast cancer patients reveals that CREs are fundamental for metastasis formation. Despite rare, events associated with CREs homozygous deletion led to increased survival, indicating these events as less advantageous for tumour growth and metastasizing. While CREs amplification is linked with reduced survival. Chromatin accessibility re-wiring may control the transcriptome in either/both CREs edition, but not in a coding deletion. Altogether, these results suggest CREs as major regulators of metastasis, rather than the *CDH1* gene itself.

In summary, this study sheds light into the role of regulatory elements in gene expression, controlling development of metastasis. Additionally, our study shows that putative CRE lacking biochemical annotations may in fact act as gene-specific enhancer, but also control genome-wide chromatin accessibility, that consequently control genome-wide gene expression.

Acknowledgments

This work was supported by: 1) European Reference Network on Genetic Tumour Risk Syndromes (ERN GENTURIS) – Project ID No. 739547. ERN GENTURIS is partly co-funded by the European Union within the framework of the Third Health Programme “ERN-2016 – Framework Partnership Agreement 2017–2021”, (2) Doctoral Programme in Biomedicine - Faculty of Medicine, Uni-versity of Porto, to CSJ; (3) Doctoral Programme in Biotech Health in Institute of Biomedical Sciences, University of Porto to RBM, SL; (4) Doctoral Programme in Computational Sciences, Faculty of Sciences, University of Porto, to MF.

Funding

This work received funding from: (1) FEDER (Fundo Europeu de Desenvolvimento Regional) funds through the COMPETE 2020 - Operacional Programme for Competitiveness and Internationalisation (POCI), Portugal 2020 and by Fundação para a Ciência e a Tecnologia (FCT): “Solving the 3D Chromatin Structure of *CDH1* Locus to Identify Disease-Associated Mechanisms (3DChrome)”, PTDC/BTM-TEC/30164/2017; POCI-01-0145-FEDER-030164 and "Life-time Risk Estimations and Genetic Modifiers Of Hereditary Diffuse Gastric Cancer (LEGOH)", PTDC/BTM-TEC/6706/2020; (2) European Union’s Horizon 2020 research and innovation programme under grant agreement No 779257: “Solve-RD”; (3) FCT with salary support for PhD studentships to CSJ (Ref. SFRH/BD/140796/2018), RBM (Ref. SFRH/BD/145132/2019), SL (Ref. 2020.05773.BD) and MF (Ref. 2020.05763.BD).

References

Abascal F, Acosta R, Addleman NJ, et al (2020) Expanded encyclopaedias of DNA elements in the human and mouse genomes. Nature 583:699–710. https://doi.org/10.1038/S41586-020-2493-4

Artus J, Hadjantonakis A-K (2011) Generation of chimeras by aggregation of embryonic stem cells with diploid or tetraploid mouse embryos. Methods Mol Biol 693:37–56. https://doi.org/10.1007/978-1-60761-974-1_3

Ayollo D V., Zhitnyak IY, Vasiliev JM, Gloushankova NA (2009) Rearrangements of the actin cytoskeleton and E-cadherin-based adherens junctions caused by neoplasic transformation change cell-cell interactions. PLoS One 4:. https://doi.org/10.1371/JOURNAL.PONE.0008027

Banerji J, Rusconi S, Schaffner W (1981) Expression of a beta-globin gene is enhanced by remote SV40 DNA sequences. Cell 27:299–308. https://doi.org/10.1016/0092-8674(81)90413-X

Barski A, Cuddapah S, Cui K, et al (2007) High-resolution profiling of histone methylations in the human genome. Cell 129:823–837. https://doi.org/10.1016/J.CELL.2007.05.009

Benton ML, Talipineni SC, Kostka D, Capra JA (2019) Genome-wide enhancer annotations differ significantly in genomic distribution, evolution, and function. BMC Genomics 20:. https://doi.org/10.1186/S12864-019-5779-X

Bessa J, Tena JJ, De La Calle-Mustienes E, et al (2009) Zebrafish Enhancer Detection (ZED) vector: A new tool to facilitate transgenesis and the functional analysis of cis-regulatory regions in zebrafish. Developmental Dynamics 238:2409–2417. https://doi.org/10.1002/dvdy.22051

Buenrostro JD, Giresi PG, Zaba LC, et al (2013) Transposition of native chromatin for fast and sensitive epigenomic profiling of open chromatin, DNA-binding proteins and nucleosome position. Nat Methods 10:1213–1218. https://doi.org/10.1038/NMETH.2688

Buenrostro JD, Wu B, Chang HY, Greenleaf WJ (2015) ATAC-seq: A Method for Assaying Chromatin Accessibility Genome-Wide. Curr Protoc Mol Biol 109:21.29.1-21.29.9. https://doi.org/10.1002/0471142727.MB2129S109

Chen A, Beetham H, Black MA, et al (2014) E-cadherin loss alters cytoskeletal organization and adhesion in non-malignant breast cells but is insufficient to induce an epithelial-mesenchymal transition. BMC Cancer 14:. https://doi.org/10.1186/1471-2407-14-552

Curtis C, Shah SP, Chin SF, et al (2012) The genomic and transcriptomic architecture of 2,000 breast tumours reveals novel subgroups. Nature 486:346–352. https://doi.org/10.1038/NATURE10983

Dagenais M, Dupaul-Chicoine J, Douglas T, et al (2017) The Interleukin (IL)-1R1 pathway is a critical negative regulator of PyMT-mediated mammary tumorigenesis and pulmonary metastasis. Oncoimmunology 6:. https://doi.org/10.1080/2162402X.2017.1287247

Das SG, Romagnoli M, Mineva ND, et al (2016) miR-720 is a downstream target of an ADAM8-induced ERK signaling cascade that promotes the migratory and invasive phenotype of triple-negative breast cancer cells. Breast Cancer Res 18:. https://doi.org/10.1186/S13058-016-0699-Z

Desai RA, Gao L, Raghavan S, et al (2009) Cell polarity triggered by cell-cell adhesion via E-cadherin. J Cell Sci 122:905–911. https://doi.org/10.1242/JCS.028183

Elisha Y, Kalchenko V, Kuznetsov Y, Geiger B (2018) Dual role of E-cadherin in the regulation of invasive collective migration of mammary carcinoma cells. Sci Rep 8:. https://doi.org/10.1038/S41598-018-22940-3

Feingold EA, Good PJ, Guyer MS, et al (2004) The ENCODE (ENCyclopedia Of DNA Elements) Project. Science 306:636–640. https://doi.org/10.1126/SCIENCE.1105136

Gasperini M, Hill AJ, McFaline-Figueroa JL, et al (2019) A Genome-wide Framework for Mapping Gene Regulation via Cellular Genetic Screens. Cell 0:1–14. https://doi.org/10.1016/J.CELL.2018.11.029

Geeven G, Teunissen H, de Laat W, de Wit E (2018) peakC: a flexible, non-parametric peak calling package for 4C and Capture-C data. Nucleic Acids Res 46:E91. https://doi.org/10.1093/NAR/GKY443

Goto T, Hara H, Nakauchi H, et al (2016) Hypomorphic phenotype of Foxn1 gene-modified rats by CRISPR/Cas9 system. Transgenic Res 25:533–544. https://doi.org/10.1007/S11248-016-9941-9

Grosveld F, Van Staalduinen J, Stadhouders R (2021) Transcriptional Regulation by (Super)Enhancers: From Discovery to Mechanisms. Annu Rev Genomics Hum Genet 22:127–146. https://doi.org/10.1146/ANNUREV-GENOM-122220-093818

Inoue F, Ahituv N (2015) Decoding enhancers using massively parallel reporter assays. Genomics 106:159–164. https://doi.org/10.1016/J.YGENO.2015.06.005

Kimmel CB, Ballard WW, Kimmel SR, et al (1995) Stages of embryonic development of the zebrafish. Dev Dyn 203:253–310. https://doi.org/10.1002/AJA.1002030302

Krijger PHL, Geeven G, Bianchi V, et al (2020) 4C-seq from beginning to end: A detailed protocol for sample preparation and data analysis. Methods 170:17–32. https://doi.org/10.1016/J.YMETH.2019.07.014

Livak KJ, Schmittgen TD (2001) Analysis of relative gene expression data using real-time quantitative PCR and the 2(-Delta Delta C(T)) Method. Methods 25:402–408. https://doi.org/10.1006/meth.2001.1262

Lobe CG, Koop KE, Kreppner W, et al (1999) Z/AP, a double reporter for cre-mediated recombination. Dev Biol 208:281–292. https://doi.org/10.1006/dbio.1999.9209

Lu HH, Lin SY, Weng RR, et al (2020) Fucosyltransferase 4 shapes oncogenic glycoproteome to drive metastasis of lung adenocarcinoma. EBioMedicine 57:. https://doi.org/10.1016/J.EBIOM.2020.102846

Lupianez DG, Kraft K, Heinrich V, et al (2015) Disruptions of topological chromatin domains cause pathogenic rewiring of gene-enhancer interactions. Cell 161:1012–1025. https://doi.org/10.1016/j.cell.2015.04.004

Minnerop M, Kurzwelly D, Wagner H, et al (2017) Hypomorphic mutations in POLR3A are a frequent cause of sporadic and recessive spastic ataxia. Brain 140:1561–1578. https://doi.org/10.1093/BRAIN/AWX095

Panigrahi A, O'Malley BW (2021) Mechanisms of enhancer action: the known and the unknown. Genome Biol 22:. https://doi.org/10.1186/S13059-021-02322-1

Perry MW, Boettiger AN, Levine M (2011) Multiple enhancers ensure precision of gap gene-expression patterns in the Drosophila embryo. Proc Natl Acad Sci U S A 108:13570–13575. https://doi.org/10.1073/PNAS.1109873108

Pinheiro H, Carvalho J, Oliveira P, et al (2012) Transcription initiation arising from E-cadherin/*CDH1* intron2: a novel protein isoform that increases gastric cancer cell invasion and angiogenesis. Hum Mol Genet 21:4253–4269. https://doi.org/10.1093/HMG/DDS248

Ran FA, Hsu PD, Wright J, et al (2013) Genome engineering using the CRISPR-Cas9 system. Nat Protoc 8:2281–2308. https://doi.org/10.1038/NPROT.2013.143

São José C, Garcia-Pelaez J, Ferreira M, et al (2023) Combined loss of *CDH1* and downstream regulatory sequences drive early-onset diffuse gastric cancer and increase penetrance of hereditary diffuse gastric cancer. Gastric Cancer. https://doi.org/10.1007/S10120-023-01395-0

Sara Rocha (2020) Deconstructing Gastric Cancer Colonization: The interplay between immune cells, gastric cancer cells and extracellular vesicles

Stemmler MP, Hecht A, Kemler R (2005) E-cadherin intron 2 contains cis-regulatory elements essential for gene expression. Development 132:965–76. https://doi.org/10.1242/dev.01662

Stemmler MP, Hecht A, Kinzel B, Kemler R (2003) Analysis of regulatory elements of E-cadherin with reporter gene constructs in transgenic mouse embryos. Developmental Dynamics 227:238–245. https://doi.org/10.1002/dvdy.10301

Tedaldi G, Molinari C, José CS, et al (2021) Genetic and Epigenetic Alterations of *CDH1* Regulatory Regions in Hereditary and Sporadic Gastric Cancer. Pharmaceuticals (Basel) 14:. https://doi.org/10.3390/PH14050457

Van De Werken HJG, De Vree PJP, Splinter E, et al (2012) 4C technology: protocols and data analysis. Methods Enzymol 513:89–112. https://doi.org/10.1016/B978-0-12-391938-0.00004-5

Will AJ, Cova G, Osterwalder M, et al (2017) Composition and dosage of a multipartite enhancer cluster control developmental expression of Ihh (Indian hedgehog). Nat Genet 49:1539–1545. https://doi.org/10.1038/NG.3939

Yadavalli S, Jayaram S, Manda SS, et al (2017) Data-Driven Discovery of Extravasation Pathway in Circulating Tumor Cells. Sci Rep 7:. https://doi.org/10.1038/SREP43710

Supplementary table 1. Primers used for 4C-seq

Viewpoint	Sequence	Orientation
Exon 2	gaatttacgcagatttggggatc	Forward
Exon 2	aacggtgggctaggtcttgag	Reverse
CREa	cagggcaggcagggagatc	Forward
CREa	agtgagtgcctgttgtgtgttag	Reverse
CREb	ctttcctacgcaggctctgatc	Forward
CREb	cctagtggtagaaaggaggacac	Reverse
Promoter	GTGAACCCTCAGCCAATCAG	Forward
Promoter	TGGCTCACTAAGACCTGGGATC	Reverse

Supplementary table 3. Primers used for mouse enhancer assay

Name	Sequence	Orientation
CREa	GTTCTTAATTAAAGTGTTTTCCACCTTTGGTCA	Forward
CREa	CTTCTTAATTAAACTCTATCTGGCCTGGATAACTT	Reverse
CREb	CATCTTAATTAACTCCCAAAGTGCTGGGATTA	Forward
CREb	GATCTTAATTAATGGCACTGTGACACACAGAA	Reverse

Supplementary table 4. sgRNAs used for CRISPR-Cas9

Name	Sequence	Location	Cell line
Exon 2	TGCCAAGAAAGGTCGTAAAT	5'	MKN74
Exon 2	AGAGTGACGTCCACTTGCTC	3'	MKN74
CREa	ATGTAAGTTTCCCTAGGCCA	5'	MKN74
CREa	GGACCGGAACGGGTTTGTTG	3'	MKN74
CREb	GGATGTAAAGCAGTGAAACG	5'	MKN74
CREb	AAGGGTTCGGGGCGCGTGGT	3'	MKN74
CREa del	GTAAAGCAGTGAAACGGGGT	5'	MKN45
CREa del	AAGGGTTCGGGGCGCGTGGT	3'	MKN45
CREb del	AGGTTGATTGTAAATGGGCG	5'	MKN45
CREb del	GTGGTACTAGCATCAAAAAA	3'	MKN45

Supplementary table 5. Primers used for genotyping

Name	Sequence	Orientation
Exon 2	ACCGCTTCCCTTCTTCCAAG	Forward
Exon 2	TGACTCCCAGGCCCCATTAAG	Reverse
CREa	GAGTCTGGAAATGGGAGGGC	Forward
CREa	CAGCATCTATCTGGCCTGGA	Reverse
CREb	TACCAGCCAGTTGACATCCA	Forward
CREb	TCTGGGATTCTAGGGCTGAA	Reverse

Supplementary table 6. Primers used for methylation assay

Name	Sequence	Orientation
CREb	GGAGGATTTGGTGTTGTTTG	Forward
CREb	CTAACAAATTCAAAATCATCCC	Reverse
Promoter	GTTTTGTTTTGGGGAGGGGTT	Forward
Promoter	ACTCCAAAAACCCATAACTAACC	Reverse

Original paper 5

3D chromatin architecture re-wiring at the *CDH3*/*CDH1 loci* contributes to E-Cadherin to P-Cadherin expression switch in gastric cancer

Celina S. José[1,2,3], Carla Pereira[1,2], Marta Ferreira[1,2,4], Ana André[1], Hugo Osório[1], Irene Gullo[1,5,6,7], Fátima Carneiro[1,5,6,7], Carla Oliveira[1,2,8]

[1] Instituto de Investigação e Inovação em Saúde, Universidade do Porto, Porto, Portugal.

[2] Instituto de Patologia e Imunologia Molecular da Universidade do Porto, Porto, Portugal

[3] Doctoral Programme in Biomedicine, Faculty of medicine, University of Porto, Porto, Portugal

[4] Dept. Computer Science Faculty of Science, University of Porto, Porto, Portugal

[5] Institute of Molecular Pathology and Immunology of the University of Porto, Porto, Portugal

[6] Department of Pathology, Faculty of Medicine, University of Porto, Porto, Portugal

[7] Department of Pathology, Centro Hospitalar Universitário São João, Porto, Portugal

[8] Faculty of Medicine of the University of Porto, Porto, Portugal

* Correspondence: carlaol@ipatimup.pt; +351-225570785

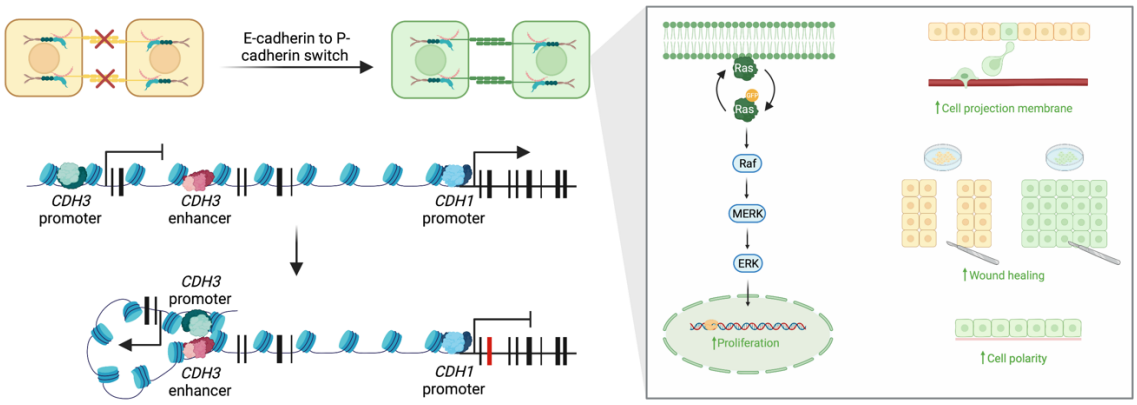
Biology 2023,12, 803 (appendix E).

Author statement

Celina São José: Investigation, methodology, conceptualization, formal analysis, writing – original draft, writing – review and editing.

Unveil the molecular mechanisms behind E- to P-cadherin switch in gastric cancer

Cadherins are cell–cell adhesion molecules, fundamental for cell architecture and polarity. E-cadherin to P-cadherin switch can rescue adherens junctions in epithelial tumours. Herein, we disclose a mechanism for E-cadherin to P-cadherin switch in gastric cancers. *CDH1* and *CDH3* mRNA expression was obtained from 42 gastric tumours' RNA-seq data. CRISPR-Cas9 was used to knock out *CDH1* and a putative regulatory element. *CDH1*-depleted and parental cells were submitted to proteomics and enrichment GO terms analysis; ATAC-seq/4C-seq with a *CDH1* promoter view- point to assess chromatin accessibility and conformation; and RT-PCR/flow cytometry to assess *CDH1*/E-cadherin and *CDH3*/P-cadherin expression. In 42% of gastric tumours analysed, *CDH1* to *CDH3* switch was observed. *CDH1* knockout triggered *CDH1*/E-cadherin complete loss and *CDH3*/P-cadherin expression increase at plasma membrane. This switch, likely rescuing adherens junctions, increased cell migration/proliferation, commonly observed in aggressive tumours. E- to P-cadherin switch accompanied increased *CDH1* promoter interactions with *CDH3*–eQTL, absent in normal stomach and parental cells. *CDH3*–eQTL deletion promotes *CDH3*/*CDH1* reduced expres- sion. These data provide evidence that loss of *CDH1*/E-cadherin expression alters the *CDH3* locus chromatin conformation, allowing a *CDH1* promoter interaction with a *CDH3*-eQTL, and promoting *CDH3*/P-cadherin expression. These data highlight a novel mechanism triggering E- to P-cadherin switch in gastric cancer.



Abstract

Cadherins are cell–cell adhesion molecules, fundamental for cell architecture and polarity. E-cadherin to P-cadherin switch can rescue adherens junctions in epithelial tumours. Herein, we disclose a mechanism for E-cadherin to P-cadherin switch in gastric cancers. *CDH1* and *CDH3* mRNA expression was obtained from 42 gastric tumours' RNA-seq data. CRISPR-Cas9 was used to knock out *CDH1* and a putative regulatory element. *CDH1*-depleted and parental cells were submitted to proteomics and enrichment GO terms analysis; ATAC-seq/4C-seq with a *CDH1* promoter view- point to assess chromatin accessibility and conformation; and RT-PCR/flow cytometry to assess *CDH1*/E-cadherin and *CDH3*/P-cadherin expression. In 42% of gastric tumours analysed, *CDH1* to *CDH3* switch was observed. *CDH1* knockout triggered *CDH1*/E-cadherin complete loss and *CDH3*/P-cadherin expression increase at plasma membrane. This switch, likely rescuing adherens junctions, increased cell migration/proliferation, commonly observed in aggressive tumours. E- to P-cadherin switch accompanied increased *CDH1* promoter interactions with *CDH3*–eQTL, absent in normal stomach and parental cells. *CDH3*–eQTL deletion promotes *CDH3*/*CDH1* reduced expres- sion. These data provide evidence that loss of *CDH1*/E-cadherin expression alters the *CDH3* locus chromatin conformation, allowing a *CDH1* promoter interaction with a *CDH3*-eQTL, and promoting *CDH3*/P-cadherin expression. These data highlight a novel mechanism triggering E- to P-cadherin switch in gastric cancer.

Introduction

Cadherins are tissue-specific transmembrane glycoproteins that play a key role in stable cell–cell adhesion, cellular polarization, maintenance of tissue architecture, and embryogenesis [1,2]. E-cadherin and P-cadherin are encoded by *CDH1* and *CDH3* genes, respectively, and present a great degree of sequence and gene structure homology [3]. Despite their homology, both proteins display tissue-specific expression patterns and functions, being often mutually exclusive at the cellular level [3]. While E-cadherin is widely expressed in epithelial tissues [4], P-cadherin expression is restricted to organ regions, such as myoepithelial cells of the breast or mesothelia [5]. Whereas E-cadherin acts as a tumour suppressor in most epithelial tissues, P-cadherin de novo expression or upregulation has been linked to mammary cells' proliferation and migration in mice [6], tumour-promotion in gastric [7,8] and breast cancer [9,10], and worse prognosis in breast cancer [11–13].

Aberrant cadherin isoform switching occurs in tumours with implications for patients' survival [14,15]. E-cadherin to N-cadherin switch is a mark of epithelial to mesenchymal transition, in which differentiated epithelial cells, expressing E-cadherin, transition to dedifferentiated mesenchymal-like cells, expressing N-cadherin, and are prone to invade and metastasize [16]. Evidence of E-cadherin to P-cadherin switching has been observed in both gastric [17] and breast cancer, with implications for survival in the latter [11].

Mechanisms for E-cadherin to P-cadherin switching have been suggested, mainly through MiR-34a/WNT1 cascade [18], but fail to explain all observed switching phenotypes. As *CDH1* and *CDH3* are juxtaposed genes, derived from gene duplication, and both belong to the same topological-associating domain (TAD) [19], their expression is most likely co-regulated and dependent on the tissue-specific 3D chromatin architecture. The 3D chromatin architecture tightly regulates gene expression in time and space through physical interactions involving promoters and regulatory elements located in non-coding regions in their vicinity [20]. These regulatory networks are generally tissue-specific [21], and their impairment may interfere with the normal 3D chromatin architecture, leading to genes' misexpression patterns, causing or aggravating disease [22,23].

Herein, we aimed to characterize a potential 3D chromatin architecture re-wiring process, occurring at the *CDH3–CDH1 loci*, that may contribute to E-cadherin to P-cadherin switch in gastric cancer.

Materials and Methods

Samples

Gastric cancers from 43 patients were selected for the study, displaying equivalent male to female ratio, range of age at diagnosis, distribution of tumour stage, and histological type. Clinicopathological features were collected from clinical records. The present project was approved by the Ethical Committee of Centro Hospitalar Universitário de São João on 16 March 2017, with internal reference CES072017.

RNA Extraction

Tumour areas with at least 75% tumour cells were selected for RNA extraction using MagMax FFPE DNA/RNA Ultra Kit (Applied Biosystems, Life technologies, Carlsbad, CA, USA), according to the manufacturer's instructions. Briefly, sections of paraffin-embedded tumour samples were de-paraffinized and protease K-digested. RNA was bonded to magnetic beads, pelleted against a magnetic stand, and the supernatant containing RNA washed and eluted. Concentrations were determined using Qubit® 2.0 Fluorometer (In- vitrogen, Life Technologies Europe BV, Porto, Portugal) with RNA high sensitivity assay, and integrity was assessed with a 2100 Bioanalyzer Instrument (Agilent Technologies, Instruments Soquimica, Lisboa, Portugal).

RNA-Seq and Whole Transcriptome Sequencing Bioinformatics Analysis

RNA was sequenced with Ion AmpliSeqTM Transcriptome Human Gene Expression Panel kit in an Ion ChefTM 550 NGS instrument. We obtained sequencing reads aligned to the GRCh37 human genome using TMAP [24] and annotated with USCS GRCh37. Only canonical transcripts were selected. *CDH1* and *CDH3* mRNA expression was considered positive for a threshold of 10 TPM or higher. Processed total RNA-seq from normal stomach tissue samples ENCSR853WOM, ENCSR471RUK, and ENCSR000AFI was collected from ENCODE [25–27].

Cell Lines and Culture

MKN74 human cell line was acquired from the JCRB Cell Bank, cultured in RPMI medium (GIBCO, Fisher Scientific, Porto Salvo, Portugal) supplemented with 10% fetal bovine serum (GIBCO, Fisher Scientific, Porto Salvo, Portugal) and maintained at 37 ◦C and 5% CO2 in a high-humidity atmosphere. STR analysis was used to confirm cell identification. Cells were free from mycoplasma contamination.

Generation of CRISPR-Cas9 E-Cadherin Deletion

Single guided RNAs (sgRNAs) were designed to target the *CDH1* exon 2 and impair E-cadherin function, resourced using the Zhang Lab tool, currently available at Benchling online platform. Individual sgRNAs (Sigma-Aldrich, Merck Life Science, Darmstadt, Germany) (Supplementary Table S1) were cloned into pSpCas9(BB)-2A-Puro (PX459) V2.0 plasmid (Addgene 62988), in BbsI restriction site (New England Biolabs), according to the method of Ran and colleagues [28]. Each sgRNA-plasmid was transformed into stbl3 cells (Invitrogen), and colonies were screened by PCR and Sanger sequenced (Supplementary Table S1). sgRNAs were transfected in MKN74 cell line. Briefly, cells were seeded in 12-well plates and grown for 24 h, and pairs of plasmids were transfected with lipofectamine 3000 (Invitrogen), according to the manufacturer's instructions. Puromycin (Merck, Merck Life Science, Darmstadt, Germany) treatment started at 48 h and was renewed every 72 h, until non-transfected cells were dead.

Genotyping of Edited Clones

gDNA was extracted using NZY Tissue gDNA Isolation kit (NZYTech, Lisboa, Portu- gal) according to the manufacturer's protocol. gDNA was amplified with multiplex PCR kit (Qiagen, Germantown, MD, USA) and primers flanking the edition sites (Supplementary Table S1, Sigma-Aldrich). Edition was confirmed by Sanger, sequenced using BigDye Terminator v.3.1 cycle sequencing kit (Thermo Fisher Scientific, Waltham, MA, USA, EUA) on an ABI-3130 Genetic Analyzer (Applied Biosystems Europe B.V., Porto, Portugal).

mRNA and Protein Expression Analysis

CDH1/3 mRNA expression was assessed by qPCR in triplicate. Briefly, RNA was extracted using mirVana RNA Isolation Kit (Invitrogen), according to manufacturer's pro- tocol. cDNA was synthesized using 1 µg of template RNA, and SuperScriptII reverse transcriptase (Invitrogen) was used to synthesize cDNA, according to the manufacturer's instructions. *CDH1* mRNA expression was analysed by qPCR with KAPA PROBE FAST qPCR Master Mix (2X) Kit (Sigma-Aldrich) and probes for *CDH1* (Hs.PT.58.3324071, IDT), *CDH3* (Hs.PT.51.5028751, IDT), and 18S (custom assay, IDT) as endogenous control. Re- actions were sequenced on a 7500 Real-Time PCR System (Applied Biosystems). Relative expression was normalized for the endogenous 18S control and quantified using the $2^{-\Delta\Delta Ct}$ method [29].

E- and P-cadherin expression was assessed by flow cytometry in triplicate. Cells were blocked for 30 min with 3% bovine serum albumin–phosphatase buffer saline (BSA, NZYTech) and incubated with primary mouse antibodies HECD-1 or

P-cadherin (1:100 dilution; 1 h at 4 ◦C; Invitrogen), followed by secondary anti-mouse Alexa Fluor 647 antibody (1:500; 45 min at 4 ◦C; Invitrogen). FACS ARIA (BD Biosciences, Franklin Lakes, NJ, USA) was used to measure mean fluorescence, and Flow Jo version 10 software was used to analyse the data.

Statistical Analysis

GraphPad Prism version 7.00 software (GraphPad Software Inc, San Diego, CA, USA) was used to perform statistical analysis. Student's t-test was used for comparison anal- ysis, assuming equal variance between clones and wild-type samples. Differences were considered significant for p-value <0.05.

Immunofluorescence

Cells were fixed for 20 min in 4% paraformaldehyde (Merck Life Science, Darmstadt, Germany) and blocked with 5% BSA (NZYTech) for 30 min. Cells were incubated with primary mouse antibody anti-E-cadherin (24E10) or anti-β-catenin (1:50; ON; 4 ◦C, Cell Signaling Technology, Danvers, MA, EUA) and primary rabbit monoclonal antibody anti- P-cadherin (#2130; 1:50; ON; 4 ◦C, Cell Signaling Technology), followed by secondary antibodies Alexa Fluor 488 Donkey Anti-Mouse (1:500; 60 min; Life Technologies) or Alexa Fluor 594 anti-rabbit (1:500, 60 min; Life Technologies). Cell nuclei were stained with DAPI (1 µg/mL in PBS; 5 min incubation; Sigma, Merck Life Science, Darmstadt, Germany). All coverslips were mounted using Vectashield mounting media (Vector Laboratories, Newark, CA, USA) and cells were analysed by fluorescence microscopy (Imager.Z1, AxioCam fluorescence microscope or Eclipse TE-2000, both from Zeiss, Oberkochen, Germany) using AxioVision software (Rockville, White Plains, NY, USA).

Protein Extraction

Cells were lysed with lysis buffer containing 1% (v/v) Triton X-100 and 1% (v/v) IGEPAL in PBS, supplemented with phosphatase (Sigma-Aldrich) and protease inhibitor cocktails (Roche, Mannheim, Germany) at 4 ◦C. Modified Bradford Assay (Bio-Rad, Her- cules, CA, USA) was used to quantify total protein content of the samples.

Proteomics Sample Processing

Proteomics was performed according to solid-phase-enhanced sample-preparation (SP3) protocol [30]. Briefly, proteins were solubilized for 10 min with 100 mM Tris pH 8.5, 1% sodium deoxycholate, 10 mM tris(2-carboxyethyl)

phosphine (TCEP), and 40 mM chloroacetamide at 95 ◦C at 1000 rpm. Proteins were digested with Trypsin/LysC (2 µg; Thermo Fisher Scientific, Waltham, MA, USA, EUA) overnight at 37 ◦C and 1000 rpm. Peptides were cleaned up and quantified.

Protein Identification and Quantification, and Enrichment GO Terms Analysis

Proteome Discoverer 2.1 software was used for identification (UniProt/SwissProt human database using Mascot (Version 2.5.1, Matrix Sciences)) and/or quantification of proteins. Carbamidomethylation of cysteines was set as fixed modification and methionine oxidation as variable modification, with trypsin as selected enzyme. High confidence master proteins with at least two unique peptides and without contaminants were consid- ered. Differentially expressed proteins were selected considering adjusted p-value ≤0.05, abundance ratio ≤0.67 (downregulated) or ≥1.5 (upregulated), and concordance between replicates. ClusterProfiler (v.4.4.4) R package was used for enrichment GO terms analysis (q-value < 0.05). Statistics were performed using R (v.4.2.0).

ATAC-Seq

ATAC-seq data from normal stomach tissue were collected from phase 3 ENCODE project [31]. ATAC-seq libraries were generated from MKN74 cell line and *CDH1* deletion clone, as described previously [32]. A total of 5 × 104 cells were lysed and chromatin was tagmented using Tn5 transposase and purified with MiniElute PCR purification kit (Qia- gen), following the manufacturer's protocol. The appropriate number of PCR cycles was determined using SYBR Green qPCR. Transposed DNA fragments were amplified by PCR and purified using QIAquick PCR purification kit (Qiagen), following the manufacturer's protocol. Quality control of ATAC-seq libraries was accessed by tapestation, and samples were sequenced in Illumina HiSeqX technology according to standard protocols. ATAC-seq peaks were mapped resourcing to an ENCODE-DCC ATAC-seq bioinformatics pipeline with default parameters for adaptors trimming (cutadapt: -e 0.1 -m 5), alignment (bowtie), quality filtering for mapq < 30, removal of duplicates and mitochondrial reads (piccard), peak calling (macs2), and replicate statistical analysis (idr, p-value < 0.05).

4C-Seq

Fresh gastric tissue from bariatric surgeries was collected and washed in HBSS 1x (GIBCO, Fisher Scientific, Porto Salvo, Portugal). Gastric mucosa cells were scraped and enzymatically dissociated with collagenase type I (Merck Life Science, Darmstadt, Ger- many), dispase (Merck), trypsin inhibitor (Sigma-Aldrish, Merck Life Science, Darmstadt, Germany), BSA (Nzytech, Lisboa,

Portugal), dithiothreitol (Invitrogen Life Technologies Europe BV, Porto, Portugal), and HBSS 1× (GIBCO, Fisher Scientific, Porto Salvo, Portugal) at 37 ◦C and 150 rpm for at least 1 h. From these, 4C-seq libraries were generated, as previously described [33,34], containing 1 × 107 cells, cross-linked in 2% paraformaldehyde and lysed. Nuclei suspensions were digested with DnpII (New England Biolabs, EVRY cedex, Évry-Courcouronnes, France) as primary and Csp6I (New England Biolabs) as secondary restriction enzymes and re-circularized with T4 DNA Ligase (Thermo Fisher Scientific). Then, 4C-seq libraries were purified using Amicon Ultra-15 10 kDa (MWCO) (Millipore) and PCR-amplified with 3.2 µg per reaction (Supplementary Table S2). The *CDH1* viewpoint was designed to target chr16: 68737018-68737215, hg38. Samples were paired-end sequenced on Illumina HiSeqX technology (150 bp reads) according to standard protocols. *CDH1* interactions on a genome scale were mapped from sequenced 4C libraries, using a bioinformatics pipeline based on Pipe4C [34] and PeakC [35] with window size 2, alpha fdr 0.1, and minimal distance 500. Collection of bariatric surgeries' material was approved by the Ethical Committee of Centro Hospitalar Universitário de São João on 8 May 2020, with internal reference CE 305-19.

Results

CDH1 to CDH3 mRNA Expression Switch Is a Frequent Event in Gastric Cancer

To understand the frequency of and explore potential mechanisms which underly E- to P-cadherin switching in gastric cancer, we started by analysing RNA-seq data from 43 gastric cancers to evaluate *CDH1* and *CDH3* mRNA expression. We found that *CDH1* downregulation and *CDH3* upregulation are the most frequent combination of events in this cohort, occurring in 42% of all cases, followed by expression of both cadherins (33%) (**Figure 1A**). Downregulation of both cadherins occurred in 19% of tumours, while only 7% of all tumours expressed *CDH1* and lacked *CDH3* expression (**Figure 1A**). In this cohort, 75% of the tumours expressed *CDH3* (**Figure 1A, upper panels**), a feature that is acquired in tumours, as normal stomach epithelial cells do not express *CDH3* (**Figure 1B**). In 42% of all cases, there was a *CDH1* to *CDH3* expression switch (**Figure 1A, upper left panel**).

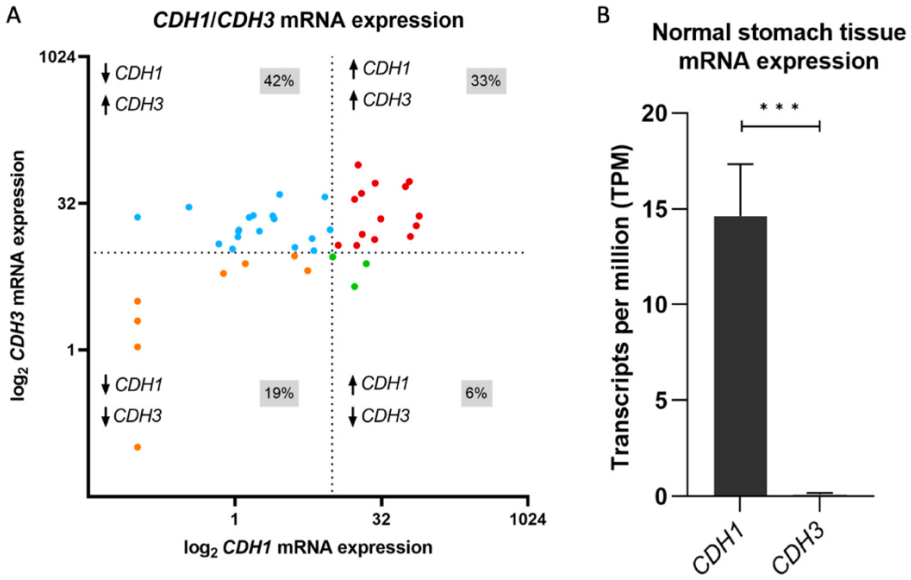


Figure 1. CDH1 and CDH3 mRNA expression in gastric tissues. *(A) Distribution of CDH1 and CDH3 mRNA expression in gastric cancer tumours. Blue dots represent gastric cancers with low CDH1 mRNA expression in gastric cancer tumours. Blue dots represent gastric cancers with low CDH1 and high CDH3 expression, red dots represent gastric cancers with high expression for both CDH1 and high CDH3 expression, red dots represent gastric cancers with high expression for both CDH1 and CDH3, orange dots represent gastric cancers with low CDH1 and CDH3 expression, green dots and CDH3, orange dots represent gastric cancers with low CDH1 and CDH3 expression, green dots represent gastric cancers with high CDH1 and low CDH3 expression. Dashed lines represent 10 TPM, represent gastric cancers with high CDH1 and low CDH3 expression. Dashed lines represent 10 TPM, the threshold for low or high expression. Data shown in transcripts per million (TPM). (B) CDH1 the threshold for low or high expression. Data shown in transcripts per million (TPM). (B) CDH1 and and CDH3 mRNA expression in stomach normal tissue from ENCODE [25–27]. Data shown in CDH3 mRNA expression in stomach normal tissue from ENCODE [25–27]. Data shown in transcripts transcripts per million (TPM) (*** p value ≤ 0.001).*

E-cadherin loss of function triggers E-cadherin to P-cadherin switch in gastric cancer

Given the abundance of gastric cancers bearing simultaneous loss of *CDH1* expression and *CDH3* upregulation, we tested whether depleting *CDH1* would trigger *CDH3* upregulation.

We used sgRNAs flanking *CDH1* exon 2 (**Supplementary Table S1**), and successful removal of exon 2 was confirmed by Sanger sequencing in the MKN74 cell line (**Supplementary Table S1**). This clone carried a homozygous deletion encompassing the *CDH1* exon 2 (**Figure 2A**) and presented complete loss of both *CDH1* mRNA and E-cadherin cadherin protein expression (**Figure 2B**). As a consequence of this deletion, both *CDH3* mRNA and P-cadherin protein expression increased by 60% (p-value = 0.05 and p-value = 0.037, respectively) (**Figure 2C**). The immunofluorescence data highlight the P-cadherin expression increase at the plasma membrane upon E-cadherin expression loss, likely rescuing adherens junctions (**Figure 2D**). Indeed, P-cadherin co-localized with adhesion partner β-catenin at the cell membrane in the *CDH1* deletion clone, while the parental cell line displayed lower (and heterogeneous) P-cadherin, as was the case in the cytoplasm (**Figure 2D**). Altogether, (**Figure 2D**). Altogether, these results support that depletion of E-cadherin, due to gene deletions, can trigger P-cadherin switch.

E-cadherin expression loss triggers downregulation of adhesion complex partners and adhesion related pathways

To further dissect the impact of E-cadherin loss, we used proteomics to evaluate the differential proteome between the parental cell line and the *CDH1* deletion clone. A differential proteome between the parental cell line and the *CDH1* deletion clone. A total of 4579 proteins were identified, from which 41 were downregulated and 39 were upregulated in the *CDH1* deletion clone (**Supplementary Table S3**). The E- to P-cadherin switch was confirmed in the *CDH1* deletion clone, with downregulation of the former (0.37-fold, p-value = 0.0002) and upregulation of the latter (1.99-fold, p-value = 0.018) (0.37-fold, p-value = 0.0002) (**Figure 2D**). In addition to E- and P-cadherin expression, other adhesion complex partners became impaired in the *CDH1* deletion clone, namely, *CTNNA1* (αE-catenin) and *CTNNB1* (β-catenin), while others remain unchanged, namely, *CTNND1* (p120) (**Figure 2D**). Gene ontology analysis revealed several adhesion-associated pathways downregulated upon *CDH1* deletion (**Figure 2E, Supplementary Table S4**), encompassing ‘adherens junction’, ‘cadherin binding’, ‘focal adhesion assembly’, and catenin- and actin-associated pathways (**Figure 2E, Supplementary Table S4**). In contrast, the ‘establishment and maintenance of cell polarity’ pathway became upregulated, likely revealing the rescue by P-cadherin. Events associated with migration, such as ‘cell projection membrane’ and ‘wound healing’, were found upregulated in

cells with E- to P-cadherin switch (**Figure 2E, Supplementary Table S4**). Downregulation of immune-associated pathways 'macrophage chemotaxis' and 'regulation of leucocyte migration' and upregulation of oncogenic pathway 'Ras protein signal transduction' were also found upon *CDH1* deletion (**Figure 2E, Supplementary Table S4**).

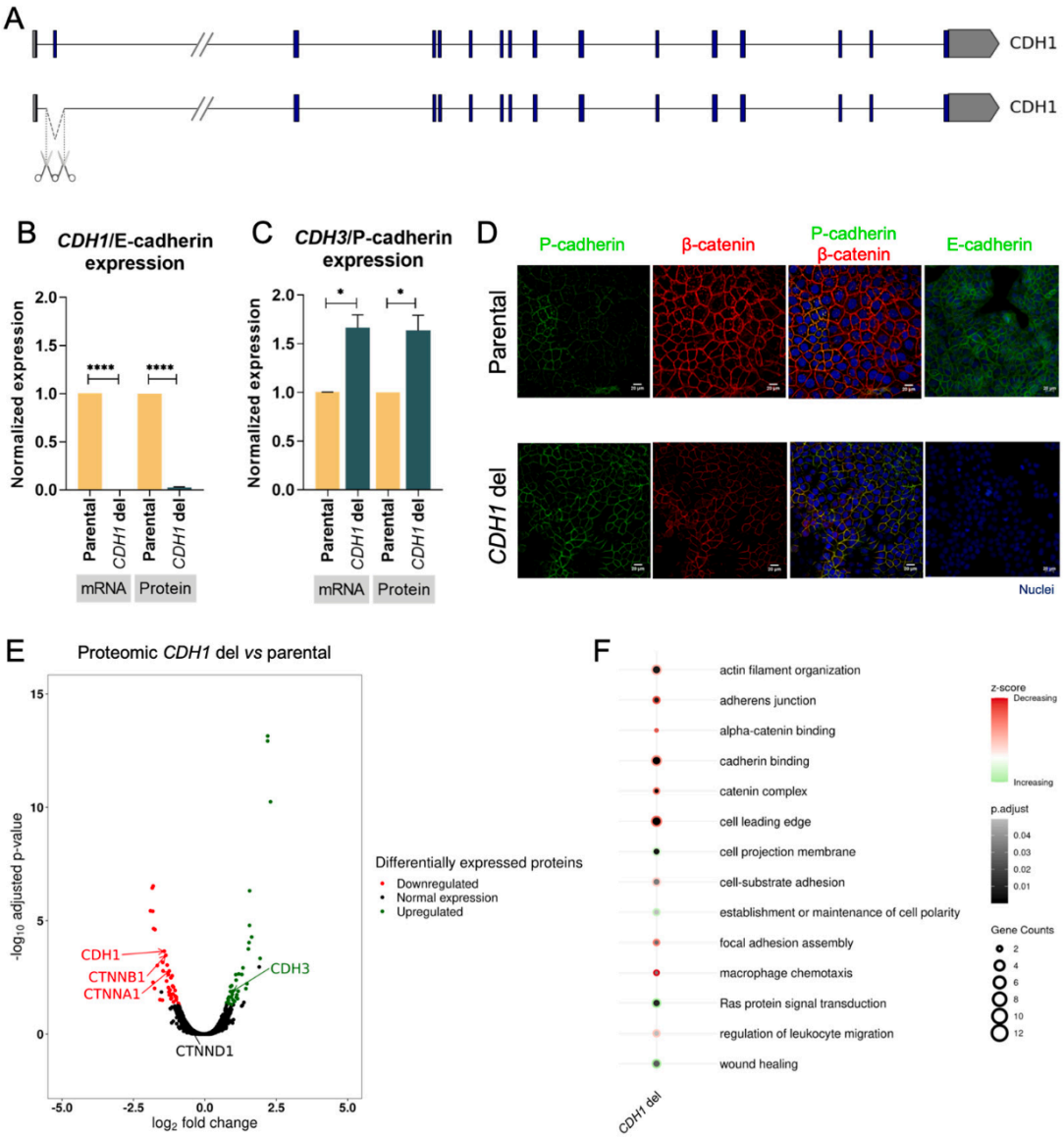


Figure 2. Characterization of E- to P-cadherin switch. *(A) CRISPR-Cas9 design strategy. (B) Characterization of CDH1 deletion in CDH1/E-cadherin mRNA and protein expression. Values represented as mean ± SEM. Protein expression represented as median fluorescence intensity (MFI) (**** p value ≤ 0.0001). (C) Characterization of CDH1 deletion in CDH3/P-cadherin mRNA and protein expression. Values represented as mean ± SEM. Protein expression represented as median fluorescence intensity (MFI) (* p value ≤ 0.05). (D) Immunofluorescence of parental cells (top) and CDH1 deletion cells (down) for P-cadherin, β-catenin, and E-cadherin. Nuclei were stained with DAPI (blue), and white scale bars represent a distance of 20 µm. (E) Volcano plot illustrating differentially expressed proteins in CDH1 deletion vs. parental cells. (F) Enrichment analysis of biological biological processes, molecular function, and cellular components in upregulated and downregulated proteins.*

3D chromatin architecture re-wiring at the CDH3-CDH1 loci contributes to E-cadherin to P-cadherin switch in gastric cancer

In an attempt to identify the mechanism triggering the E- to P-cadherin switch, upon *CDH1* depletion, we analysed regulatory elements within the *CDH3–CDH1* loci that may control expression of cadherins in the stomach. We mined the Ensembl database and found an eQTL located within the *CDH3* intron 2 (chr16: 68661867-68662578 and chr16: 68663306-68670017, hg38) (**Figure 3A**). Normal epithelial cells from bariatric surgeries, MKN74 parental cells as well as *CDH1* deletion cells, were profiled for *CDH1*-promoter interactions with the eQTL using 4C-seq, and for accessibility of chromatin using ATAC-seq. Normal stomach accessible chromatin was also collected from the ENCODE project and analysed [31]. In normal stomach, which expresses *CDH1* but not *CDH3*, the chromatin was not accessible at the *CDH3* promoter nor at the *CDH3*-eQTL, but was accessible at the *CDH1* promoter (**Figure 3A, ATAC-seq normal stomach**). Neither the *CDH3* promoter nor the *CDH3*–eQTL interacted with the *CDH1* promoter (**Figure 3A, 4C-seq normal stomach**). In contrast, upon *CDH1* deletion, availability of the chromatin increased around the eQTL sequence and was maintained as available at the *CDH1* and *CDH3* promoters (**Figure 3A, ATAC-seq parental vs. ATAC-seq *CDH1* deletion**). Additionally, new interactions were established between the *CDH1* promoter and the *CDH3*–eQTL region (**Figure 3A, 4C-seq parental vs. 4C-seq *CDH1* deletion**). These chromatin architecture changes seem to indicate that, upon *CDH1*/E-cadherin depletion, the *CDH1* promoter interacts with the *CDH3* eQTL, potentially leading to *CDH3*/P-cadherin expression increase (**Figure 2B–D**). If this is the case, deleting the *CDH3*–eQTL in a cell line expressing both *CDH3* and *CDH1* could provide information on the expression regulatory role of the *CDH3*-eQTL. To address this, we CRISPR-Cas9 deleted the eQTL in parental MKN74 cells. This led to a 50% decrease in both *CDH3* and *CDH1* mRNA expression (p-value = 0.0196 and p-value = 0.0084, respectively), and a 30% decrease in P-cadherin and E-cadherin protein expression (p-value = 0.035 and p-value = 0.0486, respectively) as compared to parental MKN74 cells (**Figure 3B**). These results pinpoint *CDH3*–eQTL as a potential enhancer of both *CDH3* and *CDH1* in this cell line. Altogether, these data provide evidence that loss of *CDH1*/E-cadherin expression alters the chromatin conformation at the *CDH3* locus, allowing interaction of the *CDH1* promoter with the *CDH3*–eQTL, which seems to potentiate *CDH3*/P-cadherin expression.

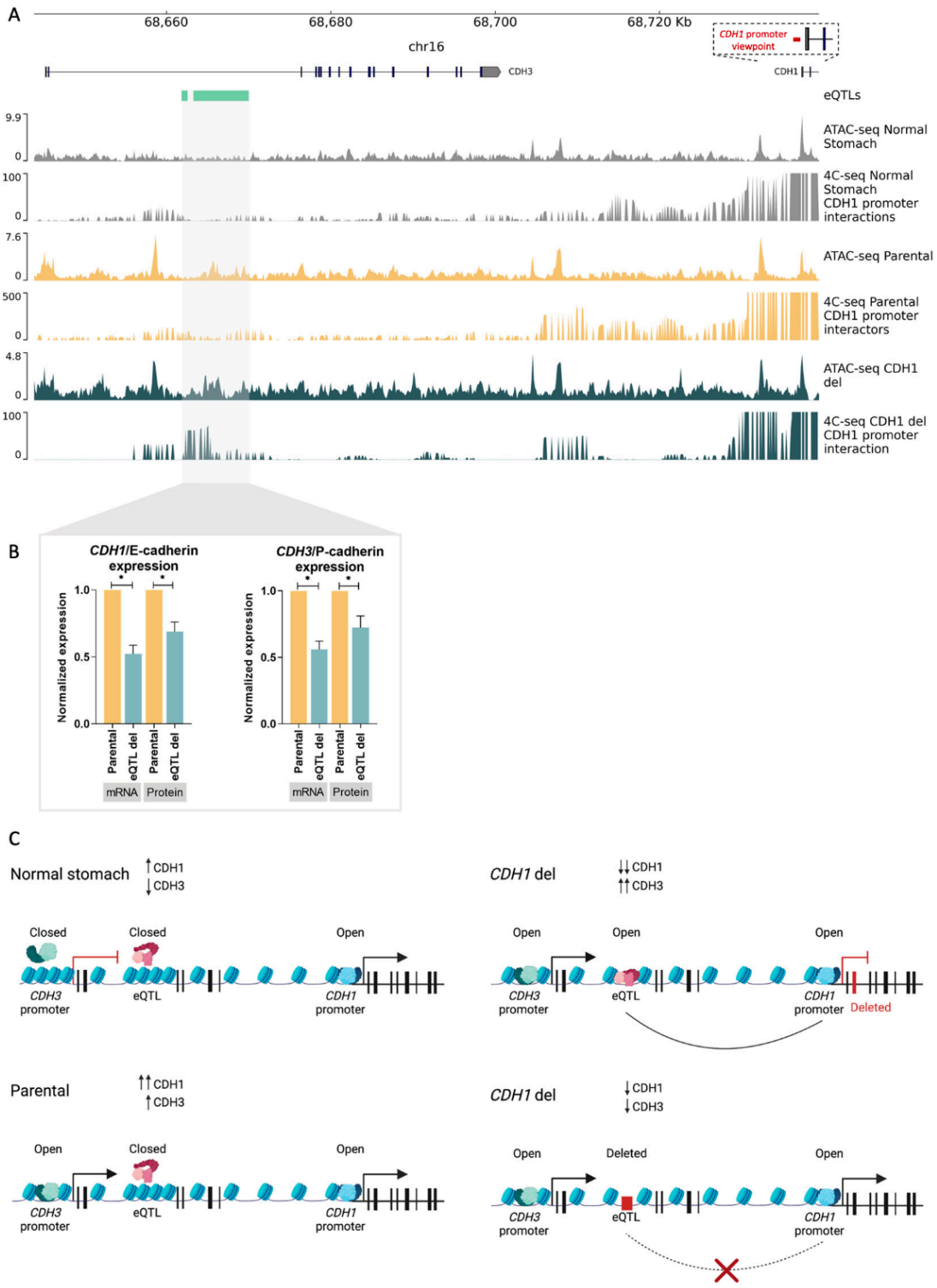


Figure 3. The 3D chromatin architecture in E- to P-cadherin switch. *(A) Accessible chromatin and CDH1-promoter interactors in normal stomach tissue, parental MKN74 cells, and CDH1 deletion CDH1 promoter viewpoint is shown in red. (B) Characterization of CDH3–eQTL deletion in CDH1/E-cadherin and CDH3/P-cadherin mRNA, and protein expression. Values represented as mean ± SEM. Protein expression represented as median fluorescence intensity (MFI) (* p value ≤ 0.05). (C) Model of the role of 3D chromatin architecture in E- to P-cadherin switch. Red rectangles represent deletions.*

Discussion

Herein, we aimed to characterize a potential 3D chromatin architecture re-wiring process, occurring at the *CDH3–CDH1 loci*, that may explain the *CDH1*/E-cadherin to *CDH3*/P-cadherin switch observed in gastric cancer. First, we demonstrated that 75% of the gastric cancers analysed overexpressed *CDH3*. Additionally, we observed that, in 42% of all cases, *CDH3* overexpression occurred concomitantly with *CDH1* downregulation, representing an E- to P-cadherin switch. We also found that an eQTL within the *CDH3* intron 2, acting as a regulatory element, was able to modulate both *CDH1* and *CDH3* expression, and that it interacted with the *CDH1* promoter upon *CDH1* loss of expression. To the best of our knowledge, this body of results highlights a novel mechanism which triggers E- to P-cadherin switch in gastric cancer.

Our data suggest that E- to P-cadherin switch is a common event in gastric cancers. The impact of P-cadherin overexpression in cancer is still controversial, as this molecule seems to act as tumour-promoting and a marker of a bad prognosis in several cancer types—namely, in breast, glioblastoma, and gastric cancer [36]—but may also act as a marker of a good prognosis—namely, in colon cancer [37]. Breast cancer cells presenting E- to P-cadherin switch, and those expressing both cadherins, reveal increased migration capacity in vitro [38], while breast cancer epithelial cells expressing E-cadherin and lacking P-cadherin expression, resembling the normal breast, and cells lacking both cadherins display reduced migration [38].

Herein, we created a model of E- to P-cadherin switch by using CRISPR-Cas9 editing to deplete *CDH1*. This model presented reduced expression of adhesion complex partners, but maintained β-catenin expression at adherens junctions, which co-localized with P-cadherin at the cell membrane. This likely represents an adherens junctions rescue by P-cadherin in E-cadherin-depleted cells, as previously reported [39]. Evidence for E- to P-cadherin switch has been linked to invasive lobular breast cancer with tubular elements, which show reduced E-cadherin expression, normal β-catenin expression, and P-cadherin upregulation, highlighting rescue of adherens junctions in breast tubular elements when E-cadherin is lost [39].

Our data on differential protein expression and gene ontology analysis in the *CDH1* deletion clone, in comparison to an isogenic parental *CDH1* cell line, highlighted downreg- ulation of adhesion-, catenin- and actin-related pathways, as well as upregulation of cell polarity, all associated with E-cadherin main functions. Despite the fact that E-cadherin defi- ciency weakens cell-adhesion, cell polarity regulation further supports P-cadherin rescue of adherens junctions (Christgen et al. 2020). Our gene ontology data also support increased migration potential of cells with E- to P-cadherin switch, namely, through activation of the ‘wound healing’ pathway, also described for breast cancer cells [38]. Our findings also suggest that cells with E- to P-cadherin switch may have increased

proliferation through the activation of Ras oncogenic signalling pathways. As in many published reports, the abovementioned findings support that E- to P-cadherin switch is associated with more aggressive tumour phenotypes.

In the present study, we used the abovementioned cell model to identify the mechanism responsible for the E- to P-cadherin switch. We raised the hypothesis that regulatory elements in the *CDH3*/*CDH1 loci* could play a role in the observed E- to P-cadherin expres- sion switch, as gene expression is tightly regulated through the physical proximity enabled by the 3D-chromatin architecture (Andrey and Mundlos 2017) [20]. We demonstrated that, upon *CDH1* depletion triggered by a gene deletion, the chromatin re-shuffles at the *CDH3*/*CDH1 loci*, and its promoter becomes physically bound to an eQTL overlapping accessible chromatin located in *CDH3* intron 2; this associates with increased *CDH3*/P-cadherin expression. The absence of interactions between the *CDH1* promoter and this eQTL in normal stomach epithelial tissue may explain why *CDH3* is not expressed in this particular tissue. While these data provide a rationale for E- to P-cadherin switch in gastric cancer, they do not inform whether the eQTL promotes *CDH3* expression. To address this question, we deleted the eQTL in a tumour cell line expressing wild-type E- and P-cadherins, preventing it from interacting with the *CDH1* promoter. This resulted in downregulation of both P- and E-cadherin expression and demonstrated that the *CDH3*–eQTL controlled P-cadherin expression. Taken together, these data provide evidence that loss of *CDH1*/E-cadherin expression alters the chromatin conformation at the *CDH3* lo-cus, allowing interaction of the *CDH1* promoter with the *CDH3*–eQTL, which seems to potentiate *CDH3*/P-cadherin expression. Our findings also suggest that the *CDH3*–eQTL acts as an enhancer regulating both P- and E-cadherin protein expression, thus regulating P-cadherin expression in the switch mechanism. Similar findings have been observed for the ZRS enhancer located within the *Lmbr1* gene, also controlling Shh gene expression [40].

Single-cell sorting of CRISPR-Cas9 clones herein described only yielded one homozygous clone for each set of sgRNAs harbouring the intended deletions, which we acknowl- edge as a study limitation.

Conclusions

Overall, this study provides insights into a noncoding regulatory element that controls E- and P-cadherin expression and triggers E- to P-cadherin switch in the presence of a *CDH1* depletion, through a re-shuffle of the *CDH1*/*CDH3 loci*.

Supplementary Materials: The following supporting information can be downloaded at https: //www.mdpi.com/article/10.3390/biology12060803/s1, Supplementary Table S1: sgRNAs and primers used for CRISPR-Cas9 and genotyping; Supplementary Table S2: Primers used for 4C-seq; Supplementary Table S3: Differentially expressed proteins in *CDH1* exon 2 del vs. parental cells; Supplementary Table S4: Enriched GO terms of differentially expressed proteins in *CDH1* exon 2 del vs. parental cells.

Acknowledgments

This work was supported by: (1) Doctoral Programme in Biomedicine—Faculty of Medicine, University of Porto, to C.S.J. and C.P.; (2) Doctoral Programme in Computational Sciences, Faculty of Sciences, University of Porto, to M.F. We would like to thank Rita Barbosa-Matos for performing E-cadherin immunohistochemistry.

Funding

This work received funding from (1) FEDER (Fundo Europeu de Desenvolvimento Re- gional) funds through the COMPETE 2020—Operacional Programme for Competitiveness and Internationalisation (POCI), Portugal 2020 and by Fundação para a Ciência e a Tecnologia (FCT): “Solving the 3D Chromatin Structure of *CDH1* Locus to Identify Disease-Associated Mechanisms (3DChrome)”, PTDC/BTM-TEC/30164/2017; POCI-01-0145-FEDER-030164 and “Life-time Risk Estimations and Genetic Modifiers Of Hereditary Diffuse Gastric Cancer (LEGOH)”, PTDC/BTM- TEC/6706/2020; (2) European Union’s Horizon 2020 research and innovation programme under grant agreement No 779257: “Solve-RD”; (3) COMPETE 2020—POCI, Lisboa Portugal Regional Operational Programme (Lisboa2020), Algarve Portugal Regional Operational Programme (CRESC Algarve2020), under the PORTUGAL 2020 Partnership Agreement, through the European Regional Development Fund (ERDF), and by FCT: “GenomePT”, POCI-01-0145-FEDER-022184; (4) NORTE2020 and FEDER funds through the COMPETE 2020—POCI and Portugal 2020: “Diabetes & obesity at the crossroads between Oncological and Cardiovascular diseases—a system analysis NETtwork towards precision medicine (DOCnet)”, NORTE-01-0145-FEDER-000003; (5) Portuguese Mass Spectrometry Network, integrated in the National Roadmap of Research Infrastructures of Strategic Relevance (ROTEIRO/0028/2013; LISBOA-01-0145-FEDER-022125); (6) FCT with salary support for PhD stu- dentships to C.S.J. (Ref. SFRH/BD/140796/2018), C.P. (Ref. SFRH/BD/113031/2015), and M.F. (Ref. 2020.05763.BD).

Institutional Review Board Statement: The study was conducted in accordance with the Declaration of Helsinki and approved by the Ethics Committee of Centro Hospitalar Universitário de São João (CES072017 on 16 March 2017 and CE 305-19 on 8 May 2020).

Informed Consent Statement: Informed consent was obtained from all subjects involved in the study. Data Availability Statement: Raw data supporting the findings of this study are available upon request from the corresponding author. Processed data are available in Supplementary Materials.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Ivanov, D.B.; Philippova, M.P.; Tkachuk, V.A. Structure and Functions of Classical Cadherins. Biochemistry 2001, 66, 1174–1186. [CrossRef] [PubMed]
2. Bhatt, T.; Rizvi, A.; Batta, S.P.R.; Kataria, S.; Jamora, C. Signaling and Mechanical Roles of E-cadherin. Cell Commun. Adhes. 2013, 20, 189–199. [CrossRef]
3. Nollet, F.; Koolsa, P.; Van Roy, F. Phylogenetic analysis of the cadherin superfamily allows identification of six major subfamilies besides several solitary members. J. Mol. Biol. 2000, 299, 551–572. [CrossRef] [PubMed]
4. Cao, J.; O'day, D.R.; Pliner, H.A.; Kingsley, P.D.; Deng, M.; Daza, R.M.; Zager, M.A.; Aldinger, K.A.; Blecher-Gonen, R.; Zhang, F.; et al. A human cell atlas of fetal gene expression. Science 2020, 370, eaba7721. [CrossRef]
5. Daniel, C.W.; Strickland, P.; Friedmann, Y. Expression and Functional Role of E- and P-Cadherins in Mouse Mammary Ductal Morphogenesis and Growth. Dev. Biol. 1995, 169, 511–519. [CrossRef] [PubMed]
6. Park, J.; Park, E.; Han, S.-W.; Im, S.-A.; Kim, T.-Y.; Kim, W.-H.; Oh, D.-Y.; Bang, Y.-J. Down-regulation of P-cadherin with PF-03732010 inhibits cell migration and tumor growth in gastric cancer. Investig. New Drugs 2012, 30, 1404–1412. [CrossRef]
7. Kim, M.A.; Jung, E.J.; Lee, H.S.; Lee, H.E.; Yang, H.-K.; Oh, D.-Y.; Bang, Y.-J.; Kim, W.H. P-cadherin expression in gastric carcinoma: Its regulation mechanism and prognostic significance. Hum. Pathol. 2010, 41, 877–885. [CrossRef]
8. Eom, B.W.; Ryu, K.W.; Yoon, H.M.; Kook, M.-C. Predictive value of E-cadherin and EpCAM for detection of metastatic lymph node in early gastric cancer. Chin. J. Cancer Res. 2020, 32, 614–620. [CrossRef]
9. Liu, N.; Yu, Q.; Liu, T.J.; Gebreamlak, E.P.; Wang, S.L.; Zhang, R.J.; Zhang, J.; Niu, Y. P-cadherin expression and basal-like subtype in breast cancers. Med. Oncol. 2012, 29, 2606–2612. [CrossRef]
10. Margan, M.M.; Cimpean, A.M.; Ceausu, A.R.; Raica, M. Differential Expression of E-Cadherin and P-Cadherin in Breast Cancer Molecular Subtypes. Anticancer Res. 2020, 40, 5557–5566. [CrossRef]
11. Turashvili, G.; E McKinney, S.; Goktepe, O.; Leung, S.C.; Huntsman, D.G.; A Gelmon, K.; Los, G.; A Rejto, P.; Aparicio, S.A.J.R. P-cadherin expression as a prognostic biomarker in a 3992 case tissue microarray series of breast cancer. Mod. Pathol. 2011, 24, 64–81. [CrossRef]
12. Faria, G.; Cardoso, M.J.; Martins, D.; Bettencourt, H.; Amendoeira, I.; Schimitt, F. P-cadherin as prognostic factor for loco-regional relapse in breast cancer. Acta Med. Port. 2012, 25, 97–105.
13. Sridhar, S.; Rajesh, C.; Jishnu, P.V.; Jayaram, P.; Kabekkodu, S.P. Increased expression of P-cadherin is an indicator of poor prognosis in breast cancer: A systematic review and meta-analysis. Breast Cancer Res. Treat. 2020, 179, 301–313. [CrossRef] [PubMed]
14. Gravdal, K.; Halvorsen, O.J.; Haukaas, S.A.; Akslen, L.A. A Switch from E-Cadherin to N-Cadherin Expression Indicates Epithelial to Mesenchymal Transition and Is of Strong and Independent Importance for the Progress of Prostate Cancer. Clin. Cancer Res. 2007, 13, 7003–7011. [CrossRef]
15. Appolloni, I.; Barilari, M.; Caviglia, S.; Gambini, E.; Reisoli, E.; Malatesta, P. A cadherin switch underlies malignancy in high-grade gliomas. Oncogene 2015, 34, 1991–2002. [CrossRef]
16. Araki, K.; Shimura, T.; Suzuki, H.; Tsutsumi, S.; Wada, W.; Yajima, T.; Kobayahi, T.; Kubo, N.; Kuwano, H. E/N-cadherin switch mediates cancer progression via TGF-beta-induced epithelial-to-mesenchymal transition in extrahepatic cholangiocarcinoma. Br. J. Cancer 2011, 105, 1885–1893. [CrossRef]
17. Barber, M.; Murrell, A.; Ito, Y.; Maia, A.-T.; Hyland, S.; Oliveira, C.; Save, V.; Carneiro, F.; Paterson, A.; Grehan, N.; et al. Mechanisms and sequelae of E-cadherin silencing in hereditary diffuse gastric cancer. J. Pathol. 2008, 216, 295–306. [CrossRef]
18. Li, B.; Guo, X.; Li, N.; Chen, Q.; Shen, J.; Huang, X.; Huang, G.; Wang, F. WNT1, a target of miR-34a, promotes cervical squamous cell carcinoma proliferation and invasion by induction of an E-P cadherin switch via the WNT/β-catenin pathway. Cell. Oncol. 2020, 43, 489–503. [CrossRef] [PubMed]
19. D'Antonio, M.; Guerra, R.F.; Cereda, M.; Marchesi, S.; Montani, F.; Nicassio, F.; Di Fiore, P.P.; Ciccarelli, F.D. Recessive Cancer Genes Engage in Negative Genetic Interactions with Their Functional Paralogs. Cell Rep. 2013, 5, 1519–1526. [CrossRef] [PubMed]
20. Andrey, G.; Mundlos, S. The three-dimensional genome: Regulating gene expression during pluripotency and development. Development 2017, 144, 3646–3658. [CrossRef]
21. Will, A.J.; Cova, G.; Osterwalder, M.; Chan, W.L.; Wittler, L.; Brieske, N.; Heinrich, V.; de Villartay, J.P.; Vingron, M.; Klopocki, E.; et al. Composition and dosage of a multipartite enhancer cluster control developmental expression of Ihh (Indian hedgehog). Nat Genet. 2017, 49, 1539–1545. [CrossRef] [PubMed]
22. Lupiáñez, D.G.; Kraft, K.; Heinrich, V.; Krawitz, P.; Brancati, F.; Klopocki, E.; Horn, D.; Kayserili, H.; Opitz, J.M.; Laxova, R.; et al. Disruptions of Topological Chromatin Domains Cause Pathogenic Rewiring of Gene-Enhancer Interactions. Cell 2015, 161, 1012–1025. [CrossRef] [PubMed]
23. Spielmann, M.; Lupiáñez, D.G.; Mundlos, S. Structural variation in the 3D genome. Nat. Rev. Genet. 2018, 19, 453–467. [CrossRef] [PubMed]
24. Tennekes, M. tmap: Thematic Maps in R. J. Stat. Softw. 2018, 84, 1–39. [CrossRef]
25. Feingold, E.A.; Good, P.J.; Guyer, M.S.; Kamholz, S.; Liefer, L.; Wetterstrand, K.; Collins, F.S.; Gingeras, T.R.; Kampa, D.; Sekinger, E.A.; et al. The ENCODE (ENCyclopedia Of DNA Elements) Project. Science 2004, 306, 636–640. [CrossRef]
26. Zhang, X.O.; Gingeras, T.R.; Weng, Z. Genome-wide analysis of polymerase III–transcribed Alu elements suggests cell-type– specific enhancer function. Genome Res. 2019, 29, 1402–1414. [CrossRef]
27. Lee, D.; Zhang, J.; Liu, J.; Gerstein, M. Epigenome-based splicing prediction using a recurrent neural network. PLoS Comput. Biol. 2020, 16, e1008006. [CrossRef]
28. Ran, F.A.; Hsu, P.D.; Wright, J.; Agarwala, V.; Scott, D.A.; Zhang, F. Genome engineering using the CRISPR-Cas9 system. Nat. Protoc. 2013, 8, 2281–2308. [CrossRef]
29. Livak, K.J.; Schmittgen, T.D. Analysis of relative gene expression data using real-time quantitative PCR and the $2^{-\Delta\Delta CT}$ Method. Methods 2001, 25, 402–408. [CrossRef]
30. Hughes, C.S.; Moggridge, S.; Müller, T.; Sorensen, P.H.; Morin, G.B.; Krijgsveld, J. Single-pot, solid-phase-enhanced sample preparation for proteomics experiments. Nat. Protoc. 2019, 14, 68–85. [CrossRef]
31. Moore, J.E.; Purcaro, M.J.; Pratt, H.E.; Epstein, C.B.; Shoresh, N.; Adrian, J.; Kawli, T.; Davis, C.A.; Dobin, A. Expanded encyclopaedias of DNA elements in the human and mouse genomes. Nature 2020, 583, 699–710. [CrossRef]
32. Buenrostro, J.D.; Wu, B.; Chang, H.Y.; Greenleaf, W.J. ATAC-seq: A Method for Assaying Chromatin Accessibility Genome-Wide. Curr. Protoc. Mol. Biol. 2015, 109, 21.29.1–21.29.9. [CrossRef] [PubMed]
33. Van de Werken, H.J.; de Vree, P.J.; Splinter, E.; Holwerda, S.J.; Klous, P.; de Wit, E.; de Laat, W. 4C technology: Protocols and data analysis. Methods Enzym. 2012, 513, 89–112. [CrossRef]
34. Krijger, P.H.; Geeven, G.; Bianchi, V.; Hilvering, C.R.; de Laat, W. 4C-seq from beginning to end: A detailed protocol for sample preparation and data analysis. Methods 2020, 170, 17–32. [CrossRef]
35. Geeven, G.; Teunissen, H.; De Laat, W.; De Wit, E. peakC: A flexible, non-parametric peak calling package for 4C and Capture-C data. Nucleic Acids Res. 2018, 46, e91. [CrossRef] [PubMed]
36. Martins, E.P.; Gonçalves, C.S.; Pojo, M.; Carvalho, R.; Ribeiro, A.S.; Miranda-Gonçalves, V.; Taipa, R.; Pardal, F.; Pinto, A.A.; Custódia, C.; et al. Cadherin-3 is a novel oncogenic biomarker with prognostic value in glioblastoma. Mol. Oncol. 2022, 16, 2611–2631. [CrossRef] [PubMed]
37. Van Marck, V.; Stove, C.; Jacobs, K.; Eynden, G.V.D.; Bracke, M. P-cadherin in adhesion and invasion: Opposite roles in colon and bladder carcinoma. Int. J. Cancer 2011, 128, 1031–1044. [CrossRef] [PubMed]
38. Ribeiro, A.S.; Sousa, B.; Carreto, L.; Mendes, N.; Nobre, A.R.; Ricardo, S.; Albergaria, A.; Cameselle-Teijeiro, J.F.; Gerhard, R.; Söderberg, O.; et al. P-cadherin functional role is dependent on E-cadherin cellular context: A proof of concept using the breast cancer model. J. Pathol. 2013, 229, 705–718. [CrossRef] [PubMed]
39. Christgen, M.; Bartels, S.; van Luttikhuizen, J.L.; Bublitz, J.; Rieger, L.U.; Christgen, H.; Stark, H.; Sander, B.; Lehmann, U.; Steinemann, D.; et al. E-cadherin to P-cadherin switching in

lobular breast cancer with tubular elements. Mod. Pathol. 2020, 33, 2483–2498. [CrossRef]
40. Paliou, C.; Guckelberger, P.; Schöpflin, R.; Heinrich, V.; Esposito, A.; Chiariello, A.M.; Bianco, S.; Annunziatella, C.; Helmuth, J.; Haas, S.; et al. Preformed chromatin topology assists transcriptional robustness of Shh during limb development. Proc. Natl. Acad. Sci. USA 2019, 116, 12390–12399. [CrossRef]

Chapter 5

General discussion



Regulatory elements play a critical role in proper gene expression in time and space (Spitz and Furlong 2012b). Despite distant in the DNA sequence, enhancer-promoter communications occur within the 3D chromatin folding. When disrupted, regulatory landscapes may result in gene misexpression and, consequently, disease (Cova et al. 2023), for which hereditary diffuse gastric cancer (HDGC) is no exception.

Single nucleotide variants (SNV) in *CDH1* and *CTNNA1* genes and copy number variants (CNV) in *CDH1* are the main HDGC genetic triggers (Guilford et al. 1998c; Majewski et al. 2013b), which predisposes to diffuse gastric cancer (DGC) and lobular breast cancer (LBC) development in early adulthood (Blair et al. 2020a). Yet, most HDGC families remain genetically unsolved, despite carrying germline *CDH1* mRNA monoallelic downregulation, a common feature observed in families that bear causing SNVs and CNVs (Pinheiro et al. 2010). Altogether, these results hint *CDH1* as the main HDGC genetic trigger. Working under the hypothesis that unsolved HDGC cases are caused by defects in regulatory elements controlling *CDH1* expression, we explored the *CDH1* regulatory network and its impact in *CDH1*/E-cadherin expression, using HDGC and sporadic gastric cancer as disease models.

Besides *CDH1*, other gastro-intestinal cancer syndromes-associated genes have been proposed as genetic triggers of gastric cancer-associated syndromes. In **review paper 1** we mined the literature for individuals with personal or family history compatible with HDGC or familial intestinal gastric cancer (FIGC) with pathogenic or likely pathogenic variants in genes associated with gastric cancer predisposition. DNA repair-associated pathways were frequently mutated in both gastric cancer-associated syndromes. Indeed, *PALB2* and *BRCA2* were the most promising candidate genes in HDGC and FIGC, respectively. Consolidation of these genes holds promise and worth further validation, including segregation analysis, second somatic hit analysis, penetrance, and life-time risk calculations. A small family branch may meet HDGC/FIGC clinical criteria but represent branches of larger families with different syndromes. **Review paper 1** also highlights the importance of collecting extended and updated family histories, to decide the most adequate genetic testing and post-testing management.

We applied a similar approach to solve the missing heritability in HDGC families and sequenced the gDNA of these families for **original paper 1**. Given the lack of SNVs and CNVs in HDGC-predisposing genes (*CDH1* and *CTNNA1*), we searched for variants in genes associated with gastric cancer syndromes. A *MLH1* CNV was found in a family with DGC biased development. Deletion of putative stomach-specific regulatory elements, absent in normal colonic mucosa, may enlighten the DGC biased development. This family could rather represent a branch of a larger family with lynch syndrome, as seen in **review paper 1**. Then, we inquired the *CDH1* regulatory network within its topological-associating domain for CNVs. We found a large deletion in the upstream gene *CDH3*, encompassing a *CDH1*-interactor with accessible chromatin and a *CDH1*-eQTL.

We mimicked the CNV and the putative regulatory elements in a gastric cancer cell line and testes their enhancer activity in mouse embryos. The *CDH3*-CNV completely knocks down *CDH1*, while edition of either regulatory element impacts *CDH1* in approximately 50%. These display enhancer activity in mouse embryos (forebrain, palate and stomach) and may compose co-operative and hypomorphic enhancers, as observed in **original paper 4** and for the *ihh locus* (Will et al. 2017). Their synergistic mode of action knocks down *CDH1* expression, while separately only reduce its expression in half. Additionally, we found a *CDH1-TANGO6* downstream intergenic region, which deleted downregulates *CDH1*/E-cadherin expression in half and displays enhancer activity. **Original paper 2** described a family with extremely early onset DGC and carrying a *CDH1-TANGO6* downstream intergenic deletion, besides *CDH1*. The germline landscape of HDGC families revealed impacted immune-associated pathways, mainly in CNVs overlapping gastric accessible chromatin of young-affected individuals. Thus, highlighting the role of stomach regulatory elements in creating the appropriate genetic background for DGC development in the presence of an unknown causing defect.

In an attempt to solve the missing heritability of a HDGC family lost for genetic follow-up, we re-analysed whole exome sequencing data of the proband and several family members in **original paper 2**. We described a family with extremely early onset DGC and high disease penetrance predisposed by a *CDH1-TANGO6* large deletion. Combined loss of *CDH1* and its downstream regulatory elements may control the high disease aggressiveness observed in this family. CRISPR-cas9 clones mimicking the *CDH1* fraction and the family CNV disclosed increased *CDH1* mRNA expression downregulation and adhesion-associated pathways in the *CDH1-TANGO6* deletion. Combined deletion of *CDH1* and its downstream regulatory elements may predispose to extremely early onset DGC in this family (**original paper 2**) and for later onset DGC in a family with a downstream regulatory elements' deletion alone (**original paper 1**). Additionally, immune-related pathways were specifically impaired in the family CNV, feature also observed in the germline landscape of HDGC patients (**original paper 1**). Despite most *CDH1* mutation carriers present early-stage intramucosal SRCCs, penetrance of advanced DGC is still incomplete (Rocha et al. 2018). SRCC progression to aggressive DGC may be related with environmental factors and modifiers, enclosing immune surveillance genes, such as variants in HLA-associated genes (Decourtye-Espiard and Guilford 2023). Herein, we proved that this progression may be germline predisposed, rather that acquired tumour variants. These impaired pathways may also be linked with increased proliferation and invasion, and decreased DNA damage response, characteristic of more aggressive tumours, than the commonly described for HDGC patients bearing *CDH1* coding variants, who generally die by dissemination to the peritoneum. **Original paper 2** also highlight the importance of retesting families with suspected tumour-risk syndromes with updated technical approaches and knowledge to confer proper management and care.

Then, we focused on sporadic gastric cancer as a disease model.

In **original paper 3** we explored the role of *CDH1* intronic regulatory elements in gastric cancer cell lines and patients. The 7 selected regulatory elements display differential methylation patterns between cells and tumours expressing and non-expressing *CDH1*. The role of two putative regulatory elements in *CDH1* expression and metastasis development have been further dissected in **original paper 4**.

Given the abovementioned roles of *CDH1* regulatory elements, we next explored E-cadherin intronic regulatory network in **original paper 4**. Working under the hypothesis that E-cadherin main functions are determined by regulatory elements within its introns, rather than its coding regions, we selected two putative regulatory elements, also described in **original paper 3**, and a coding sequence. The intronic regulatory elements display hypomorphic and co-operative mode of action, completely downregulating *CDH1* when edited simultaneously, producing a similar outcome as an exonic deletion, that separately contribute with 50% expression loss. Evidence for their co-operativity also lays with bilateral interactions between the two enhancers, requiring co-inactivation of additional regulatory elements to produce the same effect as a coding variant in *CDH1* expression. This enhancer mode of action has also been observed for *CDH1* enhancers in **original paper 1**. Given that E-cadherin is known marker of invasion (Elisha et al. 2018), we tested whether edition of one/both CREs could control migration and metastasis development, beyond their impact in E-cadherin expression. Despite a similar impact on *CDH1*/E-cadherin expression, edition of both CREs were unable to produce lung metastasis, as opposed to a coding deletion that produces a similar amount of lung metastasis as a metastatic gastric cancer cell line. These results suggest a regulatory role beyond the *CDH1 locus* itself. Again, despite a similar *CDH1* knock-down, edition of both CREs displays a much stronger impact on the transcriptome, carrying almost 17 times more misregulated genes, than a *CDH1* coding deletion. Transcriptomic analysis revealed that both CREs edition downregulates adhesion-associated pathways, while the coding deletion upregulates immune-associated pathways. Impairment of immune-associated pathways has been observed in a family with extremely early onset DGC (**original paper 2**) and at the germline landscape overlapping stomach regulatory sequenced of unsolved young-affected individuals (**original paper 1**). Despite showing increased proliferation, migration, and invasion, features commonly observed in more aggressive tumours, edition of both CREs lacks metastatic capacity in mice. Downregulation of ‘cellular extravasation’ may clarify the biological mechanism for reduced metastasis when editing both CREs, even with E-cadherin loss of function. Accordingly, CREs homozygous deletion improves breast cancer patients’ survival, while their amplification reduces. Altogether, these results suggest CREs as major regulators of metastasis, rather than the *CDH1* gene itself. Additionally, our study shows that putative CRE lacking biochemical annotations may in fact act as gene-specific enhancer, but

also control genome-wide chromatin accessibility, that consequently control genome-wide gene expression.

Original paper 5 discloses a mechanism for E-cadherin to P-cadherin switch in gastric cancers. For which we dissected the switch frequency in a gastric cancer cohort and mimicked a *CDH1* knockout to identify regulatory elements controlling *CDH1*/*3* expression. Most gastric tumours express abnormal *CDH3*, from which 43% display an E- to P-cadherin switch. *CDH1* knockdown triggered E-cadherin loss and P-cadherin expression increase, likely recuing adherens junctions, as well as increased cell migration and proliferation. E- to P-cadherin switch triggered *CDH1*-promoter interactions with a putative regulatory element, *CDH3*-eQTL, absent in normal stomach and parental cells. Working under the hypothesis that these chromatin architectural changes control *CDH3*/P-cadherin expression in *CDH1*-depleted cells, we deleted the *CDH3*-eQTL and obtained reduced expression of both *CDH1* and *CDH3*. Indeed, this *CDH3*-eQTL has been proved to harbour enhancer activity in forebrain and stomach of mouse embryos (**original paper 1**), suggesting this region as a *CDH1* and *CDH3* tissue-specific enhancer. These data highlight a novel mechanism triggering E- to P-cadherin switch in gastric cancer, providing evidence that loss of *CDH1*/E-cadherin expression alters the chromatin conformation at the *CDH3* locus allowing interaction of the *CDH1* promoter with the *CDH3*-eQTL, which seems to potentiate *CDH3*/P-cadherin expression.

From the clinical standpoint, we pinpointed novel HDGC genetic defects, solving the missing heritability in several HDGC families. These findings guide genetic testing/counselling in family members and prophylactic gastrectomy and mastectomy of at-risk individuals to prevent disease onset. From the research standpoint, we unveil novel *CDH1*-associated regulatory mechanisms controlling *CDH1* expression and driving tumorigenesis/metastasis in HDGC and its sporadic counterpart.

References

Blair VR, McLeod M, Carneiro F, et al (2020) Hereditary diffuse gastric cancer: updated clinical practice guidelines. Lancet Oncol 21:e386–e397. https://doi.org/10.1016/S1470-2045(20)30219-9

Cova G, Glaser J, Schöpflin R, et al (2023) Combinatorial effects on gene expression at the Lbx1/Fgf8 locus resolve split-hand/foot malformation type 3. Nat Commun 14:. https://doi.org/10.1038/S41467-023-37057-Z

Decourtye-Espiard L, Guilford P (2023) Hereditary Diffuse Gastric Cancer. Gastroenterology 164:719–735. https://doi.org/10.1053/J.GASTRO.2023.01.038

Elisha Y, Kalchenko V, Kuznetsov Y, Geiger B (2018) Dual role of E-cadherin in the regulation of invasive collective migration of mammary carcinoma cells. Sci Rep 8:. https://doi.org/10.1038/S41598-018-22940-3

Guilford P, Hopkins J, Harraway J, et al (1998) E-cadherin germline mutations in familial gastric cancer. Nature 392:402–405. https://doi.org/10.1038/32918

Majewski IJ, Kluijt I, Cats A, et al (2013) An alpha-E-catenin (*CTNNA1*) mutation in hereditary diffuse gastric cancer. J Pathol 229:621–629. https://doi.org/10.1002/path.4152

Pinheiro H, Bordeira-Carrico R, Seixas S, et al (2010) Allele-specific *CDH1* downregulation and hereditary diffuse gastric cancer. Hum Mol Genet 19:943–952. https://doi.org/10.1093/hmg/ddp537

Rocha JP, Gullo I, Wen X, et al (2018) Pathological features of total gastrectomy specimens from asymptomatic hereditary diffuse gastric cancer patients and implications for clinical management. Histopathology 73:878–886. https://doi.org/10.1111/HIS.13715

Spitz F, Furlong EEM (2012) Transcription factors: from enhancer binding to developmental control. Nat Rev Genet 13:613–626. https://doi.org/10.1038/NRG3207

Will AJ, Cova G, Osterwalder M, et al (2017) Composition and dosage of a multipartite enhancer cluster control developmental expression of Ihh (Indian hedgehog). Nat Genet 49:1539–1545. https://doi.org/10.1038/NG.3939

Appendix A

Gastric cancer genetic predisposition and clinical presentations: Established heritable causes and potential candidate genes

José Garcia-Pelaez*, Rita Barbosa-Matos*, Celina São José*, Sónia Sousa Irene Gulloa, Nicoline Hoogerbrugge Fátima Carneiro and Carla Oliveira

European Journal of Medical Genetics

2021

DOI: 10.1016/j.ejmg.2021.104401

European Journal of Medical Genetics 65 (2022) 104401

Contents lists available at ScienceDirect

European Journal of Medical Genetics

journal homepage: www.elsevier.com/locate/ejmg

Gastric cancer genetic predisposition and clinical presentations: Established heritable causes and potential candidate genes

José Garcia-Pelaez [a,b,c,1,2], Rita Barbosa-Matos [a,b,d,1,2], Celina São José [a,b,c,1,2], Sónia Sousa [a,b,1], Irene Gullo [a,b,e,f,1], Nicoline Hoogerbrugge [g,1], Fátima Carneiro [a,b,e,f,1], Carla Oliveira [a,b,e,1,*]

[a] i3S – Instituto de Investigação e Inovação em Saúde, Porto, Portugal
[b] IPATIMUP – Instituto de Patologia e Imunologia Molecular da Universidade do Porto, Porto, Portugal
[c] Doctoral Programme in Biomedicine, Faculty of Medicine, University of Porto, Porto, Portugal
[d] International Doctoral Programme in Molecular and Cellular Biotechnology Applied to Health Sciences from Institute of Biomedical Sciences Abel Salazar (ICBAS), University of Porto, Porto, Portugal
[e] FMUP – Faculty of Medicine of the University of Porto, Porto, Portugal
[f] Centro Hospitalar e Universitário S. João, Porto, Portugal
[g] Department of Human Genetics, Radboud University Medical Center, Radboud Institute for Molecular Life Sciences, Nijmegen, the Netherlands

ARTICLE INFO

Keywords:
Hereditary diffuse gastric cancer (HDGC)
CDH1
Gastric adenocarcinoma and proximal polyposis of the stomach (GAPPS)
Familial intestinal gastric cancer (FIGC)
Candidate genes

ABSTRACT

Tumour risk syndromes (TRS) are characterized by an increased risk of early-onset cancers in a familial context. High cancer risk is mostly driven by loss-of-function variants in a single cancer-associated gene. Presently, predisposition to diffuse gastric cancer (DGC) is explained by *CDH1* and *CTNNA1* pathogenic and likely pathogenic variants (P/LP), causing Hereditary Diffuse Gastric Cancer (HDGC); while *APC* promoter 1B single nucleotide variants predispose to Gastric Adenocarcinoma and Proximal Polyposis of the Stomach (GAPPS). Familial Intestinal Gastric Cancer (FIGC), recognized as a GC-predisposing disease, remains understudied and genetically unsolved. GC can also occur in the spectrum of other TRS. Identification of heritable causes allows defining diagnostic testing criteria, helps to clinically classify GC families into the appropriate TRS, and allows performing pre-symptomatic testing identifying at-risk individuals for downstream surveillance, risk reduction and/or treatment. However, most of HDGC, some GAPPS, and most FIGC patients/families remain unsolved, expecting a heritable factor to be discovered. The missing heritability in GC-associated tumour risk syndromes (GC-TRS) is likely explained not by a single major gene, but by a diversity of genes, some, predisposing to other TRS. This would gain support if GC-enriched small families or apparently isolated early-onset GC cases were hiding a family history compatible with another TRS. Herein, we revisited current knowledge on GC-TRS, and searched in the literature for individuals/families bearing P/LP variants predisposing for other TRS, but whose probands display a clinical presentation and/or family history also fitting GC-TRS criteria. We found 27 families with family history compatible with HDGC or FIGC, harbouring 28 P/LP variants in 16 TRS-associated genes, mainly associated with DNA repair. *PALB2* or *BRCA2* were the most frequently mutated candidate genes in individuals with family history compatible with HDGC and FIGC, respectively. Consolidation of *PALB2* and *BRCA2* as HDGC- or FIGC-associated genes, respectively, holds promise and worth additional research. This analysis further highlighted the influence, that proband's choice and small or unreported family history have, for a correct TRS diagnosis, genetic screening, and disease management. In this review, we provide a rational for identification of particularly relevant candidate genes in GC-TRS.

* Corresponding author. i3S – Instituto de Investigação e Inovação em Saúde, Porto, Portugal, Rua Alfredo Allen, 208, 4200-135, Porto, Portugal.
E-mail address: carlaol@ipatimup.pt (C. Oliveira).
[1] Full Member of the European Reference Network on Genetic Tumor Risk Syndromes (ERN GENTURIS) – Project ID No 739547.
[2] Equivalent contribution.

https://doi.org/10.1016/j.ejmg.2021.104401
Received 1 May 2021; Received in revised form 10 November 2021; Accepted 28 November 2021
Available online 3 December 2021
1769-7212/© 2021 Elsevier Masson SAS. All rights reserved.

1. Introduction

Tumour risk syndromes (TRS) are a group of diseases characterized by increased risk to develop cancer, generally at an earlier age than expected for the general population (Rahner and Steinke, 2008). In most known TRS, high cancer risk exists due to the presence of a heterozygous gene variant disrupting the function of a single important cancer-associated gene, that is transmitted to the offspring with an autosomal dominant pattern (monogenic autosomal dominant – AD) (Leoz et al., 2016). Although less frequent, clear recessive patterns of disease inheritance are also emerging in TRS (monogenic autosomal recessive – AR) (Vogt et al., 2009; Win et al., 2016). In most TRS, evidence of heritability are: (a) cancer diagnosis at early age; (b) familial clustering of one cancer type or a group of related cancers, and; (c) a patient presenting metachronous and/or synchronous primary tumours (Rahner and Steinke, 2008; Vos et al., 2019). The identification of heritable genetic causes allows defining diagnostic testing criteria, classifying families into TRS based on clinical data and identifying at-risk individuals for downstream risk-reduction and/or treatment (Vos et al., 2019).

To consolidate heritable defects as disease causative, there is an absolute need for reliable clinical and pathological data, recurrently associated with the potential heritable cause (Oliveira et al., 2015; Rahner and Steinke, 2008). The difficulty in compiling and combining this information constitutes the major drawback for the establishment of novel causes of disease. The complexity of inherited diseases increases further with the contribution of genetic and/or epigenetic modifiers, as well as environmental factors, which, individually or in combination, play an unneglectable role in cancer predisposition, disease development, age of disease onset, and disease spectrum (Cooper et al., 2013; Riordan and Nadeau, 2017). Indeed, carriers of a given causative variant may live a long life free of disease, while others succumb to severe disease at an early age (Oliveira et al., 2015). Also, while in most genetically solved TRS families, clinical presentations are highly correlated with the causal germline defect, in some cases, clinical presentations diverge from the ones expected for a given germline causal defect (Gullo et al., 2018; Rahner and Steinke, 2008). These observations support the role of disease modifiers, and challenge the classical Mendelian patterns of monogenic inheritance, raising the interest both in oligogenic inheritance, polygenic risk scores, and in pleiotropic functions of TRS-causal genes (Kuchenbaecker et al., 2017; Riordan and Nadeau, 2017; Zhu et al., 2015).

These concepts are starting to be explored by genotype-to-phenotype models (e.g., applying machine learning approaches) and by family-based approaches (e.g. using trio-sequencing) and may shed light on the so far molecularly unsolved cases (Drouin et al., 2019; Jezela-Stanek et al., 2020; Vissers and Veltman, 2015).

New challenges involving disease management and clinical criteria for genetic testing are emerging, also due to the fast evolution of next-generation sequencing technologies, multipanel/exome/whole genome testing, as well as 'direct-to-consumer' or 'over-the-counter' genetic susceptibility testing, which are available without the need for a medical referral. These multiple approaches allow the identification of germline variants, that despite sometimes occurring at relevant cancer-associated genes, lack clinical support and prognostic significance, creating a burden for family management (Blair et al., 2020; Lowstuter et al., 2017). All this results in a scenario where geneticists, genetic counsellors, and other clinicians struggle to inform on, and adjust risk perception to decrease adverse psychological effects on "potentially" at-risk individuals, without a specific, clear, or even evident family history of cancer (Cabrera et al., 2010; Gullo et al., 2021; Lowstuter et al., 2017).

In this review, we provide a description of current knowledge on the germline predisposition and clinical features from patients and families affected by gastric cancer (GC) associated TRS. We will further review published data on candidate genes tested for their potential role in heritable GC, and provide a rationale for the identification of particularly relevant candidate genes in this disease.

2. Genetic predisposition to gastric cancer

Evidence for familial predisposition to GC dates back historically to the 18th Century, was observed in the Bonaparte Family and described by Sokoloff in 1938 (Sokoloff, 1938). It took 60 years for a direct association between a germline heritable genetic defect and increased risk to diffuse gastric cancer (DGC) development to be discovered (Guilford et al., 1998). In 1999, the International Gastric Cancer Linkage Consortium (IGCLC) established for the first-time diagnostic criteria and guidelines for the clinical management of GC families (Caldas et al., 1999). In 2012, a second GC and polyposis-associated syndrome was identified (Worthley et al., 2012) with a causal germline defect identified in 2016 (Li et al., 2016). In 2017, European Reference Networks (ERN) emerged in Europe, and the ERN on Genetic Tumour Risk Syndromes, also known as GENTURIS, was created to raise awareness, standardize care offered to patients across European countries, and train European professionals on TRS, including those involving GC (Vos et al., 2019).

It is fairly established that 1–3% of all GCs are hereditary and generically classified into three main TRS: (1) hereditary diffuse gastric cancer (HDGC); (2) gastric adenocarcinoma and proximal polyposis of the stomach (GAPPS), and; (3) familial intestinal gastric cancer (FIGC) (Caldas et al., 1999; Worthley et al., 2012).

Both HDGC and GAPPS have well-recognized heritable mechanisms that confer high risk for developing mainly GC, while FIGC remains widely understudied, presenting a large disease spectrum and generally occurring at an older age (Carvalho et al., 2021). Despite their low incidence, these heritable diseases deserve special attention due to their high morbidity and mortality rates (Vos et al., 2019). In fact, according to the most recent global cancer statistics, GC is the fourth most deadly cancer worldwide (Sung et al., 2021). Therefore, it is of uttermost importance to identify patients with GC predisposition from birth, to improve their genetic testing, surveillance, and/or risk-reduction, ultimately to prevent disease onset. This can be achieved by increasing the knowledge of the inherited factors associated with GC (Pocurull et al., 2021).

It is essential to have access to detailed family history, patient's age of onset, tumours' histological classification, and reports of secondary cancers from all affected individuals, for proper allocation of isolated or familial cases of GC, clustering into a particular GC-TRS (Choi and Kim, 2016). Laurén and the World Health Organization (WHO) classifications (Lauren, 1965; WHO Classification of Tumours Editorial Board, 2019) are considered the foundation for histopathological characterization of gastric tumours according to their subtypes and should be used when collecting data from patients with GC predisposing syndromes.

3. Hereditary diffuse gastric cancer (HDGC)

HDGC is an autosomal dominant TRS, predisposing to DGC and lobular breast cancer (LBC). It was the first recognized heritable GC-TRS and is historically known for its association with germline variants in the *CDH1* gene, but recently *CTNNA1* variants were also found segregating with early-onset DGC (Fig. 1) (Guilford et al., 1998; Majewski et al., 2013). The definition of HDGC evolved over the years, concomitantly with the clinical criteria driving genetic testing. Clinical practice guidelines were updated three times so far, first to better identify at-risk individuals for DGC and later considering the development of LBC in female carriers of *CDH1* germline variants (Blair et al., 2020; Caldas et al., 1999; Fitzgerald et al., 2010; van der Post et al., 2015a). In addition, the discovery of personal or family history of cleft lip and/or cleft-palate in DGC patients or families carrying *CDH1* loss of function (LoF) variants led to the contemplation of these phenotypes, when co-occurring with DGC, as criteria for genetic testing (Blair et al., 2020). A detailed description of the evolution and the current clinical criteria

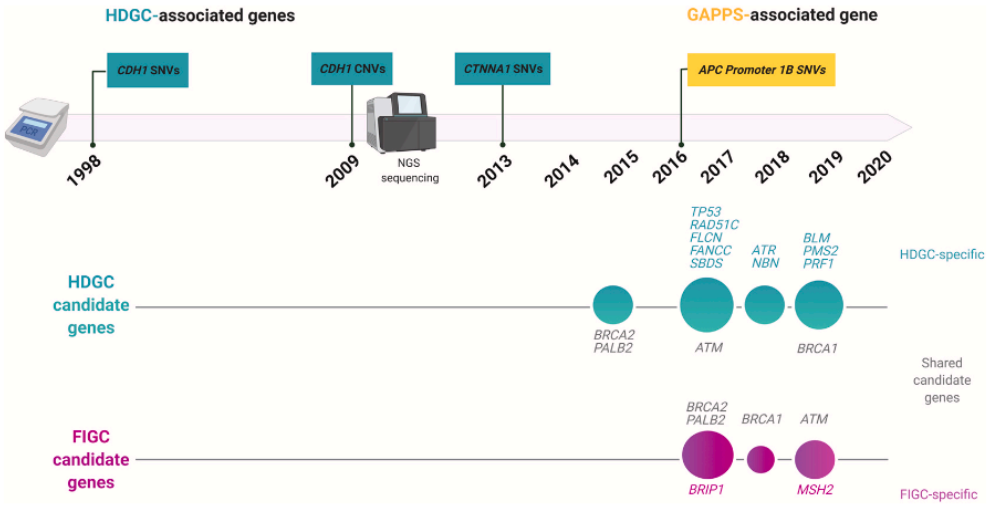


Fig. 1. Genetic timeline of *bona fide* and candidate genes for hereditary diffuse gastric cancer (HDGC), gastric adenocarcinoma and proximal polyposis of the stomach (GAPPS) and familial intestinal gastric cancer (FIGC). SNVs, single-nucleotide variants; CNVs, copy number variants.

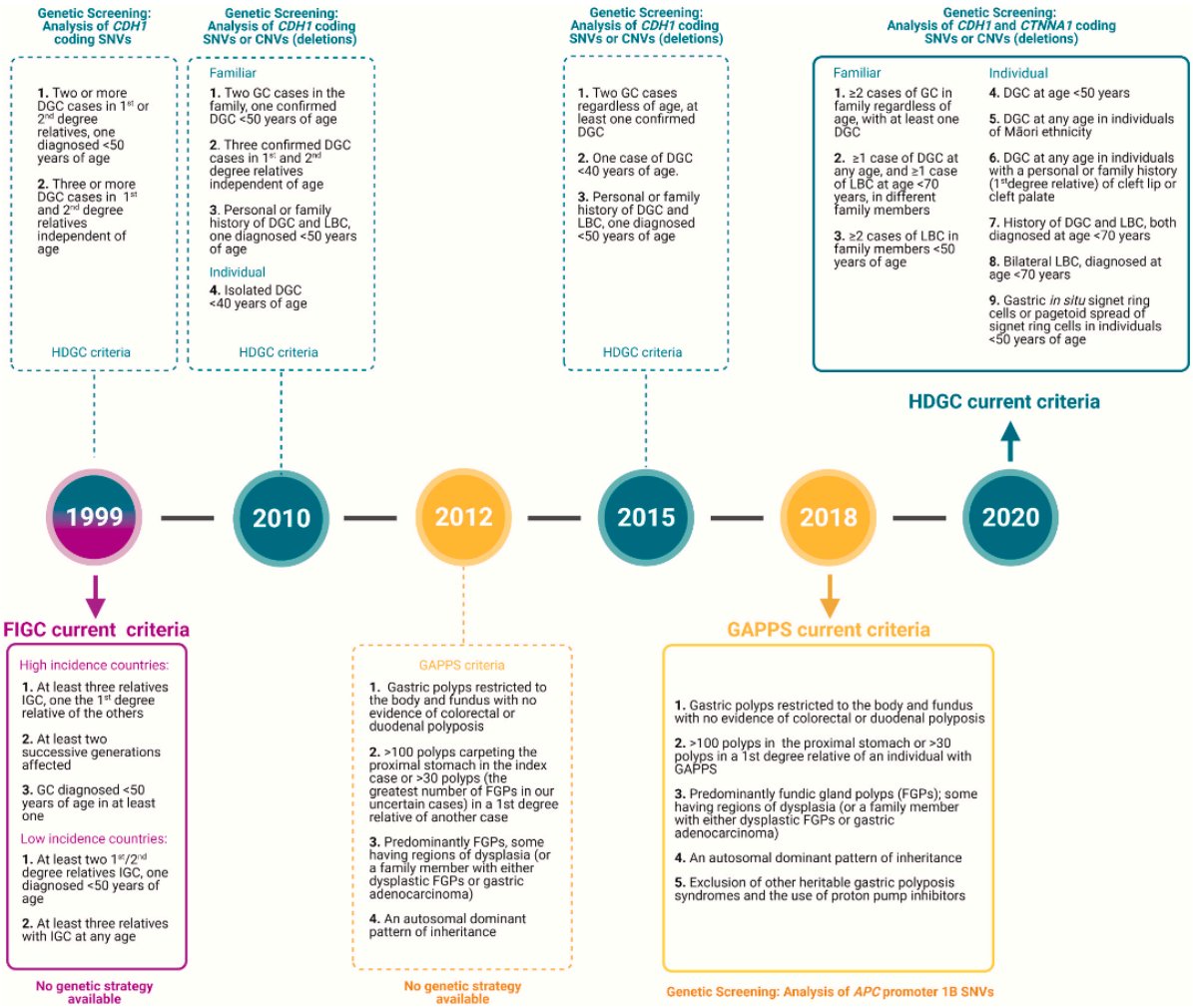


Fig. 2. Evolution of clinical criteria and genetic screening guidelines for the main GC-TRS: HDGC, GAPPS and FIGC. GC, gastric cancer; DGC, diffuse gastric cancer; LBC, lobular breast cancer; FGP(s), Fundic Gland Polyp(s); IGC, intestinal gastric cancer.

according to International Gastric Cancer Linkage Consortium (IGCLC) is depicted in Fig. 2.

To acknowledge the fact that *CDH1* variants may predispose preferentially to LBC in families lacking DGC, the 2020 clinical guidelines created a novel entity – Hereditary Lobular Breast Cancer (HLBC). HLBC families will remain under this designation, while no DGCs are detected in carriers, and their management will be discussed on a case-by-case basis (Fig. 2 (Blair et al., 2020)).

According to the American College of Medical Genetics and Genomics and the Association for Molecular Pathology (ACMG/AMP) criteria, germline *CDH1* pathogenic (P) or likely pathogenic (LP) variants, (Lee et al., 2018), or LoF variants in *CTNNA1* are considered causative of HDGC (Blair et al., 2020). Importantly, families fulfilling clinical criteria, but lacking a causal variant in *CDH1* or *CTNNA1,* are considered by IGCLC as HDGC-like families, understood as genetically unsolved families, and in need of disease surveillance (Blair et al., 2020; Caldas et al., 1999; Fitzgerald et al., 2010; van der Post et al., 2015a).

3.1. Established HDGC genes: CDH1 and CTNNA1

3.1.1. CDH1/E-Cadherin

The *CDH1* gene encodes for E-cadherin, a type I transmembrane glycoprotein, that mediates intercellular adhesion in epithelial mammalian tissues in a Ca^{2+} dependent manner (Harris and Tepass, 2010). This protein is composed of a cytoplasmatic, a transmembrane, and an extracellular domain. The extracellular domain establishes homophilic interactions with E-cadherin proteins in neighbouring cells, while the cytoplasmatic domain interacts with p120 and β-catenin proteins, promoting its stabilization. *CDH1*/E-cadherin LoF in epithelial cells often promotes cell-to-cell adhesion impairment, increased migratory and invasive potential, increased survival, and altered signalling mainly related to the interaction with stroma and the tissue microenvironment (Aberle et al., 1996; Okamoto et al., 2005; Shapiro et al., 1995). Thus, E-cadherin LoF has a close association with the infiltrative and invasive phenotypes observed in HDGC hallmark tumours (DGC and LBC) (Table 1 (Aberle et al., 1996; Frebourg et al., 2006; Humar et al., 2009; Kaurah et al., 2007)). The relationship of germline *CDH1* LoF with DGC and LBC is also reminiscent of the sporadic context, where somatic *CDH1* LoF is a common and determinant event for the initiation and development of DGC and LBC sporadic counterparts (Berx et al., 1995).

Heterozygous germline *CDH1* alterations, either Single Nucleotide Variants (SNVs) or Copy Number Variants (CNVs), occur in 10–40% of HDGC families, depending on the geographical background (Guilford et al., 1998; Oliveira et al., 2009a). Truncating SNVs in the Nonsense-mediated decays (NMD) competent region of the gene (Karam et al., 2008), enclosing frameshift, nonsense, and many splice-site variants, account for the majority of HDGC-associated *CDH1* variants (Table 1 (Oliveira et al., 2015)). Large deletions, encompassing both 5′- and 3′-regulatory or coding regions of the *CDH1* gene were identified in HDGC families in 2009, adding a novel disease mechanism predisposing for HDGC (Table 1, Fig. 1 (Oliveira et al., 2009a)). Truncating SNVs (upstream of the NMD-boundary), start loss variants and large deletions have a clear clinical significance and are thus generally considered actionable in HDGC families. These occur throughout the gene without any hotspot so far identified and follow the classical tumour suppressor gene model (Table 1). Carriers of germline heterozygous *CDH1* causative variants remain healthy until a second event affects the remaining wild-type (WT) allele. Once the WT is targeted, through a second hit mechanism, and *CDH1* somatic expression is completely lost, in either the stomach or breast, a tumour may arise often at an early age (Oliveira et al., 2015). Somatic *CDH1* promoter methylation, large deletions causing loss of heterozygosity (LOH) or a second SNV have been identified as second hit mechanisms in HDGC tumours (Barber et al., 2008; Grady et al., 2000; Oliveira et al., 2009b). The most frequent second inactivation hit in HDGC primary tumours is *CDH1* promoter hypermethylation, while LOH prevails in lymph node metastases (Table 1 (Oliveira et al., 2009b)).

Although academically established as an HDGC cause, *CDH1* variant type diversity remain a controversial territory in clinical practice. Therefore, to standardize *CDH1* variants' classification and to attribute a value to its actionability, Lee K and colleagues developed specific variant curation guidelines for the analysis of germline *CDH1* sequence variants, according to the ACMG/AMP criteria (Table 1) (Lee et al., 2018; Richards et al., 2015). This classification system allows allocating *CDH1* variants into five categories (benign - B, likely benign - LB, variant of unknown significance - VUS, likely pathogenic – LP, and pathogenic – P), and consider P or LP *CDH1* variants as actionable in the HDGC context (Blair et al., 2020; Lee et al., 2018). The definition of actionability of P/LP variants triggers a cascade that involves carrier genetic testing in extended families, surveillance, and/or prophylactic measures (Blair et al., 2020). Nevertheless, predicting VUS actionability, particularly of missense variants, is still very challenging, due to the scarce clinical data available for these type of variants. Given the high frequency of missense variants found in families with HDGC criteria, it is important to understand which of them may induce cryptic splicing with severe consequences such as premature truncation. This knowledge would very likely contribute to a better classification into more benign

Table 1
Molecular and clinical and characteristics of HDGC and GAPPS and their causal genes.

Syndrome	Gene	Spectrum	Main Molecular Variation	Accionable Variants	Hotspot	2nd hit	Penetrance			MANE Transcript (Refseq)	ACMG/AMP Criteria (Yes/No) [PMID Ref]
							GC in males	GC in females	LBC		
HDGC	*CDH1*	DGC, LBC, CLP	Truncating SNVs; large deletions	P/LP	No	Promoter methylation; LOH; 2nd SNV	42%–70% by 80 years of age	33%–55% by 80 years of age	39%–55% by 80 years of age	NM_004360.5	Yes
	CTNNA1	DGC	Truncating SNVs	P/LP	No	NA	NA		NA	NM_001903.5	No
GAPPS	*APC*	GC with Polyposis	Promoter SNVs	P/LP	Promoter 1B	NA	13%		NA	NM_000038.6	No
References (PMID):						19269290 10973239	26182300 31246251 21813476	26182300 31246251	26182300 11729114		30311375

Footnote: GC, Gastric Cancer; DGC, Diffuse Gastric Cancer; LBC, Lobular Breast Cancer; CLP, Clefts of the Lip and/or Palate; FGP(s), Fundic Gland Polyp(s); SNVs, single nucleotide variants; ACMG/AMP, American College of Medical Genetics and Genomics and the Association for Molecular Pathology; P, Pathogenic; LP, Likely Pathogenic; LOH, Loss of heterozygosity; SNV, single nucleotide variant; NA, Not Available.

or pathogenic classes (Barbosa-Matos et al., 2021; Figueiredo et al., 2021).

3.1.2. CDH1-associated disease risk

According to Hansford S. and colleagues (Hansford et al., 2015), *CDH1* variant carriers that met 2010 clinical criteria (Fitzgerald et al., 2010) had a GC cumulative incidence of 70% in males and 56% in females by 80 years of age (Table 1). Recently, Roberts and colleagues estimated the cumulative risk to develop GC by 80 years of age to be 42% in male carriers and 33% in female carriers (Roberts et al., 2019), of which 33% met the 2015 less stringent clinical criteria (Table 1) (van der Post et al., 2015a). Estimations of cumulative risk to develop LBC range between 39% and 55% in females carrying *CDH1* variants (Table 1) (Hansford et al., 2015; Pharoah et al., 2001; Roberts et al., 2019; Xicola et al., 2019). The cumulative risk differs from study to study, and largely depends on whether families included in the risk calculation were ascertained or non-ascertained, if these arose in countries with a high or low incidence of GC and/or breast cancer (BC), were exposed to different genetic and envirnomental modifiers, and importantly on the set of *CDH1* variants under study. This is quite evident from the Roberts et al. study, which has shown that the disease-associated risk decreased when non-ascertained cohorts of *CDH1* pathogenic variant carriers were used (Roberts et al., 2019).

In summary, carriers of *CDH1* P/LP variants display high but variable DGC and/or LBC associated risk, and the factors that underlie differential risks to develop DGC, LBC or both remain elusive.

3.1.3. CTNNA1/α-E-catenin

The *CTNNA1* gene encodes for the αE-catenin protein, which plays a role in adherens junctions. E-cadherin and αE-catenin act together with p120 and β-catenin, which are αE-catenin binding partners, to mediate the cytoplasmic linkage with the actin cytoskeleton (Takeichi, 2018; van Roy, 2014).

Lobo et al. recently reviewed the literature describing the occurrence of germline *CTNNA1* variants in HDGC families (Lobo et al., 2021). This study concluded that 13 HDGC families have been described to carry *CTNNA1* germline variants (Benusiglio et al., 2019; Clark et al., 2020; Hansford et al., 2015; Majewski et al., 2013; van der Post et al., 2018; Weren et al., 2018), 10 of which classified as P or LP, according to the ACMG/AMP guidelines for *CDH1* variant classification in HDGC (Lee et al., 2018). This literature review also suggested that P, but not LP variants in *CTNNA1* seem to associate with DGC, accounting as a disease cause for <2% of families/individuals fulfilling HDGC criteria (Lobo et al., 2021). Considering that HDGC is very rare in the population, germline pathogenic variants in *CTNNA1* are even rarer but quite well supported by clinical and molecular evidence.

Given the above evidence, *CTNNA1* variant analysis has been introduced in the 2020 HDGC clinical testing guidelines, and the gene is considered as a *bona fide* HDGC predisposing gene (Fig. 2 (Blair et al., 2020)). Supporting this decision is the fact that asymptomatic *CTNNA1* variant carriers were found to display DGC *foci* in prophylactic total gastrectomy specimens, raising the hypothesis that *CDH1* and *CTNNA1* pathogenic variants may have similar behaviour for DGC development (Benusiglio et al., 2019). Thus, families that fulfil HDGC clinical criteria and lack a *CDH1* P/LP SNV or CNV are now advised to undergo *CTNNA1* testing (Fig. 2 (Blair et al., 2020)). *CTNNA1*-positive families described to date display mainly DGC, and a clear association with LBC and cleft lip, with or without cleft palate, has not been reported (Table 1).

Classification using ACMG criteria is not available for *CTNNA1* variants, thus, these have been mainly classified using the general ACMG criteria (Table 1) (Richards et al., 2015). The lack of specific ACMG guidelines for the classification of these variants occurring in HDGC patients and families complicates their interpretation and prediction of actionability. Most of the germline variants carried by HDGC families, reported to date, were truncating, including frameshift and nonsense variants (Table 1) (Benusiglio et al., 2019; Clark et al., 2020; Hansford et al., 2015; Majewski et al., 2013; Weren et al., 2018). A paper addressing specific aspects on *CTNNA1* germline variants in HDGC can be found in this special issue (Lobo et al., 2021). In Lobo *et al.* manuscript, *CTNNA1* germline variants were classified according to specifications of the ACMG/AMP variant curation guidelines for the analysis of germline *CDH1* sequence variants (Lee et al., 2018)

3.2. Management: HDGC families carrying CDH1 and CTNNA1 causative variants

Individuals at risk of HDGC development are clinically managed according to the 2020 HDGC clinical guidelines (Blair et al., 2020). Upon identification of a *CDH1* causal variant, carriers are offered risk-reduction measures available to prevent cancer. These include prophylactic gastrectomy, regardless of endoscopic results, in carriers older than 18 years old, for DGC prevention in both genders; and annual breast magnetic resonance imaging (MRI) surveillance for LBC prevention in females, or bilateral risk-reducing mastectomy (Blair et al., 2020). Considering that DGC may arise at an early age in *CDH1* variant carriers, prophylactic surgery is recommended between age 20 and 30, and before 70 (Blair et al., 2020). For LBC, bilateral-risk reduction mastectomy is recommended in carriers between the ages of 30 and 60 years old (Blair et al., 2020; van der Post et al., 2015a).

CDH1 variant carriers may also undergo DGC and LBC surveillance by endoscopic/multiple biopsies and/or breast MRI while waiting for prophylactic surgeries, if unfit or upon refusal of prophylactic interventions. For DGC, yearly endoscopy performed by professionals with knowledge of HDGC is recommended and repeated after 4–6 weeks if there is a suspicion of infiltrating lesions. The clinical recommendations for surveillance by breast MRI begin from age 30, are yearly performed, and in some instances continue beyond 50 years old (Blair et al., 2020).

For *CTNNA1*-positive families, given the current lack of knowledge, surveillance is favoured over prophylactic surgery for DGC in variant carriers. As LBC predisposition due to *CTNNA1* germline variants has not been documented, BC surveillance in asymptomatic carriers should be considered on a case-by-case basis (Blair et al., 2020).

CDH1-positive HLBC families are advised to undergo annual breast surveillance with MRI or bilateral risk-reducing mastectomy and may be recategorized into HDGC, if a DGC develops or a DGC precursor lesion is identified during surveillance in a carrier family member (Blair et al., 2020).

3.3. Unknowns: CDH1-related HDGC

Although the knowledge on the implications of bearing a germline *CDH1* variant has grown considerably, since the first reports of their causality (Guilford et al., 1998), many uncertainties remain about penetrance, variant actionability, and the spectrum of the disease in families carrying different *CDH1* germline variants. Most of the studies performed to date to understand penetrance in *CDH1* carrier families (Hansford et al., 2015; Pharoah et al., 2001; Roberts et al., 2019; Xicola et al., 2019), are biased towards the previous IGCLC criteria that mainly focused on risk for DGC. The exception is the study performed by Roberts et al., (2019), which predicted penetrance using an unascertained cohort, and demonstrated that disease risk for DGC may be lower than initially predicted using ascertained cohorts (Roberts et al., 2019; van der Post et al., 2015a). However, several aspects related to penetrance in *CDH1* carrier families remain elusive, namely penetrance of confirmed DGC and LBC, penetrance modifiers and variant-specific penetrance.

The *CDH1*-associated disease spectrum, beyond DGC and LBC, also remains unclear, as some apparently HDGC-unrelated early-onset cancers may also develop in carriers. While intestinal-type GC has been completely excluded from the *CDH1* spectrum (Caldas et al., 1999; van der Post et al., 2015b), an association with colorectal cancer (CRC) and ovarian cancer (OC) has yet to be acknowledged (Figueiredo et al., 2019; Oliveira et al., 2002; Richards et al., 1999; Roberts et al., 2019; Stuebs

et al., 2018).

A few genotype-phenotype association studies have been performed, considering the molecular classification of *CDH1* variants (Lo et al., 2019; Selvanathan et al., 2020). Nonetheless, a clear correlation of cancer and non-cancer phenotypes with variant molecular types remains to be performed.

CDH1 missense variants are often identified in HDGC families, but their functional consequence is hard to predict, and therefore their potential causal effect and actionability in HDGC remain a matter of debate. Missense variants do not affect the size of the E-cadherin protein, and many times not even the protein localization, however, E-cadherin-mediated signalling seems to be modified in many instances (Lee et al., 2018; Melo et al., 2017). Whether the latter consequence is sufficient to predispose to HDGC remains unknown, and only genotype-phenotype directed to families and individuals carrying these variants will shed light on their potential causality. To date, it has been reported that missense variants seem to be enriched in HLBC cases (Corso et al., 2018).

Recently, genetic modifiers, epigenetic alterations, or environmental factors have been suggested to play a role in hereditary cancer predisposition (Taeubner et al., 2018). However, how specific genetic and environmental factors, or their combination, contribute to increasing HDGC predisposition, severity, or disease spectrum remains unexplained.

H. pylori and *Epstein-Barr* virus infections have been reported to be shared among individuals in the same environment, and sometimes from the same family (Ellison-Loschmann et al., 2017), however, their contribution to HDGC remains understudied.

3.4. Unknowns: CTNNA1-related HDGC

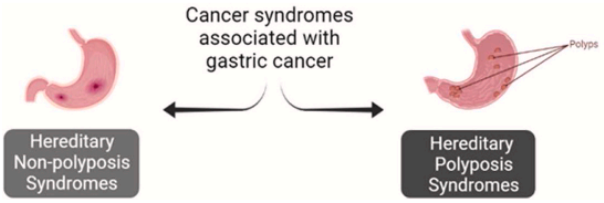
Knowledge on *CTNNA1*-associated HDGC is still very scarce, therefore most of what has been realized for *CDH1*-associated HDGC is yet to be discovered in *CTNNA1*-HDGC. *CTNNA1* carriers are predisposed to DGC, but whether LBC and cleft lip with or without cleft palate are also part of the disease spectrum remains unknown. Disease-risk estimations, variant actionability, genotype-phenotype correlations, and general approaches for surveillance and prophylactic surgery are all currently missing items in knowledge about *CTNNA1*-associated HDGC.

4. Gastric adenocarcinoma and proximal polyposis of the stomach (GAPPS)

GAPPS is a GC-predisposing syndrome with an autosomal dominant inheritance pattern with tropism to the proximal stomach (Worthley et al., 2012).

Predisposition to develop gastric polyps distinguishes GAPPS from other GC syndromes. The spectrum of disease includes fundic gland polyps (FGP), which may show focal foveolar-type dysplasia, hyperplastic polyps, hyperproliferative aberrant pits, adenomas with gastric-type dysplasia, and preinvasive lesions with mixed-type features.

Table 2
TRS without primary predisposition to GC but presenting GC as part of their disease spectrum.



	Syndrome	Inheritance	Cancer Spectrum	Major Associated Genes	References (PMID)
Hereditary Non-polyposis syndromes	Hereditary breast and ovarian cancer syndrome (HBOC)	AD	Breast, ovarian, prostate, pancreatic and gastric cancer	*BRCA1* *BRCA2*	32259785 16369438 12373604
	Li-Fraumeni syndrome (LFS)	AD	Soft tissue sarcoma, female breast cancer, osteosarcoma, CNS tumor, gastric, lung, and adrenocortical cancers; schoroid plexus carcinoma, leukemia and rhabdomyosarcoma of embryonal anaplastic subtype	*TP53*	32457520 32259785 3409256 26014290 21552135
	Lynch syndrome (LS)	AD	Colorectal, endometrial, ovarian, gastric, pancreatic, urothelial, pancreatic, brain (usually glioblastoma), biliary tract and small-bowel cancers; sebaceous adenomas/carcinomas and keratoacanthomas	*MLH1* *PMS2* *MSH2* *EPCAM* *MSH6*	32930401 31487681 28754778
Hereditary Polyposis syndromes	Familial adenomatous polyposis (FAP)	AD	Colorectal, duodenal, gastric, and thyroid cancer; desmoid tumours, osteomas, odontomas, hepatoblastomas and medulloblastomas	*APC*	28185118 1316858 2896968 * NCCN 2020
	Juvenile polyposis syndrome (JPS)	AD	Colorectal, gastric, small intestine and pancreatic cancer	*SMAD4* *BMPR1A*	31108454 25645574 * NCCN 2020
	MUTYH-associated adenomatous polyposis (MAP)	AR	Colorectal and duodenal cancer; gastric polyps and gastric cancer	*MUTYH*	19732775 27194394 * NCCN 2020
	Peutz-Jeghers syndrome (PJS)	AD	Colorectal, gastric, breast, small intestine, pancreatic, ovarian, uterus, cervical, and lung cancer	*STK11*	24143323 20051941 20581245 * NCCN 2020
	PTEN hamartoma tumor syndrome (PHTS)	AD	Breast, endometrial and follicular thyroid cancer; gastrointestinal hamartomas or ganglioneuromas, colon cancer, lipomas, renal cell carcinoma and melanoma	*PTEN*	32533092 32259785

Footnote: AD, Autosomal Dominant; AR, Autosomal Recessive; CNS, Central Nervous System; NCCN 2020 (National Comprehensive Cancer Network, 2021b).

Adenocarcinomas occurring in GAPPS are of intestinal and/or mixed-type (de Boer et al., 2018; Gullo et al., 2021; WHO Classification of Tumours Editorial Board, 2019; Worthley et al., 2012).

Fundic gland polyposis with multifocal dysplasia have been detected as early as 10 years old, however, GC age of onset ranges from 23 to 75 years (Repak et al., 2016; Worthley et al., 2012). GAPPS families display a heterogeneous phenotype, mainly in the number of polyps, that ranges between 30 to hundreds (Rudloff, 2018; Worthley et al., 2012). Intestinal and mixed type GC have been detected in 13% of patients from families with a GAPPS diagnosis ((Worthley et al., 2012) (Table 1)).

Current diagnostic criteria for GAPPS are depicted in Fig. 2 (Rudloff, 2018). The most recent updates in clinical criteria exclude other heritable gastric polyposis syndromes (Table 2), and individuals that use proton pump inhibitors, which may be associated with sporadic FGPs. GAPPS has been considered a variant of Familial Adenomatous Polyposis (FAP) which is restricted to the stomach (Rudloff, 2018).

GAPPS has been shown to be caused by SNVs at the *APC* promoter 1B, and for this reason, *APC* promoter 1B is currently tested upon clinical suspicion of GAPPS (Fig. 1). To date, germline variants at the *APC* promoter 1B hotspot are considered the main, and only known, cause of GAPPS.

4.1. Bona fide GAPPS genes: APC

The *APC* gene encodes the Adenomatous Polyposis Coli protein, a negative regulator of β-catenin and, thus, of the WNT signalling pathway. *APC* downregulation leads to an increase in cytoplasmatic β-catenin, which is translocated to the nucleus, promoting increased proliferation (Ilyas and Tomlinson, 1997).

Germline SNVs in *APC*, with a hotspot in promoter 1B (YY1 Transcription Factor Binding motif), have been identified as the main trigger to develop GAPPS (Table 1 (Li et al., 2016)). In the stomach, promoter 1B has ~15-fold higher transcription than promoter 1A, due to methylation of the latter in cells from the gastric mucosa (Hosoya et al., 2009; Rohlin et al., 2011). Therefore, lack of YY1 transcription factor binding may lead to reduced promoter 1B activity and reduced *APC* tumour suppressive phenotype in the stomach. In the colon, there is normal promoter 1A-driven transcription, which compensates for the lack of promoter 1B-driven transcription. This provides the rationale for a GAPPS-restricted, or at least preferential, gastric phenotype (Rudloff, 2018). However, identical promoter 1B variants have been identified in <1% of classical FAP families with both gastric and duodenal involvement (Snow et al., 2015).

In 2016, Li J. and colleagues found three germline SNVs and one small deletion in *APC* promoter 1B in an Australian family (c.-195A>C co-segregated with c.-125delA), an American family (c.-191T>C), and a Canadian family (c.-192A>G) (Li et al., 2016). These variants were considered P/LP and accepted as disease-causing (Landrum et al., 2014). A combination of variants at the *APC* promoter 1B (e.g., c.-195A>C and c.-125delA) was found associated with severe phenotypes, but their individual contribution remains unknown (Li et al., 2016). The variant c.-191T>C was found in 12 additional families from different ancestries, indicating a potential hotspot (Foretová et al., 2019; Kanemitsu et al., 2021; Mitsui et al., 2018; Repak et al., 2016).

GAPPS-associated phenotypes and disease expressivity may be heterogeneous, even among families carrying the same causal variant. While some patients may develop very early onset GC (Repak et al., 2016), old carriers remain disease-free (Foretová et al., 2019). Supporting incomplete penetrance is the fact that patients from the 2nd and 3rd generations present with severe phenotypes, while the 1st generation obligate carriers remain unaffected (Worthley et al., 2012). Altogether, these observations suggest that genetic modifiers or environmental factors may play a role in GAPPS cancer predisposition and penetrance.

Currently, ACMG clinical criteria for *APC* variants classification are only available for FAP syndrome (Table 1 (Hegde et al., 2014)).

4.2. Management: GAPPS

Given the low number of GAPPS cases reported and patient heterogeneity, the management of GAPPS families is mostly carried out on a case-by-case basis (Tacheci et al., 2021). Endoscopic surveillance with biopsies is recommended and prophylactic gastrectomy is offered to GAPPS families without a clear timing guideline (Li et al., 2016; Repak et al., 2016; Worthley et al., 2012). Nevertheless, prolonged surveillance in FGP patients is not recommended, due to a potential rapid progression. Further, first-degree relatives of families diagnosed with GAPPS, carrying the disease-associated genotype, are advised to perform upper gastrointestinal endoscopy and a colonoscopy (Rudloff, 2018; Worthley et al., 2012).

4.3. Unknowns: APC-related GAPPS

Twenty-five families, encompassing 110 patients that fulfilled GAPPS criteria, have been reported across the globe, namely from Australia, North America, Europe, and Japan (Beer et al., 2017; Declich et al., 2013; Foretová et al., 2019; Kanemitsu et al., 2021; Kunovsky et al., 2019; Li et al., 2016; Mitsui et al., 2018; Repak et al., 2016; Rudloff, 2018; Tacheci et al., 2017, 2021; Worthley et al., 2012; Yanaru-Fujisawa et al., 2012). The number of polyposis-affected patients in GAPPS families has likely been underestimated due to a bias in surveillance and diagnosis towards the stomach (Tacheci et al., 2021), and because colonic polyposis cases were not considered as part of the GAPPS disease spectrum (Li et al., 2016). For this reason, the prevalence and penetrance of GAPPS-related phenotypes remain vastly unknown (Jasperson et al., 1993; Li et al., 2016; Tacheci et al., 2021).

Although the somatic landscape of GAPPS has started to be disclosed, more studies are needed to distinguish between sporadic FGPs and GAPPS-associated FGPs (Kadiyska et al., 2014; Lagarde et al., 2010; Pavicic et al., 2014).

Even though *H. pylori* infection is thought to be absent in GAPPS patients, sparse data is available highlighting the lack of knowledge regarding the role of this and other environmental factors in tumour development in the context of this disease (Rudloff, 2018; Tacheci et al., 2021; Worthley et al., 2012).

5. Familial intestinal gastric cancer (FIGC)

FIGC has been generally understood as an autosomal dominant disease with one or several family members affected by intestinal type GC (IGC) (Caldas et al., 1999). However, no global consensus has been reached regarding clinical criteria for this disease. It is generally accepted that criteria analogous to the Amsterdam criteria for hereditary non-polyposis colorectal cancer are used in countries where GC incidence is high (e.g., Japan, China, Korea, Portugal). These criteria become more restrictive in countries with low GC incidence (e.g., USA, Canada, UK) (Caldas et al., 1999), as depicted in Fig. 2.

Whist germline variants at *CDH1*, *CTNNA1,* and *APC* genes are well established as the main triggers to develop HDGC and GAPPS, FIGC remains almost completely genetically unexplained, with an unknown tumour spectrum and age of onset. Moreover, given the lack of FIGC-causal genes, clinical criteria have not been validated nor revisited since 1999 (Caldas et al., 1999).

To date, rare patients or families suspected of FIGC or displaying isolated IGC at an early age have been shown to display germline potentially causative variants in *BRCA2*, *BRCA1*, *BRIP1,* and *MSH2* (Carvalho et al., 2021; Ichikawa et al., 2018; Lee et al., 2000; Sahasrabudhe et al., 2017; Slavin et al., 2017; Tedaldi et al., 2019; Weren et al., 2018). However, both the diversity of genes apparently involved, and the fact that these genes predispose to different diseases, have hampered the classification of these candidate genes as FIGC-associated.

A recent study pinpointed FIGC as an independent clinical and molecular entity, differing at the molecular level, from both sporadic IGC

and HDGC (Carvalho et al., 2021). The average age of onset of IGC in FIGC families was shown to be considerably higher than that of HDGC, but at least 10 years lower than that of sporadic IGC. In addition, the germline, and somatic mutational landscapes of FIGC patients and tumours were shown to be distinctive from those of sporadic IGC. Specifically, FIGC probands carried far more *TP53* germline common variants than sporadic IGC cases; and far more somatic common variants in *BRCA2*, *ATM*, *FOXF1*, *FHIT*, *SDHB*, *MSH6*, *CTNNA1,* and *PXN*. Also, the MSI frequency was significantly higher in FIGC tumours, as compared to the sporadic counterpart (Carvalho et al., 2021). The tumour spectrum of FIGC families was shown to be broad, predisposing primarily to the stomach, but very often to CRC and BC (Carvalho et al., 2021). As high penetrance gene defects were rare in FIGC, Carvalho et al. suggested FIGC as a polygenic disease (Carvalho et al., 2021), a thesis that needs further research to be proven.

Altogether, these findings claim for a proposal to redefine FIGC clinical criteria. Carvalho and colleagues suggested that a minimum of two cases of GC diagnosed at any age, with one GC being histologically confirmed IGC, would be needed for clinical confirmation of FIGC (Carvalho et al., 2021).

5.1. Management

Due to the scarce information about genetic causes and disease spectrum, the management of FIGC patients is not clear. Notwithstanding, Sereno and colleagues recommended periodic endoscopy in first-degree relatives, starting 10 years earlier than the age of onset of the youngest family member diagnosed with IGC (Sereno et al., 2011). According to Corso et al., 2013, yearly gastroduodenoscopy starting at 40 years of age or 5 years earlier than the age of onset of the youngest family member diagnosis, could also be accepted. Given the high rate of *H. pylori* infection, eradication is recommended in these families (Corso et al., 2013).

5.2. Unknowns: FIGC

FIGC remains vastly unknown and genetically unresolved to date. Consequently, no objective and targeted clinical guidelines or criteria for testing can be designed. Prevalence and penetrance are unknown, and management is suboptimal. The tumour spectrum also remains unclear, and no consensus exists on the most probable age of onset of IGC in the context of FIGC, so that cancer screening and surveillance can be optimized.

6. Inherited gastric cancer in other cancer-associated syndromes

GC and further gastric lesions can still be observed in other inherited cancer disorders as part of their tumour spectrum such as hereditary breast and ovarian cancer (HBOC), Li–Fraumeni syndrome (LFS), Lynch syndrome (LS), familial adenomatous polyposis (FAP), juvenile polyposis syndrome (JPS), MUTYH-associated polyposis (MAP), Peutz-Jeghers syndrome (PJS) and PTEN hamartoma tumour syndrome (PHTS) (Table 2). Despite not being the primary cancer, a noticeable clustering of GC cases has been observed in subsets of families with these TRS. The lifetime risk for developing GC in these heritable syndromes is also highly variable. Based on the syndromes presented in Table 2, the risk for GC can be elevated for JPS, PJS, and LS, ranging from 10% to nearly 30%. On the other hand, current literature suggests that GC is likely to occur rather infrequently in FAP, LFS, MAP, PHTS, or HBOC, and be largely dependent on environmental risk factors, linked or not with geographical background (Setia et al., 2015). This has recently been described for LFS, where a comparison of cancer spectrum in *TP53* germline carrier individuals from Japan and France revealed stomach cancer being significantly more frequent in Japan than in France (7.8% vs 1.3%) (Funato et al., 2021). Besides distinct disease spectra and mutated genes, the geographic origin of the population where these TRS occur also affects the GC risk (Capelle et al., 2010; Ikenoue et al., 2019). GC surveillance measures may be stimulated in cases where GC occurs more frequently than expected in families diagnosed with one of the above described TRS, or if patients and their relatives (affected by a TRS above) live in a country with high GC incidence (Vasen et al., 2013).

6.1. Cancer predisposition in HDGC and FIGC: candidate genes

GC, genetically determined from birth, is a rare occurrence, and the well-established predisposing genes described above are still insufficient to explain disease in many of these rare patients and families. Indeed, 60–90% of HDGC, some GAPPS, and almost all FIGC patients/families remain unsolved and expect their cancer heritability to be discovered. So, one of the most important questions in the field is how additional genes (candidate genes), with the potential to confer a high predisposition to GC, could be found in molecularly unexplained families. One possibility is that the monogenic nature of disease in unsolved cases, is explained not by the major gene in that syndrome (e.g., *CDH1* in HDGC), but by an array of different predisposing genes, some of them known to predispose to other TRS (e.g., different MMR genes in LS). On the other hand, GC may occur as part of the cancer spectrum in other syndromes (as described in Table 2). Therefore, GC isolated cases, of either intestinal or diffuse histology, detected at an early age with unknown or unreported family history, prior to genetic testing, may hide family histories compatible with classification into other TRS.

To address these possibilities, we screened the literature for reports on next-generation sequencing (NGS) studies, of either gene-panels, exomes, or whole genomes which described potentially actionable germline variants in families that could fulfil criteria of GC-associated TRS. This analysis has two main limitations. One is related to FIGC: because no clinical criteria exist to select families at risk of FIGC, a few cases have been submitted for NGS studies. The other limitation is related to GAPPS: as most GAPPS families are reported to carry promoter 1B *APC* variants, and GAPPS is very rare, few families remain unsolved.

The search strategy herein used relied on very specific requirements. Selected cases needed to be considered genetically solved and present sufficient clinical information for allocation in at least one of the following criteria: (1) families/individuals fulfilling IGCLC-HDGC criteria 2020 (Blair et al., 2020); (2) isolated cases of IGC <70 years old or families with $\geq$ 2 GC cases one confirmed IGC, herein classified as FIGC, and; (3) families/individuals fulfilling GAPPS 2018 criteria. Selected studies are depicted in Table 3 and Supplementary Table S1. Data is presented by affected gene and its association with the above-listed criteria. Selection of publications was performed with PUBMED.

In PUBMED, we used as QUERY "Familial gastric cancer germline mutations" to search for manuscripts from 1993 to 2021. We retrieved 595 publications that were manually inspected to identify the above listed requirements. Fourteen publications were selected: 26182300 (Hansford et al., 2015), 28024868 (Sahasrabudhe et al., 2017), 25340522 (Gaston et al., 2014), 29706558 (Fewings et al., 2018), 28875981 (Vogelaar et al., 2017), 29330337 (Weren et al., 2018), 11042681 (Lee et al., 2000), 31608315 (Ichikawa et al., 2018), 29025585 (Slavin et al., 2017), 25576241 (Donner et al., 2015), 12920089 (Kim et al., 2003),15173255 (Keller et al., 2004), 32066632 (Carvalho et al., 2021), 33525650 (Herrera-Pariente et al., 2021).

Additionally, two manuscripts were added manually, one was *in press* ((Te Paske et al., 2021), PMID 34075207)) and the other was published but failed to be retrieved with the above query ((Tedaldi et al., 2019), PMID 31514334). The authors, due to co-authorship, knew both manuscripts.

We identified 71 distinct families, fulfilling criteria for either HDGC or FIGC, and bearing 66 variants in 34 different genes. After classifying all variants, using the ACMG guidelines (Richards et al., 2015), we ended up with 27 families carrying 28 different P/LP variants across 16 candidate genes (Fig. 3, Table 3). Thirteen out of these 16 (81%)

Table 3
Summary, description and classification of variants found in individuals fulfilling HDGC criteria or FIGC crieria (with familial aggregation of GC with one case being IGC or isolated cases of IGC <70). Phenotypes present in different inviduals are separated by '; ' and phenotypes present in the same individual are separated by ',''. Probands are depicted in bold and '+' means a carrier of the variant not proband.

Gene	mRNA position HGVS	αα position HGVS	Molecular type	Primary syndrome	Alternative syndromes	Spectrum	Probands information	Relatives information	ACMG criteria	ACMG classification	References (PMID)
ATM	c.7705_7706del	p.Asp2569*	Frameshift	HDGC	NA	GC	**DGC 66y**	GC; GC; GC; GC; GC; GC; GC	PVS1, PM2, PS4	P	29025585
	c.1564_1565del	p. Glu522Ilefs*43	Frameshift	HDGC	FIGC	GC	IGC 54y	≥2 GC cases regardless of age, 1 confirmed DGC; ≥2 GC cases ≤60 years in 1st-degree relatives; 1 IGC 54y	PVS1, PM2, PS4	P	31514334
	c.2192dup	p.Tyr731*	Nonsense	HDGC	NA	GC	**DGC 32y**	**No additional phenotypes reported**	PVS1, PM2	LP	31514334
ATR	c.6075T>A	Tyr2025*	Nonsense	HDGC	NA	GC/BC	**DGC 28y (#)**	SRCC 45y; DGC 60y^{+}; SRCC; SRCC; CRC 75y; PRSC 60s; BC 40s	PVS1, PM2	LP	29706558
BLM	c.2395del	p. Cys799Valfs*16	Frameshift	HDGC	NA	GC/BC	**LBC 50y, DGC 54y**	≥2 GC cases regardless of age, 1 confirmed DGC	PM2, PVS1	LP	31514334
BRCA1	c.188T>A[a]	p.Leu63*	Nonsense	FIGC	HBOC	GC/BC	**IGC 66y**	BC 37y; CXC 44y; LC; GC	PM2, PS4, PVS1; PP5	P	31608315
	c.406del	p. Arg136Aspfs*27	Frameshift	HDGC	NA	GC	**DGC 47y**	GC	PVS1, PM2. PP5	P	31514334
BRCA2	c.6037A>T	p.Lys2013*	Nonsense	FIGC	NA	GC	**IGC 60y**	GC	PM2, PS4, PVS1	P	31514334
	c.89dupA	p.Leu29fs	Frameshift	FIGC	NA	GC/BC	**IGC 73y**	GC 50y; GC 55y; BC 50y	PM2, PVS1	LP	31608315
	c.6922A>T[a]	p.Lys2308*	Nonsense	FIGC	HBOC	GC/BC	**IGC 66y**	BC 37y; CXC 44y; LC; GC	PM2, PVS1	LP	31608315
	c.2808_2811del	p. Ala938Profs*21	Frameshift	FIGC	NA	GC	**IGC 58y, ESC 58y**	**No additional phenotypes reported**	PM2, PS4, PVS1, PP5	P	29025585
	c.3860_3863delATAA	p. Asn1287Ilefs*5	Frameshift	HDGC	NA	GC/BC	**DGC 64y**	GC 21y; BC 50y	PM2, PVS1, PP5	P	26182300
BRIP1	c.124_127TGTT/ c.128_131delTGTT	p.Leu43fs*11	Frameshift	FIGC	NA	GC	**IGC 19y**	**No additional phenotypes reported**	PM2, PVS1; PS4_mod	P	29025585
FANCC	c.1642C>T	p.Arg548*	Nonsense	HDGC	NA	GC	**DGC 50y (poorly differentiated)**	GC	PVS1, PS3, PM4, PM2	P	29025585
FLCN	c.59del	p.Phe20fs*35	Frameshift	HDGC	NA	GC	**DGC 17y**	**No additional phenotypes reported**	PVS1, PM2, PS4	P	29025585
MSH2	c.367-?_645+?del	NA	Frameshift	FIGC	NA	GC	**IGC 57y**	CRC	PM2, PVS1, PP5	P	31514334
NBN	c.1124+1G>C	NA	Splice-site	HDGC	NA	GC/BC	**DGC 28y (#)**	SRCC 45y; DGC 60y^{+}; SRCC; SRCC; CRC 75y; PRSC 60s; BC 40s	PVS1, PM2	LP	29706558
PALB2	c.1240C>T	p.Arg414*	Nonsense	FIGC	NA	GC	**IGC 69y**	**No further information available**	PVS1, PM2, PS4	P	28024868
	c.1192delG	p.Val398fs	Frameshift	HDGC	HBOC	GC/BC	**DGC 45y**	GC 38y; BC 42y; BC 45y; CRC 59y; PRSC 75y	PVS1, PM2	LP	26182300
	c.1240C>T	p.Arg414*	Nonsense	HDGC	NA	Unk	**DGC 48y**	**No further information available**	PVS1, PM2, PS4	P	28024868
	c.3201+1G>T	NA	Splice-site	HDGC	LS	GC	**DGC 50y**	CRC 60y; GC 60y; GC 50y; GC; GC 50y; GC 52y; GC 57y; GC 58y; GC 58y; GC 50y; GC 52y; GC 50y	PVS1, PM2	LP	28024868
	c.755_756CT	p.Leu253fs	Frameshift	HDGC	NA	GC/BC	**DGC 55y**	DGC 53y; DGC 44y; BC; GC; LC; LRYC	PVS1; PM2, PS4	P	29706558
	c.535C>T	p.Gln179*	Nonsense	HLBC	NA	BC	**BLBC 62y, 66y**	**No additional phenotypes reported**	PVS1, PM2	LP	31514334
	c.2718G>A	p.Trp906*	Nonsense	HLBC	NA	BC	**LBC 62y**	Bilateral LBC or family history of ≥ 2 LBC cases <50y	PVS1, PM2, PS4	P	31514334
PMS2	c.2182_2183delAC	p.Thr728Serfs*7	Frameshift	HDGC	NA	GC	**DGC 73y**	GC	PVS1, PM2	LP	31514334

(*continued on next page*)

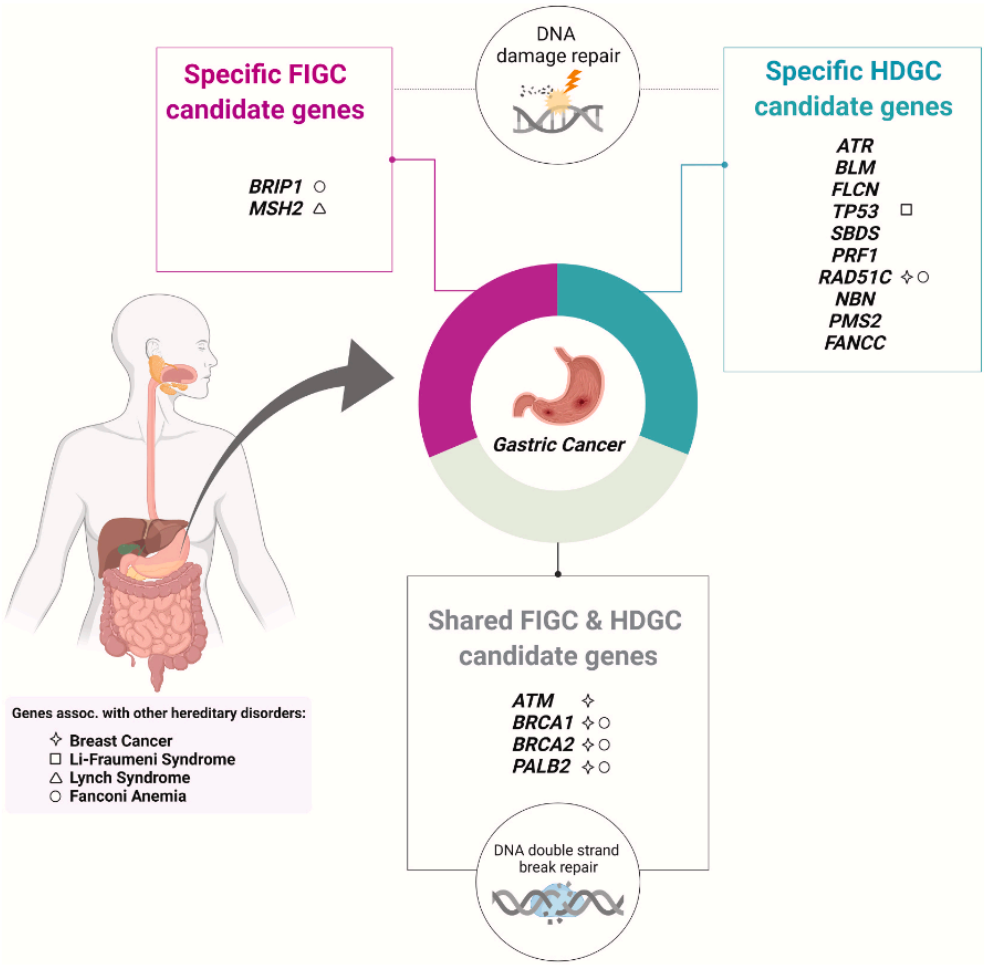


Fig. 3. List of candidate genes retrieved from our literature review and its associated biological processes.

early-onset GC, as a moderate-penetrant gene (Setia et al., 2020; Zhou et al., 2020). *ATM* encoded protein is a cell cycle checkpoint kinase that regulates DNA damage response through p53 and BRCA1 (Jin and Oh, 2019). Three germline *ATM* variants were identified and classified as P/LP (Table 3). From those, two variants were identified in families meeting HDGC clinical criteria, while one was found in a family that met both HDGC and FIGC selection criteria (Table 3). *ATM,* therefore, accounted for 3% (3/94) of HDGC families, and somatic variants were also found in sporadic DGC (9%; 24/268; Table 4, Supplementary Table S2) (Garcia-Pelaez et al., 2021). The importance of *ATM* variants in GC has been a question in debate since several cancer patients have been reported to carry both germline and somatic *ATM* variants, namely lymphomas, leukemia, hepatocellular carcinoma, CRC, GC, BC, among others (Beà et al., 2013; Choi et al., 2016; Hansford et al., 2015; Lin et al., 2019; Navrkalova et al., 2013; Pearlman et al., 2017; Tung et al., 2015, 2016). Although these data support *ATM* involvement in HDGC, its recognized pleiotropy hampers further conclusions.

The *PRF1* mapped in chromosome 10 is an additional candidate gene. It encodes Perforin 1, an important protein to resolve virus infections and to remove malignant cells from the body through a Ca^{2+}-dependent pore-forming mechanism (Law et al., 2010; Praper et al., 2011). *PRF1* variants are considered predisposing to hereditary cancer predisposition and lymphoid diseases (OMIM N° 609135; 603553; 605027), although further studies are required to explore the role of *PRF1* in GC predisposition. Besides a P variant herein found in an HDGC family (Table 3), other GC cases were described in families harbouring *PRF1* germline variants (Chaudhry et al., 2016; El Abed et al., 2011). Recently, *PRF1* levels were found to be reduced in DGC patients and associated with a less functional $CD8^{+}$T cells population in DGC tumours (Li et al., 2020). The frequency of *PRF1* variants is similar between sporadic DGC (4%; 4/106) and IGC (6%; 1/18) (Table 4, Supplementary Table S2) (Garcia-Pelaez et al., 2021). This single finding does not allow a conclusion on the involvement of *PRF1* in HDGC but collected evidence worth further research on this candidate gene.

We also found five P/LP variants in *BRCA2*, one occurring in a family fulfilling HDGC criteria and presenting BC involvement; and four fulfilling FIGC criteria from which two had BC involvement. Germline *BRCA2* (13q13.1) and *BRCA1* (17q21.31) alterations are the hallmark of autosomal dominant HBOC predisposition (Table 2; OMIM N° 604370, 612555, 114480), and associated with autosomal recessive Fanconi Anaemia (OMIM N° 617883, 605724). Further, both genes have been linked with an increased risk of developing GC before 65 years of age (Jakubowska et al., 2002; Johannsson et al., 1999; Roy et al., 2011; Takaoka and Miki, 2018). The fact that there are 4/23 (17%) FIGC families carrying *BRCA2* P/LP variants, supports *BRCA2* as a predisposing gene in FIGC (Table 3). Further supporting this relationship, is the fact that *BRCA2* somatic variants are found in 9% (25/279) of sporadic IGC and only in 5% (15/298) of sporadic DGC (Table 4, Supplementary Table S2) (Garcia-Pelaez et al., 2021).

Regarding *BRCA1*, we found two P/LP variants in two families, one

Table 4
Summary of candidate genes identified in HDGC and FIGC families.

Candidate Genes			Total nº of families screened	GC-associated syndromes in carriers of P/LP variants per gene	HDGC				FIGC				Non-assertained		References (PMID)
Gene bearing germline variant	Main described function	Main associated syndrome (OMIM)			Families screened per gene fulfilling HDGC clinical criteria	HDGC families bearing P/LP variants	Frequency of P/LP variants in HDGC families per gene (%)	Mutation frequency in sporadic DGC cases [PMID: 33724653] (%)	Families screened per gene fulfilling FIGC clinical criteria	FIGC families bearing P/LP variants	Frequency of P/LP variants FIGC families per gene (%)	Mutation frequency in sporadic IGC cases [PMID: 33724653] (%)	Nº of non-ascertain families	Non-ascertained families bearing P/LP variants	
PALB2	TS; Homologous recombination and repair of DNA, colocalized with *BRCA2*	Breast cancer, susceptibility to (AD; OMIM ID: 114480); Fanconi anemia, complementation group N (OMIM ID: 610832); Pancreatic cancer, susceptibility to, 3 (OMIM ID: 613348)	462	HDGC, HLBC, FIGC, HBOC, LS	275	6	2%	NA	11	1	9%	1% (1/75; single publication)	7	1	31514334, 26182300, 28024868, 29706558
FANCC	Delays the onset of apoptosis and promotes homologous recombination and repair of DNA	Fanconi anemia, complementation group C (AR; OMIM ID: 227645)	51	HDGC	16	1	6%	2.9% (2/72; single publication)	3	0	0%	NA	14	0	29025585
BRIP1	Double-strand break repair function	Breast cancer, early-onset, susceptibility to (AD; OMIM ID: 114480); Fanconi anemia, complementation group J (OMIM ID: 609054)	51	FIGC	16	0	0%	3% (1/34; single publication)	6	1	17%	NA	14	0	29025585
BRCA1	TS; Genome stability maintenance, specifically the homologous recombination and repair of DNA	Breast-ovarian cancer, familial, 1 (AD; OMIM ID: 604370); Fanconi anemia, complementation group S (AR; OMIM ID: 617883); Pancreatic cancer, susceptibility to, 4 (OMIM ID: 614320)	103	HDGC, FIGC, HBOC	78	1	1%	4% (4/93)	17	1	6%	8% (7/93)	7	0	31514334, 31608315
BRCA2	TS; Genome stability maintenance, specifically the homologous recombination and repair of DNA	Breast cancer, male, susceptibility to (AD; OMIM ID: 114480); Breast-ovarian cancer, familial, 2 (AD; OMIM ID: 612555); Fanconi anemia, complementation group D1 (AR; OMIM ID: 605724); Glioblastoma 3 (AR; OMIM ID: 613029); Medulloblastoma (AD; OMIM ID: 155255); Pancreatic cancer 2 (OMIM ID: 613347); Prostate cancer (AD; OMIM ID: 176807); Wilms tumor (OMIM ID:194070)	298	FIGC, HDGC, HBOC	238	1	0%	5% (15/298)	23	4	17%	9% (25/279)	21	0	31514334, 26182300, 31608315, 29025585

(*continued on next page*)

Table 4 (*continued*)

Candidate Genes			Total n° of families screened	GC-associated syndromes in carriers of P/LP variants per gene	HDGC				FIGC				Non-assertained		References (PMID)
Gene bearing germline variant	Main described function	Main associated syndrome (OMIM)			Families screened per gene fulfilling HDGC clinical criteria	HDGC families bearing P/LP variants	Frequency of P/LP variants in HDGC families per gene (%)	Mutation frequency in sporadic DGC cases [PMID: 33724653] (%)	Families screened per gene fulfilling FIGC clinical criteria	FIGC families bearing P/LP variants	Frequency of P/LP variants FIGC families per gene (%)	Mutation frequency in sporadic IGC cases [PMID: 33724653] (%)	N° of non-ascertain families	Non-ascertained families bearing P/LP variants	
RAD51C	Homologous recombination and repair of DNA	Breast-ovarian cancer, familial, susceptibility to, 3 (OMIM ID: 613399); Fanconi anemia, complementation group O (AR; OMIM ID: 613390)	200	HDGC, LS	31	1	3%	NA	0	0	NA	NA	0	0	28024868
ATM	Cell cycle checkpoint kinase required for cell response to DNA damage and for genome stability.	Ataxia-telangiectasia (AR; OMIM ID: 208900); Breast cancer, susceptibility to (AD; OMIM ID: 114480)	147	HDGC, FIGC	94	3	3%	9% (24/268)	14	0	0%	8% (11/140)	21	0	31514334, 29025585
ATR	DNA damage response	Seckel syndrome 1 (AR; OMIM ID: 210600); Cutaneous telangiectasia and cancer syndrome, familial (AD; OMIM ID: 614564)	22	HDGC	22	1	5%	4% (4/106)	0	0	NA	8% (6/75; single publication)	0	0	29706558
MSH2	TS; Component of the post-replicative DNA mismatch repair system	Colorectal cancer, hereditary nonpolyposis, type 1 (AD; OMIM ID: 120435); Mismatch repair cancer syndrome 2 (AR; OMIM ID: 619096); Muir-Torre syndrome (AD; OMIM ID: 158320)	96	FIGC	78	0	0%	2% (2/106)	11	1	9%	6% (1/18; single publication)	7	0	31514334
PMS2	Component of the post-replicative DNA mismatch repair system	Mismatch repair cancer syndrome 4 (AR; OMIM ID: 619101); Colorectal cancer, hereditary nonpolyposis, type 4 (OMIM ID: 614337)	96	HDGC	78	1	1%	2% (2/122; single publication)	11	0	0%	3% (4/140)	7	0	31514334
SBDS	Essential role in ribosome biogenesis	Shwachman-Diamond syndrome (AR; OMIM ID:260400); Aplastic anemia, susceptibility to (OMIM ID: 609135)	51	HDGC	16	1	6%	3% (2/72; single publication)	3	0	0%	NA	14	0	29025585
BLM	Genome stability maintenance	Bloom syndrome (AR; OMIM ID: 210900)	96	HDGC	78	1	1%	4% (4/106)	11	0	0%	4% (4/93)	7	0	31514334
PRF1	Involved in cytolysis and in the regulation of the immune system.	Hemophagocytic lymphohistiocytosis, familial, 2 (AR; OMIM ID: 603553); Aplastic anemia (OMIM ID: 609135);	96	HDGC	78	1	1%	4% (4/106)	11	0	0%	6% (1/18; single publication)	7	0	31514334

(*continued on next page*)

Table 4 (*continued*)

Candidate Genes			Total nº of families screened	GC-associated syndromes in carriers of P/LP variants per gene	HDGC				FIGC				Non-assertained		References (PMID)
Gene bearing germline variant	Main described function	Main associated syndrome (OMIM)			Families screened per gene fulfilling HDGC clinical criteria	HDGC families bearing P/LP variants	Frequency of P/LP variants in HDGC families per gene (%)	Mutation frequency in sporadic DGC cases [PMID: 33724653] (%)	Families screened per gene fulfilling FIGC clinical criteria	FIGC families bearing P/LP variants	Frequency of P/LP variants FIGC families per gene (%)	Mutation frequency in sporadic IGC cases [PMID: 33724653] (%)	Nº of non-ascertain families	Non-ascertained families bearing P/LP variants	
		Lymphoma, non-Hodgkin (OMIM ID: 605027)													
FLCN	TS; Cellular response to amino acid availability, modulation of EGFR and mTOR signalling pathways and mTOR signalling pathways	Birt-Hogg-Dube syndrome (AD; OMIM ID: 135150); Pneumothorax, primary spontaneous (AD; OMIM ID: 173600)	51	HDGC	16	1	6%	1% (1/72; single publication)	3	0	0%	5% (5/93)	14	0	29025585
TP53	TS; Regulates cell cycle arrest, apoptosis, senescence, DNA repair, or changes in metabolism; controls the expression of target genes	Li-Fraumeni syndrome (AD; OMIM ID: 151623); Adrenocortical carcinoma, pediatric (AD; OMIM ID: 202300); Basal cell carcinoma 7 (AD; OMIM ID: 614740); Choroid plexus papilloma (AD; OMIM ID:260500); Colorectal cancer (AD; OMIM ID: 114500); Glioma susceptibility 1 (AD; OMIM ID: 137800); Bone marrow failure syndrome 5 (AD; OMIM ID: 618165)	51	HDGC, HBOC	16	1	6%	41% (167/407)	3	0	0%	57% (132/232)	14	0	29025585
TOTAL						20				8				1	

Footnote: Approximate mutation frequencies for sporadic DGC/IGC are represented by the average frequencies across the studies reported by García-Pelaez et al. (2021). Information for the main function and associated syndromes of the abovementioned genes was collected using the websites: https://www.omim.org/; https://www.genecards.org/and https://www.uniprot.org/. HDGC, Hereditary diffuse gastric cancer; FIGC, Familial intestinal gastric cancer; HBOC, Hereditary breast and ovarian cancer; LS, Lynch syndrome; LFS, Li-Fraumeni syndrome; TS, Tumour suppressor; AR, Autosomal recessive; AD, Autosomal dominant; NA, non-applicable.

HDGC family without BC involvement and one FIGC family (Table 3). This FIGC family harbours co-occurring P/LP variants in *BRCA1* and *BRCA2*, where the proband was a 66 years old IGC patient, but also displaying a family history compatible with HBOC, as a family member developed BC at 37 years of age (Table 3). This specific case poses an important question since a different proband choice (for instance the 37-years BC case) would have potentially triggered the suspicion of a different TRS and, consequently, different clinical management. A similar outcome was identified in a *MSH2* variant for LS, where a deeper pedigree and segregation analysis revealed co-segregation with different cancers of the LS spectrum, including IGC and CRC (Table 3) (Tedaldi et al., 2019). Further, our analysis identified a single HDGC family carrying a LP in *PMS2*, and lacking CRC in the reported family pedigree. It is possible that the DGC proband carrying the *PMS2* LP variant and its relative with GC, belong to a branch of a LS family, where GC presents as part of the spectrum, but such information is not available in the original publication (Table 3). Additionally, this could also be the case of both, early-onset DGC and BC diagnosis, in a proband with a *TP53* P variant, (Table 3). This well-known tumour suppressor located on chromosome 17 is the cause of LFS. Its encoded protein is responsible for the regulation of cellular responses to stress, including DNA repair, cell cycle arrest, and apoptosis (Vaseva et al., 2012). Several human cancers have been associated with *TP53* variants (OMIM Nº 114480, 114550, 607107, 260350) including hereditary cancers (Table 2, OMIM Nº 202300, 614740, 260500, 114500, 137800, 259500, 618165, 151623). Regarding our analysis, as *TP53* is considered a high penetrant susceptibility gene for BC, LFS syndrome could be implied (Daly et al., 2021). Given the lack of further family information, we support the possibility of a *TP53* variant in the context of early-onset BC or LFS, with involvement of DGC, rather than claiming *TP53* as a candidate gene for HDGC (Kim et al., 2004).

Altogether, these observations reinforce the usefulness of panel testing/exome sequencing for syndromes with overlapping cancer spectra, and the need for collection and publication of extended family pedigrees.

Finally, an IGC case at 19 years of age was found in a proband carrying a *BRIP1* (17q23.2) P variant (Table 3). *BRIP1* is also known to be associated with Fanconi Anaemia following OMIM (OMIM Nº 609054). This gene interacts with *BRCA1* and regulates *TP53*. In sporadic GC cases, *BRIP1* is specifically mutated in DGC cases (3%; 1/34), but not in IGC, which does not add to its potential role in FIGC (Table 4, Supplementary Table S2) (Garcia-Pelaez et al., 2021).

In summary, 16 genes bearing 28 P/LP variants were found in HDGC and/or FIGC representing potential candidate genes, and mainly associated with DNA repair pathways. Homologous recombination-related genes were most often targeted than mismatch repair pathway genes in both GC-TRS. Moreover, *PALB2* and *BRCA2* were the most frequently and consistently mutated candidate genes in HDGC (2%) and FIGC (17%), respectively. Evidence for a direct correlation with either FIGC or HDGC for other genes was not evident from our analysis and supports the possibility of GC developing in the frame of families with unreported susceptibility to another TRS, and with GC involvement.

6.2. *Impact of candidate gene analysis for clinical management of carriers*

This study highlights a reasonable number of GC cases in families bearing germline variants in genes from the DNA repair machinery. This potential susceptibility may be a challenge for the management of these families, as it adds a new organ to be surveilled. This would be the case for *PALB2*, *ATM*, and *BRCA2* variant carriers, in case these become formally associated with GC risk in larger and deeper studies. As we stand, it may be adequate to offer gastric surveillance with annual endoscopy in families bearing P/LP *PALB2*, *ATM*, and *BRCA2* in case family history reveals a high number of GC cases, independently of the histological subtype and especially in countries of high incidence.

7. Conclusion

The acknowledgement of hereditary GC is crucial to reduce GC late diagnosis and, consequently its mortality. While SNVs and CNVs have been identified in HDGC (*CDH1* and *CTNNA1* genes) and GAPPS families (*APC* promoter 1B), FIGC families remain genetically undetermined. Thus, efforts should be made to identify the underlying FIGC germline cause, establish a genetic screening protocol, and update clinical guidelines, to improve the management of these families. For HDGC and GAPPS, further research is still crucial to identify variant-specific penetrance and the significance of genetic modifiers, epigenetic alterations, or environmental factors in disease-associated risk.

Considering the complexity of these diseases independently of the GC syndrome, it is relevant to reinforce the need for patients' management to include multidisciplinary teams.

PALB2 was the most promising candidate gene in HDGC, while *BRCA2* was the most promising candidate gene in FIGC. Nevertheless, the association of these candidate genes with these TRS needs further validation, including segregation analysis, second somatic hit analysis, penetrance, and life-time risk calculations.

As we anticipated, the lack of extended family pedigrees prior to genetic testing prescription likely hampers the correct diagnosis of a TRS. This is likely the reason why so many different genes are found apparently mutated in small HDGC and FIGC families, or isolated early-onset GC cases, which likely represent branches of larger families with different TRS.

This review highlights the importance of collecting extended and updated family histories, to decide on the most adequate genetic testing to be offered, and the most adequate post-testing management. For FIGC, as a purer cause remains to be identified, it is important to consider these families for panel-testing including DNA damage repair genes, as these emerge as great candidates in this disease.

Author statement

José G. Pelaez, Rita Barbosa-Matos, and Celina S. José: Investigation and Methodology; **Carla Oliveira:** Supervision; Project administration; Funding acquisition; **José G. Pelaez, Rita Barbosa-Matos, Celina S. José and Carla Oliveira:** Conceptualization; Formal analysis; Writing - Original Draft; **All authors:** Writing – Review & Editing.

Funding

This research was funded by FEDER - Fundo Europeu de Desenvolvimento Regional funds through the COMPETE 2020-Operacional Programme for Competitiveness and Internationalisation (POCI), Portugal 2020, and by Portuguese funds through FCT/Ministério da Ciência, Tecnologia e Inovação in the framework of the projects Ref. PTDC/BTM-TEC/30164/2017 and Ref. PTDC/BTM-TEC/6706/2020. the European Union's Horizon 2020 research and innovation programme under grant agreement No. 779257 (Solve-RD: Ref. H2020-SC1-2017-Single-Stage-RTD) providing salary support through a Ph.D studentship to José G. Pelaez; This work was further supported by the European Union's Horizon 2020 research and innovation programme under grant agreement No. 779257 (Solve-RD: Ref. H2020-SC1-2017-Single-Stage-RTD) providing a Ph.D studentship to José G. Pelaez; and by Fundação para a Ciência e Tecnologia (FCT) with support for the PhD studentships to Rita Barbosa-Matos (Ref. SFRH/BD/145132/2019) and Celina S. José (Ref. SFRH/BD/140796/2018).

Declaration of competing interest

The authors disclosure no conflicts of interest.

Acknowledgments

This research is supported by the European Reference Network on Genetic Tumour Risk Syndromes (ERN GENTURIS) – Project ID No. 739547. ERN GENTURIS is partly co-funded by the European Union within the framework of the Third Health Programme "ERN-2016 – Framework Partnership Agreement 2017–2021". University of Porto supported this work through the Doctoral Pprogramme in Biomedicine of the Faculty of Medicine to José G. Pelaez and Celina S. José, and through the International Doctoral Programme in Molecular and Cellular Biotechnology Applied to Health Sciences from Institute of Biomedical Sciences Abel Salazar to Rita Barbosa-Matos.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.ejmg.2021.104401.

References

Aberle, H., Schwartz, H., Kemler, R., 1996. Cadherin-catenin complex: protein interactions and their implications for cadherin function. J. Cell. Biochem. 61, 514–523.

Antoniou, A.C., Casadei, S., Heikkinen, T., Barrowdale, D., Pylkäs, K., Roberts, J., Lee, A., Subramanian, D., De Leeneer, K., Fostira, F., Tomiak, E., Neuhausen, S.L., Teo, Z.L., Khan, S., Aittomäki, K., Moilanen, J.S., Turnbull, C., Seal, S., Mannermaa, A., Kallioniemi, A., Lindeman, G.J., Buys, S.S., Andrulis, I.L., Radice, P., Tondini, C., Manoukian, S., Toland, A.E., Miron, P., Weitzel, J.N., Domchek, S.M., Poppe, B., Claes, K.B., Yannoukakos, D., Concannon, P., Bernstein, J.L., James, P.A., Easton, D. F., Goldgar, D.E., Hopper, J.L., Rahman, N., Peterlongo, P., Nevanlinna, H., King, M. C., Couch, F.J., Southey, M.C., Winqvist, R., Foulkes, W.D., Tischkowitz, M., 2014. Breast-cancer risk in families with mutations in PALB2. N. Engl. J. Med. 371, 497–506.

Barber, M., Murrell, A., Ito, Y., Maia, A.T., Hyland, S., Oliveira, C., Save, V., Carneiro, F., Paterson, A.L., Grehan, N., Dwerryhouse, S., Lao-Sirieix, P., Caldas, C., Fitzgerald, R. C., 2008. Mechanisms and sequelae of E-cadherin silencing in hereditary diffuse gastric cancer. J. Pathol. 216, 295–306.

Barbosa-Matos, R., Leal Silva, R., Garrido, L., Aguiar, A.C., Garcia-Pelaez, J., André, A., Seixas, S., Sousa, S.P., Ferro, L., Vilarinho, L., Gullo, I., Devezas, V., Oliveira, R., Fernandes, S., Costa, S.C., Magalhães, A., Baptista, M., Carneiro, F., Pinheiro, H., Castedo, S., Oliveira, C., 2021. The CDH1 c.1901C>T variant: a founder variant in the Portuguese population with severe impact in mRNA splicing. Cancers 13.

Beà, S., Valdés-Mas, R., Navarro, A., Salaverria, I., Martín-Garcia, D., Jares, P., Giné, E., Pinyol, M., Royo, C., Nadeu, F., Conde, L., Juan, M., Clot, G., Vizán, P., Di Croce, L., Puente, D.A., López-Guerra, M., Moros, A., Roue, G., Aymerich, M., Villamor, N., Colomo, L., Martínez, A., Valera, A., Martín-Subero, J.I., Amador, V., Hernández, L., Rozman, M., Enjuanes, A., Forcada, P., Muntañola, A., Hartmann, E.M., Calasanz, M. J., Rosenwald, A., Ott, G., Hernández-Rivas, J.M., Klapper, W., Siebert, R., Wiestner, A., Wilson, W.H., Colomer, D., López-Guillermo, A., López-Otín, C., Puente, X.S., Campo, E., 2013. Landscape of somatic mutations and clonal evolution in mantle cell lymphoma. Proc. Natl. Acad. Sci. U. S. A. 110, 18250–18255.

Beer, A., Streubel, B., Asari, R., Dejaco, C., Oberhuber, G., 2017. Gastric adenocarcinoma and proximal polyposis of the stomach (GAPPS) – a rare recently described gastric polyposis syndrome – report of a case. Z. Gastroenterol. 55, 1131–1134.

Benusiglio, P.R., Colas, C., Guillerm, E., Canard, A., Delhomelle, H., Warcoin, M., Bellanger, J., Eyries, M., Zizi, M., Netter, J., Soubrier, F., Parc, Y., Mourregot, A., Maran Gonzalez, A., Cusin, V., Denis, J.A., Coupier, I., Svrcek, M., Coulet, F., 2019. Clinical implications of CTNNA1 germline mutations in asymptomatic carriers. Gastric Cancer 22, 899–903.

Berx, G., Cleton-Jansen, A.M., Nollet, F., de Leeuw, W.J., van de Vijver, M., Cornelisse, C., van Roy, F., 1995. E-cadherin is a tumour/invasion suppressor gene mutated in human lobular breast cancers. EMBO J. 14, 6107–6115.

Blair, V.R., McLeod, M., Carneiro, F., Coit, D.G., D'Addario, J.L., van Dieren, J.M., Harris, K.L., Hoogerbrugge, N., Oliveira, C., van der Post, R.S., Arnold, J., Benusiglio, P.R., Bisseling, T.M., Boussioutas, A., Cats, A., Charlton, A., Schreiber, K. E.C., Davis, J.L., Pietro, M.d., Fitzgerald, R.C., Ford, J.M., Gamet, K., Gullo, I., Hardwick, R.H., Huntsman, D.G., Kaurah, P., Kupfer, S.S., Latchford, A., Mansfield, P.F., Nakajima, T., Parry, S., Rossaak, J., Sugimura, H., Svrcek, M., Tischkowitz, M., Ushijima, T., Yamada, H., Yang, H.-K., Claydon, A., Figueiredo, J., Paringatai, K., Seruca, R., Bougen-Zhukov, N., Brew, T., Busija, S., Carneiro, P., DeGregorio, L., Fisher, H., Gardner, E., Godwin, T.D., Holm, K.N., Humar, B., Lintott, C.J., Monroe, E.C., Muller, M.D., Norero, E., Nouri, Y., Paredes, J., Sanches, J.M., Schulpen, E., Ribeiro, A.S., Sporle, A., Whitworth, J., Zhang, L., Reeve, A.E., Guilford, P., 2020. Hereditary diffuse gastric cancer: updated clinical practice guidelines. Lancet Oncol. 21, e386–e397.

Cabrera, E., Blanco, I., Yagüe, C., Zabalegui, A., 2010. The impact of genetic counseling on knowledge and emotional responses in Spanish population with family history of breast cancer. Patient Educ. Counsel. 78, 382–388.

Caldas, C., Carneiro, F., Lynch, H.T., Yokota, J., Wiesner, G.L., Powell, S.M., Lewis, F.R., Huntsman, D.G., Pharoah, P.D., Jankowski, J.A., MacLeod, P., Vogelsang, H., Keller, G., Park, K.G., Richards, F.M., Maher, E.R., Gayther, S.A., Oliveira, C., Grehan, N., Wight, D., Seruca, R., Roviello, F., Ponder, B.A., Jackson, C.E., 1999. Familial gastric cancer: overview and guidelines for management. J. Med. Genet. 36, 873–880.

Capelle, L.G., Van Grieken, N.C., Lingsma, H.F., Steyerberg, E.W., Klokman, W.J., Bruno, M.J., Vasen, H.F., Kuipers, E.J., 2010. Risk and epidemiological time trends of gastric cancer in Lynch syndrome carriers in The Netherlands. Gastroenterology 138, 487–492.

Carvalho, J., Oliveira, P., Senz, J., São José, C., Hansford, S., Teles, S.P., Ferreira, M., Corso, G., Pinheiro, H., Lemos, D., Pascale, V., Roviello, F., Huntsman, D., Oliveira, C., 2021. Redefinition of familial intestinal gastric cancer: clinical and genetic perspectives. J. Med. Genet. 58, 1–11.

Chaudhry, M.S., Gilmour, K.C., House, I.G., Layton, M., Panoskaltsis, N., Sohal, M., Trapani, J.A., Voskoboinik, I., 2016. Missense mutations in the perforin (PRF1) gene as a cause of hereditary cancer predisposition. OncoImmunology 5, e1179415.

Choi, Y.J., Kim, N., 2016. Gastric cancer and family history. Korean J. Intern. Med. (Engl. Ed.) 31, 1042–1053.

Choi, M., Kipps, T., Kurzrock, R., 2016. ATM mutations in cancer: therapeutic implications. Mol. Cancer Therapeut. 15, 1781–1791.

Clark, D.F., Michalski, S.T., Tondon, R., Nehoray, B., Ebrahimzadeh, J., Hughes, S.K., Soper, E.R., Domchek, S.M., Rustgi, A.K., Pineda-Alvarez, D., Anderson, M.J., Katona, B.W., 2020. Loss-of-function variants in CTNNA1 detected on multigene panel testing in individuals with gastric or breast cancer. Genet. Med. 22, 840–846.

Cooper, D.N., Krawczak, M., Polychronakos, C., Tyler-Smith, C., Kehrer-Sawatzki, H., 2013. Where genotype is not predictive of phenotype: towards an understanding of the molecular basis of reduced penetrance in human inherited disease. Hum. Genet. 132, 1077–1130.

Corso, G., Roncalli, F., Marrelli, D., Carneiro, F., Roviello, F., 2013. History, pathogenesis, and management of familial gastric cancer: original study of John XXIII's family. BioMed Res. Int. 2013, 385132.

Corso, G., Figueiredo, J., La Vecchia, C., Veronesi, P., Pravettoni, G., Macis, D., Karam, R., Lo Gullo, R., Provenzano, E., Toesca, A., Mazzocco, K., Carneiro, F., Seruca, R., Melo, S., Schmitt, F., Roviello, F., De Scalzi, A.M., Intra, M., Feroce, I., De Camilli, E., Villardita, M.G., Trentin, C., De Lorenzi, F., Bonanni, B., Galimberti, V., 2018. Hereditary lobular breast cancer with an emphasis on E-cadherin genetic defect. J. Med. Genet. 55, 431–441.

Daly, M.B., Pal, T., Berry, M.P., Buys, S.S., Dickson, P., Domchek, S.M., Elkhanany, A., Friedman, S., Goggins, M., Hutton, M.L., Karlan, B.Y., Khan, S., Klein, C., Kohlmann, W., Kurian, A.W., Laronga, C., Litton, J.K., Mak, J.S., Menendez, C.S., Merajver, S.D., Norquist, B.S., Offit, K., Pederson, H.J., Reiser, G., Senter-Jamieson, L., Shannon, K.M., Shatsky, R., Visvanathan, K., Weitzel, J.N., Wick, M.J., Wisinski, K.B., Yurgelun, M.B., Darlow, S.D., Dwyer, M.A., 2021. Genetic/familial high-risk assessment: breast, ovarian, and pancreatic, version 2.2021, NCCN clinical practice guidelines in oncology. J. Natl. Compr. Cancer Netw. 19, 77–102.

de Boer, W.B., Ee, H., Kumarasinghe, M.P., 2018. Neoplastic lesions of gastric adenocarcinoma and proximal polyposis syndrome (GAPPS) are gastric phenotype. Am. J. Surg. Pathol. 42, 1–8.

Declich, P., Arrigoni, G.S., Omazzi, B., Devani, M., Tavani, E., Sioli, A., Bellone, S., Gozzini, C., Bortoli, A., Prada, A., 2013. Sporadic fundic gland polyps and proximal polyposis associated with gastric adenocarcinoma share a common antral G cell hyperplasia. Gut 62, 1088–1089.

Donner, I., Kiviluoto, T., Ristimäki, A., Aaltonen, L.A., Vahteristo, P., 2015. Exome sequencing reveals three novel candidate predisposition genes for diffuse gastric cancer. Fam. Cancer 14, 241–246.

Drouin, A., Letarte, G., Raymond, F., Marchand, M., Corbeil, J., Laviolette, F., 2019. Interpretable genotype-to-phenotype classifiers with performance guarantees. Sci. Rep. 9, 4071.

El Abed, R., Bourdon, V., Voskoboinik, I., Omri, H., Youssef, Y.B., Laatiri, M.A., Huiart, L., Eisinger, F., Rabayrol, L., Frenay, M., Gesta, P., Demange, L., Dreyfus, H., Bonadona, V., Dugast, C., Zattara, H., Faivre, L., Zaier, M., Jemni, S.Y., Noguchi, T., Sobol, H., Soua, Z., 2011. Molecular study of the perforin gene in familial hematological malignancies. Hered. Cancer Clin. Pract. 9, 9.

Ellison-Loschmann, L., Sporle, A., Corbin, M., Cheng, S., Harawira, P., Gray, M., Whaanga, T., Guilford, P., Koea, J., Pearce, N., 2017. Risk of stomach cancer in Aotearoa/New Zealand: a Māori population based case-control study. PLoS One 12, e0181581.

Fewings, E., Larionov, A., Redman, J., Goldgraben, M.A., Scarth, J., Richardson, S., Brewer, C., Davidson, R., Ellis, I., Evans, D.G., Halliday, D., Izatt, L., Marks, P., McConnell, V., Verbist, L., Mayes, R., Clark, G.R., Hadfield, J., Chin, S.F., Teixeira, M.R., Giger, O.T., Hardwick, R., di Pietro, M., O'Donovan, M., Pharoah, P., Caldas, C., Fitzgerald, R.C., Tischkowitz, M., 2018. Germline pathogenic variants in PALB2 and other cancer-predisposing genes in families with hereditary diffuse gastric cancer without CDH1 mutation: a whole-exome sequencing study. Lancet Gastroenterol Hepatol 3, 489–498.

Figueiredo, J., Melo, S., Carneiro, P., Moreira, A.M., Fernandes, M.S., Ribeiro, A.S., Guilford, P., Paredes, J., Seruca, R., 2019. Clinical spectrum and pleiotropic nature of CDH1 germline mutations. J. Med. Genet. 56, 199–208.

Figueiredo, J., Mercadillo, F., Melo, S., Barroso, A., Gonçalves, M., Díaz-Tasende, J., Carneiro, P., Robles, L., Colina, F., Ibarrola, C., Perea, J., Morais-de-Sá, E., Seruca, R., Urioste, M., 2021. Germline CDH1 G212E missense variant: combining clinical. In: Vitro and in Vivo Strategies to Unravel Disease Burden. Cancers (Basel). vol. 13.

Fitzgerald, R.C., Hardwick, R., Huntsman, D., Ca... Chung, D.C., Norton, J., Ragunath, K., Van K... Caldas, C., 2010. Hereditary diffuse gastric ca... clinical management and directions for future research. J. Med. Genet. 47, 436–444.

http://refhub.elsevier.com/S1769-7212(21)00267-6

Foretová, L., Navrátilová, M., Svoboda, M., Grell, P., Nemec, L., Sirotek, L., Obermannová, R., Novotný, I., Sachlova, M., Fabian, P., Kroupa, R., Vasickova, P., Házová, J., Šťahlová, E.H., Machackova, E., 2019. GAPPS - gastric adenocarcinoma and proximal polyposis of the stomach syndrome in 8 families tested at masaryk memorial cancer Institute - prevention and prophylactic gastrectomies. Klin. Onkol. 32, 109–117.

Frebourg, T., Oliveira, C., Hochain, P., Karam, R., Manouvrier, S., Graziadio, C., Vekemans, M., Hartmann, A., Baert-Desurmont, S., Alexandre, C., Lejeune Dumoulin, S., Marroni, C., Martin, C., Castedo, S., Lovett, M., Winston, J., Machado, J.C., Attié, T., Jabs, E.W., Cai, J., Pellerin, P., Triboulet, J.P., Scotte, M., Le Pessot, F., Hedouin, A., Carneiro, F., Blayau, M., Seruca, R., 2006. Cleft lip/palate and CDH1/E-cadherin mutations in families with hereditary diffuse gastric cancer. J. Med. Genet. 43, 138–142.

Funato, M., Tsunematsu, Y., Yamazaki, F., Tamura, C., Kumamoto, T., Takagi, M., Kato, S., Sugimura, H., Tamura, K., 2021. Characteristics of Li-Fraumeni syndrome in Japan: a review study by the special committee of JSHT. Cancer Sci. 112, 2821–2834.

Garcia-Pelaez, J., Barbosa-Matos, R., Gullo, I., Carneiro, F., Oliveira, C., 2021. Histological and Mutational Profile of Diffuse Gastric Cancer: Current Knowledge and Future Challenges. Molecular Oncology 15, 2841–2867.

Gaston, D., Hansford, S., Oliveira, C., Nightingale, M., Pinheiro, H., Macgillivray, C., Kaurah, P., Rideout, A.L., Steele, P., Soares, G., Huang, W.Y., Whitehouse, S., Blowers, S., LeBlanc, M.A., Jiang, H., Greer, W., Samuels, M.E., Orr, A., Fernandez, C.V., Majewski, J., Ludman, M., Dyack, S., Penney, L.S., McMaster, C.R., Huntsman, D., Bedard, K., 2014. Germline mutations in MAP3K6 are associated with familial gastric cancer. PLoS Genet. 10, e1004669.

Grady, W.M., Willis, J., Guilford, P.J., Dunbier, A.K., Toro, T.T., Lynch, H., Wiesner, G., Ferguson, K., Eng, C., Park, J.G., Kim, S.J., Markowitz, S., 2000. Methylation of the CDH1 promoter as the second genetic hit in hereditary diffuse gastric cancer. Nat. Genet. 26, 16–17.

Guilford, P., Hopkins, J., Harraway, J., McLeod, M., McLeod, N., Harawira, P., Taite, H., Scoular, R., Miller, A., Reeve, A.E., 1998. E-cadherin germline mutations in familial gastric cancer. Nature 392, 402–405.

Gullo, I., Devezas, V., Baptista, M., Garrido, L., Castedo, S., Morais, R., Wen, X., Rios, E., Pinheiro, J., Pinto-Ribeiro, I., Ferreira, R.M., Preto, J., Santos-Antunes, J., Marques, M., Campos, M., Almeida, F., Espinheira, M.D.C., Amil Dias, J., Figueiredo, C., Oliveira, C., Trindade, E., Carneiro, F., 2018. Phenotypic heterogeneity of hereditary diffuse gastric cancer: report of a family with early-onset disease. Gastrointest. Endosc. 87, 1566–1575.

Gullo, I., van der Post, R.S., Carneiro, F., 2021. Recent advances in the pathology of heritable gastric cancer syndromes. Histopathology 78, 125–147.

Hansford, S., Kaurah, P., Li-Chang, H., Woo, M., Senz, J., Pinheiro, H., Schrader, K.A., Schaeffer, D.F., Shumansky, K., Zogopoulos, G., Santos, T.A., Claro, I., Carvalho, J., Nielsen, C., Padilla, S., Lum, A., Talhouk, A., Baker-Lange, K., Richardson, S., Lewis, I., Lindor, N.M., Pennell, E., MacMillan, A., Fernandez, B., Keller, G., Lynch, H., Shah, S.P., Guilford, P., Gallinger, S., Corso, G., Roviello, F., Caldas, C., Oliveira, C., Pharoah, P.D., Huntsman, D.G., 2015. Hereditary diffuse gastric cancer syndrome: CDH1 mutations and beyond. JAMA Oncol. 1, 23–32.

Harris, T.J., Tepass, U., 2010. Adherens junctions: from molecules to morphogenesis. Nat. Rev. Mol. Cell Biol. 11, 502–514.

Hegde, M., Ferber, M., Mao, R., Samowitz, W., Ganguly, A., 2014. ACMG technical standards and guidelines for genetic testing for inherited colorectal cancer (Lynch syndrome, familial adenomatous polyposis, and MYH-associated polyposis). Genet. Med. 16, 101–116.

Herrera-Pariente, C., Capó-García, R., Díaz-Gay, M., Carballal, S., Muñoz, J., Llach, J., Sánchez, A., Bonjoch, L., Arnau-Collell, C., Soares de Lima, Y., Golubicki, M., Jung, G., Lozano, J.J., Castells, A., Balaguer, F., Bujanda, L., Castellví-Bel, S., Moreira, L., 2021. Identification of new genes involved in germline predisposition to early-onset gastric cancer. Int. J. Mol. Sci. 22.

Hosoya, K., Yamashita, S., Ando, T., Nakajima, T., Itoh, F., Ushijima, T., 2009. Adenomatous polyposis coli 1A is likely to be methylated as a passenger in human gastric carcinogenesis. Cancer Lett. 285, 182–189.

Huang, K.L., Mashl, R.J., Wu, Y., Ritter, D.I., Wang, J., Oh, C., Paczkowska, M., Reynolds, S., Wyczalkowski, M.A., Oak, N., Scott, A.D., Krassowski, M., Cherniack, A.D., Houlahan, K.E., Jayasinghe, R., Wang, L.B., Zhou, D.C., Liu, D., Cao, S., Kim, Y.W., Koire, A., McMichael, J.F., Hucthagowder, V., Kim, T.B., Hahn, A., Wang, C., McLellan, M.D., Al-Mulla, F., Johnson, K.J., Lichtarge, O., Boutros, P.C., Raphael, B., Lazar, A.J., Zhang, W., Wendl, M.C., Govindan, R., Jain, S., Wheeler, D., Kulkarni, S., Dipersio, J.F., Reimand, J., Meric-Bernstam, F., Chen, K., Shmulevich, I., Plon, S.E., Chen, F., Ding, L., 2018. Pathogenic germline variants in 10,389 adult cancers. Cell 173, 355–370 e314.

Humar, B., Blair, V., Charlton, A., More, H., Martin, I., Guilford, P., 2009. E-cadherin deficiency initiates gastric signet-ring cell carcinoma in mice and man. Cancer Res. 69, 2050–2056.

Ichikawa, H., Wakai, T., Nagahashi, M., Shimada, Y., Hanyu, T., Kano, Y., Muneoka, Y., Ishikawa, T., Takizawa, K., Tajima, Y., Sakata, J., Kobayashi, T., Kemeyama, H., Yabusaki, H., Nakagawa, S., Sato, N., Kawasaki, T., Homma, K., Okuda, S., Lyle, S., Takabe, K., 2018. Pathogenic germline BRCA1/2 mutations and familial predisposition to gastric cancer. JCO Precis Oncol. 2.

Ikenoue, T., Arai, M., Ishioka, C., Iwama, T., Kaneko, S., Matsubara, N., Moriya, Y., Nomizu, T., Sugano, K., Tamura, K., Tomita, N., Yoshida, T., Sugihara, K., Naruse, H., Yamaguchi, K., Nojima, M., Nakamura, Y., Furukawa, Y., 2019. Importance of gastric cancer for the diagnosis and surveillance of Japanese Lynch syndrome patients. J. Hum. Genet. 64, 1187–1194.

Ilyas, M., Tomlinson, I.P., 1997. The interactions of APC, E-cadherin and beta-catenin in tumour development and progression. J. Pathol. 182, 128–137.

Jakubowska, A., Nej, K., Huzarski, T., Scott, R.J., Lubiński, J., 2002. BRCA2 gene mutations in families with aggregations of breast and stomach cancers. Br. J. Cancer 87, 888–891.

Jasperson, K.W., Patel, S.G., Ahnen, D.J., 1993. APC-associated polyposis conditions. In: Adam, M.P., Ardinger, H.H., Pagon, R.A., Wallace, S.E., Bean, L.J.H., Mirzaa, G., Amemiya, A. (Eds.), GeneReviews(®) Copyright © 1993-2021, University of Washington, Seattle. GeneReviews is a registered trademark of the University of Washington, Seattle. All rights reserved., Seattle (WA). University of Washington, Seattle.

Jezela-Stanek, A., Ciara, E., Jurkiewicz, D., Kucharczyk, M., Jędrzejowska, M., Chrzanowska, K.H., Krajewska-Walasek, M., Żemojtel, T., 2020. The phenotype-driven computational analysis yields clinical diagnosis for patients with atypical manifestations of known intellectual disability syndromes. Mol Genet Genomic Med. 8, e1263.

Jin, M.H., Oh, D.Y., 2019. ATM in DNA repair in cancer. Pharmacol. Ther. 203, 107391.

Johannsson, O., Loman, N., Möller, T., Kristoffersson, U., Borg, A., Olsson, H., 1999. Incidence of malignant tumours in relatives of BRCA1 and BRCA2 germline mutation carriers. Eur. J. Cancer 35, 1248–1257.

Kadiyska, T.K., Todorov, T.P., Bichev, S.N., Vazharova, R.V., Nossikoff, A.V., Savov, A.S., Mitev, V.I., 2014. APC promoter 1B deletion in familial polyposis–implications for mutation-negative families. Clin. Genet. 85, 452–457.

Kanemitsu, K., Iwatsuki, M., Yamashita, K., Komohara, Y., Morinaga, T., Iwagami, S., Eto, K., Nagai, Y., Kurashige, J., Baba, Y., Yoshida, N., Baba, H., 2021. Two Asian families with gastric adenocarcinoma and proximal polyposis of the stomach successfully treated via laparoscopic total gastrectomy. Clin. J. Gastroenterol. 14, 92–97.

Karam, R., Carvalho, J., Bruno, I., Graziadio, C., Senz, J., Huntsman, D., Carneiro, F., Seruca, R., Wilkinson, M.F., Oliveira, C., 2008. The NMD mRNA surveillance pathway downregulates aberrant E-cadherin transcripts in gastric cancer cells and in CDH1 mutation carriers. Oncogene 27, 4255–4260.

Kaurah, P., MacMillan, A., Boyd, N., Senz, J., De Luca, A., Chun, N., Suriano, G., Zaor, S., Van Manen, L., Gilpin, C., Nikkel, S., Connolly-Wilson, M., Weissman, S., Rubinstein, W.S., Sebold, C., Greenstein, R., Stroop, J., Yim, D., Panzini, B., McKinnon, W., Greenblatt, M., Wirtzfeld, D., Fontaine, D., Coit, D., Yoon, S., Chung, D., Lauwers, G., Pizzuti, A., Vaccaro, C., Redal, M.A., Oliveira, C., Tischkowitz, M., Olschwang, S., Gallinger, S., Lynch, H., Green, J., Ford, J., Pharoah, P., Fernandez, B., Huntsman, D., 2007. Founder and recurrent CDH1 mutations in families with hereditary diffuse gastric cancer. Jama 297, 2360–2372.

Keller, G., Vogelsang, H., Becker, I., Plaschke, S., Ott, K., Suriano, G., Mateus, A.R., Seruca, R., Biedermann, K., Huntsman, D., Döring, C., Holinski-Feder, E., Neutzling, A., Siewert, J.R., Höfler, H., 2004. Germline mutations of the E-cadherin (CDH1) and TP53 genes, rather than of RUNX3 and HPP1, contribute to genetic predisposition in German gastric cancer patients. J. Med. Genet. 41, e89.

Kim, I.J., Park, J.H., Kang, H.C., Shin, Y., Lim, S.B., Ku, J.L., Yang, H.K., Lee, K.U., Park, J.G., 2003. A novel germline mutation in the MET extracellular domain in a Korean patient with the diffuse type of familial gastric cancer. J. Med. Genet. 40, e97.

Kim, I.J., Kang, H.C., Shin, Y., Park, H.W., Jang, S.G., Han, S.Y., Lim, S.K., Lee, M.R., Chang, H.J., Ku, J.L., Yang, H.K., Park, J.G., 2004. A TP53-truncating germline mutation (E287X) in a family with characteristics of both hereditary diffuse gastric cancer and Li-Fraumeni syndrome. J. Hum. Genet. 49, 591–595.

Kuchenbaecker, K.B., McGuffog, L., Barrowdale, D., Lee, A., Soucy, P., Dennis, J., Domchek, S.M., Robson, M., Spurdle, A.B., Ramus, S.J., Mavaddat, N., Terry, M.B., Neuhausen, S.L., Schmutzler, R.K., Simard, J., Pharoah, P.D.P., Offit, K., Couch, F.J., Chenevix-Trench, G., Easton, D.F., Antoniou, A.C., 2017. Evaluation of polygenic risk scores for breast and ovarian cancer risk prediction in BRCA1 and BRCA2 mutation carriers. J. Natl. Cancer Inst. 109.

Kunovsky, L., Kala, Z., Potrusil, M., Novotny, I., Kubes, V., Prochazka, V., 2019. A Central European family with gastric adenocarcinoma and proximal polyposis of the stomach. Gastrointest. Endosc. 90, 523–525.

Lagarde, A., Rouleau, E., Ferrari, A., Noguchi, T., Qiu, J., Briaux, A., Bourdon, V., Rémy, V., Gaildrat, P., Adélaïde, J., Birnbaum, D., Lidereau, R., Sobol, H., Olschwang, S., 2010. Germline APC mutation spectrum derived from 863 genomic variations identified through a 15-year medical genetics service to French patients with FAP. J. Med. Genet. 47, 721–722.

Landrum, M.J., Lee, J.M., Riley, G.R., Jang, W., Rubinstein, W.S., Church, D.M., Maglott, D.R., 2014. ClinVar: public archive of relationships among sequence variation and human phenotype. Nucleic Acids Res. 42, D980–D985.

Lauren, P., 1965. The two histological main types OF gastric carcinoma: diffuse and SO-called intestinal-type carcinoma. AN attempt at a histo-clinical classification. Acta Pathol. Microbiol. Scand. 64, 31–49.

Law, R.H., Lukoyanova, N., Voskoboinik, I., Caradoc-Davies, T.T., Baran, K., Dunstone, M.A., D'Angelo, M.E., Orlova, E.V., Coulibaly, F., Verschoor, S., Browne, K.A., Ciccone, A., Kuiper, M.J., Bird, P.I., Trapani, J.A., Saibil, H.R., Whisstock, J.C., 2010. The structural basis for membrane binding and pore formation by lymphocyte perforin. Nature 468, 447–451.

Lee, J.H., Han, S.U., Cho, H., Jennings, B., Gerrard, B., Dean, M., Schmidt, L., Zbar, B., Vande Woude, G.F., 2000. A novel germ line juxtamembrane Met mutation in human gastric cancer. Oncogene 19, 4947–4953.

Lee, K., Krempely, K., Roberts, M.E., Anderson, M.J., Carneiro, F., Chao, E., Dixon, K., Figueiredo, J., Ghosh, R., Huntsman, D., Kaurah, P., Kesserwan, C., Landrith, T., Li, S., Mensenkamp, A.R., Oliveira, C., Pardo, C., Pesaran, T., Richardson, M., Slavin, T.P., Spurdle, A.B., Trapp, M., Witkowski, L., Yi, C.S., Zhang, L., Plon, S.E., Schrader, K.A., Karam, R., 2018. Specifications of the ACMG/AMP variant curation guidelines for the analysis of germline CDH1 sequence variants. Hum. Mutat. 39, 1553–1568.

Leoz, M.L., Sánchez, A., Carballal, S., Ruano, L., Ocaña, T., Pellisé, M., Castells, A., Balaguer, F., Moreira, L., 2016. [Hereditary gastric and pancreatic cancer predisposition syndromes]. Gastroenterol. Hepatol. 39, 481–493.
Li, J., Woods, S.L., Healey, S., Beesley, J., Chen, X., Lee, J.S., Sivakumaran, H., Wayte, N., Nones, K., Waterfall, J.J., Pearson, J., Patch, A.M., Senz, J., Ferreira, M.A., Kaurah, P., Mackenzie, R., Heravi-Moussavi, A., Hansford, S., Lannagan, T.R.M., Spurdle, A.B., Simpson, P.T., da Silva, L., Lakhani, S.R., Clouston, A.D., Bettington, M., Grimpen, F., Busuttil, R.A., Di Costanzo, N., Boussioutas, A., Jeanjean, M., Chong, G., Fabre, A., Olschwang, S., Faulkner, G.J., Bellos, E., Coin, L., Rioux, K., Bathe, O.F., Wen, X., Martin, H.C., Neklason, D.W., Davis, S.R., Walker, R. L., Calzone, K.A., Avital, I., Heller, T., Koh, C., Pineda, M., Rudloff, U., Quezado, M., Pichurin, P.N., Hulick, P.J., Weissman, S.M., Newlin, A., Rubinstein, W.S., Sampson, J.E., Hamman, K., Goldgar, D., Poplawski, N., Phillips, K., Schofield, L., Armstrong, J., Kiraly-Borri, C., Suthers, G.K., Huntsman, D.G., Foulkes, W.D., Carneiro, F., Lindor, N.M., Edwards, S.L., French, J.D., Waddell, N., Meltzer, P.S., Worthley, D.L., Schrader, K.A., Chenevix-Trench, G., 2016. Point mutations in exon 1B of APC reveal gastric adenocarcinoma and proximal polyposis of the stomach as a familial adenomatous polyposis variant. Am. J. Hum. Genet. 98, 830–842.
Li, R., Zhang, H., Cao, Y., Liu, X., Chen, Y., Qi, Y., Wang, J., Yu, K., Lin, C., Liu, H., He, H., Li, H., Chen, L., Shen, Z., Qin, J., Zhang, W., Sun, Y., Xu, J., 2020. Lauren classification identifies distinct prognostic value and functional status of intratumoral CD8(+) T cells in gastric cancer. Cancer Immunol. Immunother. 69, 1327–1336.
Lin, J., Shi, J., Guo, H., Yang, X., Jiang, Y., Long, J., Bai, Y., Wang, D., Yang, X., Wan, X., Zhang, L., Pan, J., Hu, K., Guan, M., Huo, L., Sang, X., Wang, K., Zhao, H., 2019. Alterations in DNA damage repair genes in primary liver cancer. Clin. Cancer Res. 25, 4701–4711.
Lo, W., Zhu, B., Sabesan, A., Wu, H.-H., Powers, A., Sorber, R.A., Ravichandran, S., Chen, I., McDuffie, L.A., Quadri, H.S., Beane, J.D., Calzone, K., Miettinen, M.M., Hewitt, S.M., Koh, C., Heller, T., Wacholder, S., Rudloff, U., 2019. Associations of CDH1 germline variant location and cancer phenotype in families with hereditary diffuse gastric cancer (HDGC). J. Med. Genet. 56, 370–379.
Lobo, S., Benusiglio, P.R., Coulet, F., Boussemart, L., Golmard, L., Spier, I., Hüneburg, R., Aretz, S., Colas, C., Oliveira, C., 2021. Cancer predisposition and germline CTNNA1 variants. Eur. J. Med. Genet. 64, 104316.
Lowstuter, K., Espenschied, C.R., Sturgeon, D., Ricker, C., Karam, R., LaDuca, H., Culver, J.O., Dolinsky, J.S., Chao, E., Sturgeon, J., Speare, V., Ma, Y., Kingham, K., Melas, M., Idos, G.E., McDonnell, K.J., Gruber, S.B., 2017. Unexpected CDH1 mutations identified on multigene panels pose clinical management challenges. JCO Precision Oncology 1–12.
Lu, C., Xie, M., Wendl, M.C., Wang, J., McLellan, M.D., Leiserson, M.D., Huang, K.L., Wyczalkowski, M.A., Jayasinghe, R., Banerjee, T., Ning, J., Tripathi, P., Zhang, Q., Niu, B., Ye, K., Schmidt, H.K., Fulton, R.S., McMichael, J.F., Batra, P., Kandoth, C., Bharadwaj, M., Koboldt, D.C., Miller, C.A., Kanchi, K.L., Eldred, J.M., Larson, D.E., Welch, J.S., You, M., Ozenberger, B.A., Govindan, R., Walter, M.J., Ellis, M.J., Mardis, E.R., Graubert, T.A., Dipersio, J.F., Ley, T.J., Wilson, R.K., Goodfellow, P.J., Raphael, B.J., Chen, F., Johnson, K.J., Parvin, J.D., Ding, L., 2015. Patterns and functional implications of rare germline variants across 12 cancer types. Nat. Commun. 6, 10086.
Majewski, I.J., Kluijt, I., Cats, A., Scerri, T.S., de Jong, D., Kluin, R.J., Hansford, S., Hogervorst, F.B., Bosma, A.J., Hofland, I., Winter, M., Huntsman, D., Jonkers, J., Bahlo, M., Bernards, R., 2013. An α-E-catenin (CTNNA1) mutation in hereditary diffuse gastric cancer. J. Pathol. 229, 621–629.
Melo, S., Figueiredo, J., Fernandes, M.S., Gonçalves, M., Morais-de-Sá, E., Sanches, J.M., Seruca, R., 2017. Predicting the functional impact of CDH1 missense mutations in hereditary diffuse gastric cancer. Int. J. Mol. Sci. 18.
Mitsui, Y., Yokoyama, R., Fujimoto, S., Kagemoto, K., Kitamura, S., Okamoto, K., Muguruma, N., Bando, Y., Eguchi, H., Okazaki, Y., Ishida, H., Takayama, T., 2018. First report of an Asian family with gastric adenocarcinoma and proximal polyposis of the stomach (GAPPS) revealed with the germline mutation of the APC exon 1B promoter region. Gastric Cancer 21, 1058–1063.
National Comprehensive Cancer Network, 2021a. In: Network, N.C.C. (Ed.), NCCN Clinical Practice Guidelines in Oncology (NCCN Guidelines). Genetic/Familial High-Risk Assessment: Breast, Ovarian, and Pancreatic. Version 2.2021.
National Comprehensive Cancer Network, 2021b. In: Network, N.C.C. (Ed.), NCCN Clinical Practice Guidelines in Oncology (NCCN Guidelines). Genetic/Familial High-Risk Assessment: Colorectal. Version 1.2020.
Navrkalova, V., Sebejova, L., Zemanova, J., Kminkova, J., Kubesova, B., Malcikova, J., Mraz, M., Smardova, J., Pavlova, S., Doubek, M., Brychtova, Y., Potesil, D., Nemethova, V., Mayer, J., Pospisilova, S., Trbusek, M., 2013. ATM mutations uniformly lead to ATM dysfunction in chronic lymphocytic leukemia: application of functional test using doxorubicin. Haematologica 98, 1124–1131.
Okamoto, R., Irie, K., Yamada, A., Katata, T., Fukuhara, A., Takai, Y., 2005. Recruitment of E-cadherin associated with alpha- and beta-catenins and p120ctn to the nectin-based cell-cell adhesion sites by the action of 12-O-tetradecanoylphorbol-13-acetate in MDCK cells. Gene Cell. 10, 435–445.
Oliveira, C., Bordin, M.C., Grehan, N., Huntsman, D., Suriano, G., Machado, J.C., Kiviluoto, T., Aaltonen, L., Jackson, C.E., Seruca, R., Caldas, C., 2002. Screening E-cadherin in gastric cancer families reveals germline mutations only in hereditary diffuse gastric cancer kindred. Hum. Mutat. 19, 510–517.
Oliveira, C., Senz, J., Kaurah, P., Pinheiro, H., Sanges, R., Haegert, A., Corso, G., Schouten, J., Fitzgerald, R., Vogelsang, H., Keller, G., Dwerryhouse, S., Grimmer, D., Chin, S.-F., Yang, H.-K., Jackson, C.E., Seruca, R., Roviello, F., Stupka, E., Caldas, C., Huntsman, D., 2009a. Germline CDH1 deletions in hereditary diffuse gastric cancer families. Hum. Mol. Genet. 18, 1545–1555.
Oliveira, C., Sousa, S., Pinheiro, H., Karam, R., Bordeira-Carriço, R., Senz, J., Kaurah, P., Carvalho, J., Pereira, R., Gusmão, L., Wen, X., Cipriano, M.A., Yokota, J., Carneiro, F., Huntsman, D., Seruca, R., 2009b. Quantification of epigenetic and genetic 2nd hits in CDH1 during hereditary diffuse gastric cancer syndrome progression. Gastroenterology 136, 2137–2148.
Oliveira, C., Pinheiro, H., Figueiredo, J., Seruca, R., Carneiro, F., 2015. Familial gastric cancer: genetic susceptibility, pathology, and implications for management. Lancet Oncol. 16, e60–70.
Pavicic, W., Nieminen, T.T., Gylling, A., Pursiheimo, J.P., Laiho, A., Gyenesei, A., Järvinen, H.J., Peltomäki, P., 2014. Promoter-specific alterations of APC are a rare cause for mutation-negative familial adenomatous polyposis. Genes Chromosomes Cancer 53, 857–864.
Pearlman, R., Frankel, W.L., Swanson, B., Zhao, W., Yilmaz, A., Miller, K., Bacher, J., Bigley, C., Nelsen, L., Goodfellow, P.J., Goldberg, R.M., Paskett, E., Shields, P.G., Freudenheim, J.L., Stanich, P.P., Lattimer, I., Arnold, M., Liyanarachchi, S., Kalady, M., Heald, B., Greenwood, C., Paquette, I., Prues, M., Draper, D.J., Lindeman, C., Kuebler, J.P., Reynolds, K., Brell, J.M., Shaper, A.A., Mahesh, S., Buie, N., Weeman, K., Shine, K., Haut, M., Edwards, J., Bastola, S., Wickham, K., Khanduja, K.S., Zacks, R., Pritchard, C.C., Shirts, B.H., Jacobson, A., Allen, B., de la Chapelle, A., Hampel, H., 2017. Prevalence and spectrum of germline cancer susceptibility gene mutations among patients with early-onset colorectal cancer. JAMA Oncol. 3, 464–471.
Pharoah, P.D., Guilford, P., Caldas, C., 2001. Incidence of gastric cancer and breast cancer in CDH1 (E-cadherin) mutation carriers from hereditary diffuse gastric cancer families. Gastroenterology 121, 1348–1353.
Pocurull, A., Herrera-Pariente, C., Carballal, S., Llach, J., Sánchez, A., Carot, L., Botargues, J.M., Cuatrecasas, M., Ocaña, T., Balaguer, F., Bujanda, L., Moreira, L., 2021. Clinical, molecular and genetic characteristics of early onset gastric cancer: analysis of a large multicenter study. Cancers 13, 3132.
Praper, T., Sonnen, A., Viero, G., Kladnik, A., Froelich, C.J., Anderluh, G., Dalla Serra, M., Gilbert, R.J., 2011. Human perforin employs different avenues to damage membranes. J. Biol. Chem. 286, 2946–2955.
Rahner, N., Steinke, V., 2008. Hereditary cancer syndromes. Dtsch Arztebl Int. 105, 706–714.
Repak, R., Kohoutova, D., Podhola, M., Rejchrt, S., Minarik, M., Benesova, L., Lesko, M., Bures, J., 2016. The first European family with gastric adenocarcinoma and proximal polyposis of the stomach: case report and review of the literature. Gastrointest. Endosc. 84, 718–725.
Richards, F.M., McKee, S.A., Rajpar, M.H., Cole, T.R., Evans, D.G., Jankowski, J.A., McKeown, C., Sanders, D.S., Maher, E.R., 1999. Germline E-cadherin gene (CDH1) mutations predispose to familial gastric cancer and colorectal cancer. Hum. Mol. Genet. 8, 607–610.
Richards, S., Aziz, N., Bale, S., Bick, D., Das, S., Gastier-Foster, J., Grody, W.W., Hegde, M., Lyon, E., Spector, E., Voelkerding, K., Rehm, H.L., 2015. Standards and guidelines for the interpretation of sequence variants: a joint consensus recommendation of the American College of medical genetics and Genomics and the association for molecular pathology. Genet. Med. 17, 405–424.
Riordan, J.D., Nadeau, J.H., 2017. From peas to disease: modifier genes, Network resilience, and the genetics of health. Am. J. Hum. Genet. 101, 177–191.
Roberts, M.E., Ranola, J.M.O., Marshall, M.L., Susswein, L.R., Graceffo, S., Bohnert, K., Tsai, G., Klein, R.T., Hruska, K.S., Shirts, B.H., 2019. Comparison of CDH1 penetrance estimates in clinically ascertained families vs families ascertained for multiple gastric cancers. JAMA Oncol. 5, 1325–1331.
Rohlin, A., Engwall, Y., Fritzell, K., Göransson, K., Bergsten, A., Einbeigi, Z., Nilbert, M., Karlsson, P., Björk, J., Nordling, M., 2011. Inactivation of promoter 1B of APC causes partial gene silencing: evidence for a significant role of the promoter in regulation and causative of familial adenomatous polyposis. Oncogene 30, 4977–4989.
Roy, R., Chun, J., Powell, S.N., 2011. BRCA1 and BRCA2: different roles in a common pathway of genome protection. Nat. Rev. Cancer 12, 68–78.
Rudloff, U., 2018. Gastric adenocarcinoma and proximal polyposis of the stomach: diagnosis and clinical perspectives. Clin. Exp. Gastroenterol. 11, 447–459.
Sahasrabudhe, R., Lott, P., Bohorquez, M., Toal, T., Estrada, A.P., Suarez, J.J., Brea-Fernández, A., Cameselle-Teijeiro, J., Pinto, C., Ramos, I., Mantilla, A., Prieto, R., Corvalan, A., Norero, E., Alvarez, C., Tapia, T., Carvallo, P., Gonzalez, L.M., , Cock-Rada, A., Solano, A., Neffa, F., Della Valle, A., Yau, C., Soares, G., Borowsky, A., Hu, N., He, L.J., Han, X.Y., Taylor, P.R., Goldstein, A.M., Torres, J., Echeverry, M., Ruiz-Ponte, C., Teixeira, M.R., Carvajal-Carmona, L.G., 2017. Germline mutations in PALB2, BRCA1, and RAD51C, which regulate DNA recombination repair, in patients with gastric cancer. Gastroenterology 152, 983–986 e986.
Savitsky, K., Bar-Shira, A., Gilad, S., Rotman, G., Ziv, Y., Vanagaite, L., Tagle, D.A., Smith, S., Uziel, T., Sfez, S., Ashkenazi, M., Pecker, I., Frydman, M., Harnik, R., Patanjali, S.R., Simmons, A., Clines, G.A., Sartiel, A., Gatti, R.A., Chessa, L., Sanal, O., Lavin, M.F., Jaspers, N.G., Taylor, A.M., Arlett, C.F., Miki, T., Weissman, S. M., Lovett, M., Collins, F.S., Shiloh, Y., 1995. A single ataxia telangiectasia gene with a product similar to PI-3 kinase. Science 268, 1749–1753.
Selvanathan, A., Nixon, C.Y., Zhu, Y., Scietti, L., Forneris, F., Uribe, L.M.M., Lidral, A.C., Jezewski, P.A., Mulliken, J.B., Murray, J.C., Buckley, M.F., Cox, T.C., Roscioli, T., 2020. CDH1 mutation distribution and type suggests genetic differences between the etiology of orofacial clefting and gastric cancer. Genes 11.
Sereno, M., Aguayo, C., Ponce, C.G., Gómez-Raposo, C., Zambrana, F., Gómez-López, M., Casado, E., 2011. Gastric tumours in hereditary cancer syndromes: clinical features, molecular biology and strategies for prevention. Clin. Transl. Oncol. 13, 599.
Setia, N., Clark, J.W., Duda, D.G., Hong, T.S., Kwak, E.L., Mullen, J.T., Lauwers, G.Y., 2015. Familial gastric cancers. Oncol. 20, 1365–1377.

Setia, N., Wang, C.X., Lager, A., Maron, S., Shroff, S., Arndt, N., Peterson, B., Kupfer, S.S., Ma, C., Misdraji, J., Catenacci, D., Hart, J., 2020. Morphologic and molecular analysis of early-onset gastric cancer. Cancer 127, 103–114.
Shapiro, L., Fannon, A.M., Kwong, P.D., Thompson, A., Lehmann, M.S., Grübel, G., Legrand, J.F., Als-Nielsen, J., Colman, D.R., Hendrickson, W.A., 1995. Structural basis of cell-cell adhesion by cadherins. Nature 374, 327–337.
Slavin, T., Neuhausen, S.L., Rybak, C., Solomon, I., Nehoray, B., Blazer, K., Niell-Swiller, M., Adamson, A.W., Yuan, Y.C., Yang, K., Sand, S., Castillo, D., Herzog, J., Wu, X., Tao, S., Chavez, T., Woo, Y., Chao, J., Mora, P., Horcasitas, D., Weitzel, J., 2017. Genetic gastric cancer susceptibility in the international clinical cancer Genomics community research Network. Cancer Genet. 216–217, 111–119.
Snow, A.K., Tuohy, T.M., Sargent, N.R., Smith, L.J., Burt, R.W., Neklason, D.W., 2015. APC promoter 1B deletion in seven American families with familial adenomatous polyposis. Clin. Genet. 88, 360–365.
Sokoloff, B., 1938. Predisposition to cancer in the Bonaparte family. Am. J. Surg. 40, 673–678.
Stuebs, F., Heidemann, S., Caliebe, A., Mundhenke, C., Arnold, N., 2018. CDH1 mutation screen in a BRCA1/2-negative familial breast-/ovarian cancer cohort. Arch. Gynecol. Obstet. 297, 147–152.
Sung, H., Ferlay, J., Siegel, R.L., Laversanne, M., Soerjomataram, I., Jemal, A., Bray, F., 2021. Global cancer statistics 2020: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. Ca - Cancer J. Clin.
Tacheci, I., Podhola, M., Lesko, M., Minarik, M., Benesova, L., Kopacova, M., Rejchrt, S., Bures, J., 2017. Gastric adenocarcinoma and proximal polyposis of the stomach. Agenetic study of a newly diagnosed family. United European Gastroenterol J. 5.
Tacheci, I., Repak, R., Podhola, M., Benesova, L., Cyrany, J., Bures, J., Kohoutova, D., 2021. Gastric adenocarcinoma and proximal polyposis of the stomach (GAPPS) – a Helicobacter-opposite point. Best Pract. Res. Clin. Gastroenterol. 101728.
Taeubner, J., Wieczorek, D., Yasin, L., Brozou, T., Borkhardt, A., Kuhlen, M., 2018. Penetrance and expressivity in inherited cancer predisposing syndromes. Trends Cancer 4, 718–728.
Takaoka, M., Miki, Y., 2018. BRCA1 gene: function and deficiency. Int. J. Clin. Oncol. 23, 36–44.
Takeichi, M., 2018. Multiple functions of α-catenin beyond cell adhesion regulation. Curr. Opin. Cell Biol. 54, 24–29.
Te Paske, I., Garcia-Pelaez, J., Sommer, A.K., Matalonga, L., Starzynska, T., Jakubowska, A., van der Post, R.S., Lubinski, J., Oliveira, C., Hoogerbrugge, N., de Voer, R.M., 2021. A mosaic PIK3CA variant in a young adult with diffuse gastric cancer: case report. Eur. J. Hum. Genet.
Tedaldi, G., Pirini, F., Tebaldi, M., Zampiga, V., Cangini, I., Danesi, R., Arcangeli, V., Ravegnani, M., Abou Khouzam, R., Molinari, C., Oliveira, C., Morgagni, P., Saragoni, L., Bencivenga, M., Ulivi, P., Amadori, D., Martinelli, G., Falcini, F., Ranzani, G.N., Calistri, D., 2019. Multigene panel testing increases the number of loci associated with gastric cancer predisposition. Cancers 11.
Tischkowitz, M., Xia, B., 2010. PALB2/FANCN: recombining cancer and Fanconi anemia. Cancer Res. 70, 7353–7359.
Tung, N., Battelli, C., Allen, B., Kaldate, R., Bhatnagar, S., Bowles, K., Timms, K., Garber, J.E., Herold, C., Ellisen, L., Krejdovsky, J., DeLeonardis, K., Sedgwick, K., Soltis, K., Roa, B., Wenstrup, R.J., Hartman, A.R., 2015. Frequency of mutations in individuals with breast cancer referred for BRCA1 and BRCA2 testing using next-generation sequencing with a 25-gene panel. Cancer 121, 25–33.
Tung, N., Lin, N.U., Kidd, J., Allen, B.A., Singh, N., Wenstrup, R.J., Hartman, A.R., Winer, E.P., Garber, J.E., 2016. Frequency of germline mutations in 25 cancer susceptibility genes in a sequential series of patients with breast cancer. J. Clin. Oncol. 34, 1460–1468.
van der Post, R.S., Vogelaar, I.P., Carneiro, F., Guilford, P., Huntsman, D., Hoogerbrugge, N., Caldas, C., Schreiber, K.E., Hardwick, R.H., Ausems, M.G., Bardram, L., Benusiglio, P.R., Bisseling, T.M., Blair, V., Bleiker, E., Boussioutas, A., Cats, A., Coit, D., DeGregorio, L., Figueiredo, J., Ford, J.M., Heijkoop, E., Hermens, R., Humar, B., Kaurah, P., Keller, G., Lai, J., Ligtenberg, M.J., O'Donovan, M., Oliveira, C., Pinheiro, H., Ragunath, K., Rasenberg, E., Richardson, S., Roviello, F., Schackert, H., Seruca, R., Taylor, A., Ter Huurne, A., Tischkowitz, M., Joe, S.T., van Dijck, B., van Grieken, N.C., van Hillegersberg, R., van Sandick, J.W., Vehof, R., van Krieken, J.H., Fitzgerald, R.C., 2015a. Hereditary diffuse gastric cancer: updated clinical guidelines with an emphasis on germline CDH1 mutation carriers. J. Med. Genet. 52, 361–374.
van der Post, R.S., Vogelaar, I.P., Manders, P., van der Kolk, L.E., Cats, A., van Hest, L.P., Sijmons, R., Aalfs, C.M., Ausems, M.G., Gómez García, E.B., Wagner, A., Hes, F.J., Arts, N., Mensenkamp, A.R., van Krieken, J.H., Hoogerbrugge, N., Ligtenberg, M.J., 2015b. Accuracy of hereditary diffuse gastric cancer testing criteria and outcomes in patients with a germline mutation in CDH1. Gastroenterology 149, 897–906 e819.
van der Post, R.S., van Dieren, J., Grelack, A., Hoogerbrugge, N., van der Kolk, L.E., Snaebjornsson, P., Lansdorp-Vogelaar, I., van Krieken, J.H., Bisseling, T.M., Cats, A., 2018. Outcomes of screening gastroscopy in first-degree relatives of patients fulfilling hereditary diffuse gastric cancer criteria. Gastrointest. Endosc. 87, 397–404 e392.
van Roy, F., 2014. Beyond E-cadherin: roles of other cadherin superfamily members in cancer. Nat. Rev. Cancer 14, 121–134.
Vasen, H.F., Blanco, I., Aktan-Collan, K., Gopie, J.P., Alonso, A., Aretz, S., Bernstein, I., Bertario, L., Burn, J., Capella, G., Colas, C., Engel, C., Frayling, I.M., Genuardi, M., Heinimann, K., Hes, F.J., Hodgson, S.V., Karagiannis, J.A., Lalloo, F., Lindblom, A., Mecklin, J.P., Møller, P., Myrhoj, T., Nagengast, F.M., Parc, Y., Ponz de Leon, M., Renkonen-Sinisalo, L., Sampson, J.R., Stormorken, A., Sijmons, R.H., Tejpar, S., Thomas, H.J., Rahner, N., Wijnen, J.T., Järvinen, H.J., Möslein, G., 2013. Revised guidelines for the clinical management of Lynch syndrome (HNPCC): recommendations by a group of European experts. Gut 62, 812–823.
Vaseva, A.V., Marchenko, N.D., Ji, K., Tsirka, S.E., Holzmann, S., Moll, U.M., 2012. p53 opens the mitochondrial permeability transition pore to trigger necrosis. Cell 149, 1536–1548.
Velázquez, C., Esteban-Cardeñosa, E.M., Lastra, E., Abella, L.E., de la Cruz, V., Lobatón, C.D., Durán, M., Infante, M., 2019. A PALB2 truncating mutation: implication in cancer prevention and therapy of Hereditary Breast and Ovarian Cancer. Breast 43, 91–96.
Vissers, L.E., Veltman, J.A., 2015. Standardized phenotyping enhances Mendelian disease gene identification. Nat. Genet. 47, 1222–1224.
Vogelaar, I.P., van der Post, R.S., van Krieken, J.H.J., Spruijt, L., van Zelst-Stams, W.A., Kets, C.M., Lubinski, J., Jakubowska, A., Teodorczyk, U., Aalfs, C.M., van Hest, L.P., Pinheiro, H., Oliveira, C., Jhangiani, S.N., Muzny, D.M., Gibbs, R.A., Lupski, J.R., de Ligt, J., Vissers, L., Hoischen, A., Gilissen, C., van de Vorst, M., Goeman, J.J., Schackert, H.K., Ranzani, G.N., Molinaro, V., Gómez García, E.B., Hes, F.J., Holinski-Feder, E., Genuardi, M., Ausems, M., Sijmons, R.H., Wagner, A., van der Kolk, L.E., Bjørnevoll, I., Høberg-Vetti, H., van Kessel, A.G., Kuiper, R.P., Ligtenberg, M.J.L., Hoogerbrugge, N., 2017. Unraveling genetic predisposition to familial or early onset gastric cancer using germline whole-exome sequencing. Eur. J. Hum. Genet. 25, 1246–1252.
Vogt, S., Jones, N., Christian, D., Engel, C., Nielsen, M., Kaufmann, A., Steinke, V., Vasen, H.F., Propping, P., Sampson, J.R., Hes, F.J., Aretz, S., 2009. Expanded extracolonic tumor spectrum in MUTYH-associated polyposis. Gastroenterology 137, 1976–1985 e1971-1910.
Vos, J.R., Giepmans, L., Röhl, C., Geverink, N., Hoogerbrugge, N., 2019. Boosting care and knowledge about hereditary cancer: European reference Network on genetic tumour risk syndromes. Fam. Cancer 18, 281–284.
Weren, R.D.A., van der Post, R.S., Vogelaar, I.P., van Krieken, J.H., Spruijt, L., Lubinski, J., Jakubowska, A., Teodorczyk, U., Aalfs, C.M., van Hest, L.P., Oliveira, C., Kamping, E.J., Schackert, H.K., Ranzani, G.N., Gómez García, E.B., Hes, F.J., Holinski-Feder, E., Genuardi, M., Ausems, M., Sijmons, R.H., Wagner, A., van der Kolk, L.E., Cats, A., Bjørnevoll, I., Hoogerbrugge, N., Ligtenberg, M.J.L., 2018. Role of germline aberrations affecting CTNNA1, MAP3K6 and MYD88 in gastric cancer susceptibility. J. Med. Genet. 55, 669–674.
WHO Classification of Tumours Editorial Board, 2019. Digestive System Tumours, 5th Ed. WHO Classification of Tumours Series. International Agency for Research on Cancer, Lyon (France).
Win, A.K., Reece, J.C., Dowty, J.G., Buchanan, D.D., Clendenning, M., Rosty, C., Southey, M.C., Young, J.P., Cleary, S.P., Kim, H., Cotterchio, M., Macrae, F.A., Tucker, K.M., Baron, J.A., Burnett, T., Le Marchand, L., Casey, G., Haile, R.W., Newcomb, P.A., Thibodeau, S.N., Hopper, J.L., Gallinger, S., Winship, I.M., Lindor, N.M., Jenkins, M.A., 2016. Risk of extracolonic cancers for people with biallelic and monoallelic mutations in MUTYH. Int. J. Cancer 139, 1557–1563.
Worthley, D.L., Phillips, K.D., Wayte, N., Schrader, K.A., Healey, S., Kaurah, P., Shulkes, A., Grimpen, F., Clouston, A., Moore, D., Cullen, D., Ormonde, D., Mounkley, D., Wen, X., Lindor, N., Carneiro, F., Huntsman, D.G., Chenevix-Trench, G., Suthers, G.K., 2012. Gastric adenocarcinoma and proximal polyposis of the stomach (GAPPS): a new autosomal dominant syndrome. Gut 61, 774–779.
Xicola, R.M., Li, S., Rodriguez, N., Reinecke, P., Karam, R., Speare, V., Black, M.H., LaDuca, H., Llor, X., 2019. Clinical features and cancer risk in families with pathogenic CDH1 variants irrespective of clinical criteria. J. Med. Genet. 56, 838–843.
Yanaru-Fujisawa, R., Nakamura, S., Moriyama, T., Esaki, M., Tsuchigame, T., Gushima, M., Hirahashi, M., Nagai, E., Matsumoto, T., Kitazono, T., 2012. Familial fundic gland polyposis with gastric cancer. Gut 61, 1103–1104.
Yang, C., Arnold, A.G., Trottier, M., Sonoda, Y., Abu-Rustum, N.R., Zivanovic, O., Robson, M.E., Stadler, Z.K., Walsh, M.F., Hyman, D.M., Offit, K., Zhang, L., 2016. Characterization of a novel germline PALB2 duplication in a hereditary breast and ovarian cancer family. Breast Cancer Res. Treat. 160, 447–456.
Zhou, J., Zhao, Z., Zhang, Y., Bao, C., Cui, L., Cai, S., Bai, Y., Shen, L., Zhang, X., 2020. Pathogenic germline mutations in Chinese patients with gastric cancer identified by next-generation sequencing. Oncology 98, 583–588.
Zhu, X., Petrovski, S., Xie, P., Ruzzo, E.K., Lu, Y.F., McSweeney, K.M., Ben-Zeev, B., Nissenkorn, A., Anikster, Y., Oz-Levi, D., Dhindsa, R.S., Hitomi, Y., Schoch, K., Spillmann, R.C., Heimer, G., Marek-Yagel, D., Tzadok, M., Han, Y., Worley, G., Goldstein, J., Jiang, Y.H., Lancet, D., Pras, E., Shashi, V., McHale, D., Need, A.C., Goldstein, D.B., 2015. Whole-exome sequencing in undiagnosed genetic diseases: interpreting 119 trios. Genet. Med. 17, 774–781.

Appendix B

Combined loss of *CDH1* and downstream regulatory sequences drive early-onset diffuse gastric cancer and increases penetrance of Hereditary Diffuse Gastric Cancer

Celina São José*, José Garcia-Pelaez*, Marta Ferreira, Oscar Arrieta, Ana André, Nelson Martins, Samantha Solís, Braulio Martínez-Benítez, Mª Luisa Ordóñez-Sánchez, Maribel Rodríguez-Torres, Anna K. Sommer, Iris B.A.W. te Paske, Carlos Caldas, Marc Tischkowitz, Maria Teresa Tusié, Solve-RD DITF-GENTURIS, Nicoline Hoogerbrugge, German Demidov, Richarda M. de Voer, Steve Laurie, Carla Oliveira

Gastric cancer

2023

DOI: 10.1007/s10120-023-01395-0

Gastric Cancer
https://doi.org/10.1007/s10120-023-01395-0

ORIGINAL ARTICLE

Combined loss of *CDH1* and downstream regulatory sequences drive early-onset diffuse gastric cancer and increase penetrance of hereditary diffuse gastric cancer

Celina São José · José Garcia-Pelaez · Marta Ferreira · Oscar Arrieta · Ana André · Nelson Martins, et al. ***[full author details at the end of the article]***

Received: 8 March 2023 / Accepted: 30 April 2023
© The Author(s) 2023

Abstract

Background Germline *CDH1* pathogenic or likely pathogenic variants cause hereditary diffuse gastric cancer (HDGC). Once a genetic cause is identified, stomachs' and breasts' surveillance and/or prophylactic surgery is offered to asymptomatic *CDH1* carriers, which is life-saving. Herein, we characterized an inherited mechanism responsible for extremely early-onset gastric cancer and atypical HDGC high penetrance.

Methods Whole-exome sequencing (WES) re-analysis was performed in an unsolved HDGC family. Accessible chromatin and *CDH1* promoter interactors were evaluated in normal stomach by ATAC-seq and 4C-seq, and functional analysis was performed using CRISPR-Cas9, RNA-seq and pathway analysis.

Results We identified a germline heterozygous 23 Kb *CDH1-TANGO6* deletion in a family with eight diffuse gastric cancers, six before age 30. Atypical HDGC high penetrance and young cancer-onset argued towards a role for the deleted region downstream of *CDH1*, which we proved to present accessible chromatin, and *CDH1* promoter interactors in normal stomach. CRISPR-Cas9 edited cells mimicking the *CDH1-TANGO6* deletion display the strongest *CDH1* mRNA downregulation, more impacted adhesion-associated, type-I interferon immune-associated and oncogenic signalling pathways, compared to wild-type or *CDH1*-deleted cells. This finding solved an 18-year family odyssey and engaged carrier family members in a cancer prevention pathway of care.

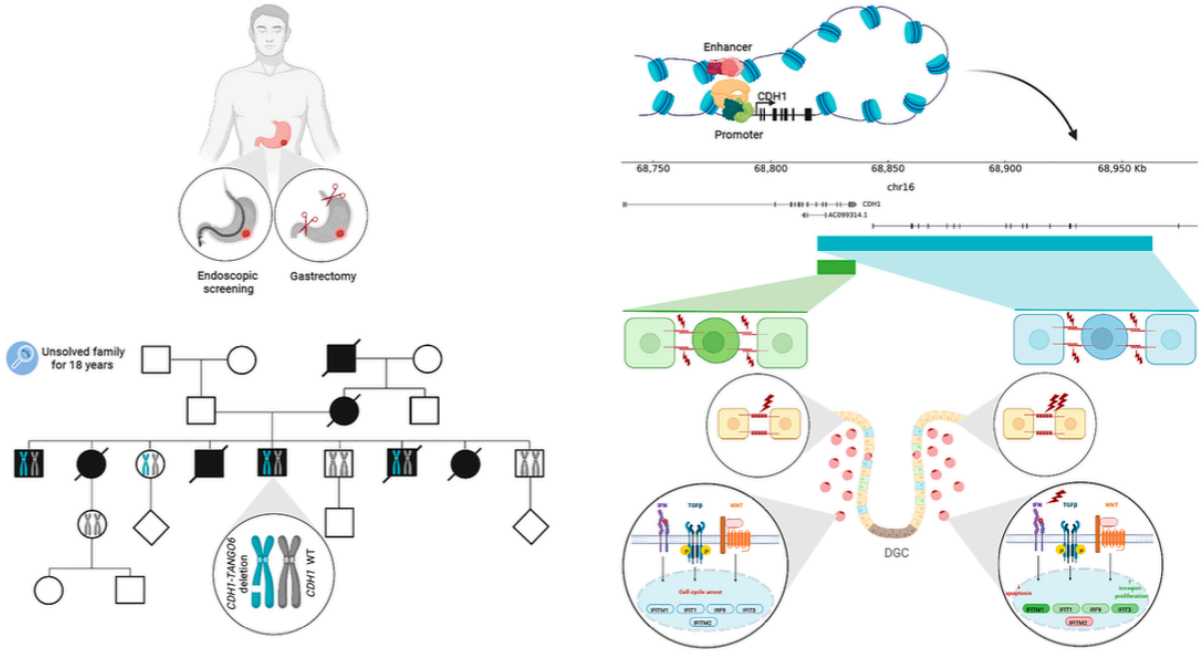
Conclusion In this work, we demonstrated that regulatory elements lying down-stream of *CDH1* are part of a chromatin network that control *CDH1* expression and influence cell transcriptome and associated signalling pathways, likely explaining high disease penetrance and very young cancer-onset. This study highlights the importance of incorporating scientific–technological updates and clinical guidelines in routine diagnosis, given their impact in timely genetic diagnosis and disease prevention.

Celina São José and José Garcia-Pelaez contributed equally to this work.

The collaborators names associated with Solve-RD DITF-GENTURIS are given in acknowledgements sections.

Published online: 30 May 2023

Graphical abstract



Keywords CDH1 · Hereditary diffuse gastric cancer · Regulatory elements · Copy number variants · Deletion · Type-I interferon immune response

Introduction

Pathogenic and likely pathogenic germline truncating single-nucleotide (SNVs) and copy number variants (CNVs) affecting the *CDH1* coding sequence [1–3], as well as, truncating *CTNNA1* SNVs [4, 5] cause hereditary diffuse gastric cancer (HDGC) syndrome [3, 5, 6]. *CDH1* and *CTNNA1* genes encode epithelial adherens junctions proteins, E-cadherin and αE-catenin respectively [7, 8], whose loss of function reduces cell–cell adhesion and increases invasiveness and cell survival, amongst other consequences [7, 9]. HDGC is autosomal-dominant and predisposes to diffuse gastric cancer (DGC) and lobular breast cancer (LBC) [1, 3].

In *CDH1*-associated HDGC, the cumulative DGC incidence ranges from 42 to 72% in males [10, 11], and from 33 to 55% in females, whilst the risk of LBC ranges from 39 to 55% in females, by 80 years of age [10–13]. Even taking into account that estimates are likely to be biassed, either due to patient geographical origin or due to clinical ascertainment [11], the risk of developing DGC or LBC remains high across different cohorts and studies [3]. Individuals with suspected HDGC are identified and referred for genetic testing, according to guidelines and clinical criteria, which have been evolving since 1999 (supplementary Fig. 1) [14–17], based on technological advances and recognition of novel HDGC-related clinical presentations [3]. Identification of a *CDH1* variant in the proband, classified as pathogenic or likely pathogenic, according to the American College of Medical Genetics and Genomics and the Association for Molecular Pathology (ACMG/AMP) variant curation guidelines for *CDH1* germline variants [18], leads to cascade genetic testing in other family members [17]. Once carrier status is confirmed in asymptomatic relatives, endoscopic surveillance and/or prophylactic surgery are offered according to guidelines to prevent DGC in both genders and/or LBC in females [17].

Many families fulfilling the 2020 HDGC clinical criteria remain genetically unexplained, despite the use of candidate-gene, whole-exome (WES), and whole-genome (WGS) sequencing approaches [6, 19]. Missing hereditability in such families may derive from screening approaches excluding CNV analysis or by technical limitations still existing in next-generation sequencing (NGS) protocols and analyses pipelines [20, 21]. In addition, missing hereditability may be hidden in regulatory elements or unscreened genomic regions [22]. The 3D chromatin architecture, and enhancers,

may drive gene expression across large genomic distances through physical proximity to the promoter [23]. These mechanisms are often cell-type specific and control overall gene expression, including that of important disease-causing genes [24, 25]. In fact, the gene expression profile of some disease-causing genes is controlled by regulatory elements located in gene bodies, intergenic regions in their vicinity, and other regions of the causal-gene topological-associating domain [26]. CNVs or other type of aberrations affecting these regulatory regions, which consequently interfere with the normal 3D chromatin architecture, may lead to misexpression patterns, causing or aggravating disease [26].

Given the advances in knowledge on all above-described layers, retesting families who tested negative prior to a syndrome-specific breakthrough, implementation of relevant technological improvements and integration of well-validated novel disease-causative mechanisms in routine testing are important aspects in diagnosis and management [27–29].

Herein, we wish to characterize a novel inherited mechanism, that is simultaneously disease-causing and likely responsible for extremely early-onset gastric cancer and atypical high penetrance of HDGC.

Materials/subjects and methods

Subjects

A HDGC-suspected family from Mexico was enrolled for meeting the 1999 HDGC clinical criteria [14]. The family presents eight DGC with signet-ring cells. WES of individuals III-1, III-3, III-5, III-6, III-7 and III-9 was available for CNVs re-analysis. Multiplex ligation-dependent probe amplification (MLPA) validation was performed for individuals III-1, III-5, III-6, III-9, IV-1 and IV-3.

Whole-exome sequencing (WES) re-analysis

WES from HDGC-suspected family [19] was re-analyzed and screened for CNVs using three variant callers (ExomeDepth [30], ClinCNV [31] and Manta [32]) in tumour risk syndromes-associated genes (supplementary table 1) and annotated using AnnotSV [33]. Candidate calls were prioritized considering quality scores, calling by more than one caller and visual evaluation using Integrated Genome Viewer (IGV).

Multiplex ligation-dependent probe amplification (MLPA)

Specific *CDH1* probes (SALSA MLPA-Probemix P083 *CDH1*, MRC-Holland) were used to identify gene dosage alterations, as previously described [2, 34] and following manufacturer's recommendations. Probemix includes 20 *CDH1*-specific probes, 2 probes in *CDH1* flanking genes and 13 reference probes in relatively copy number stable regions. Data analysis was performed using Coffalyser.Net (MRC-Holland).

ATAC-seq and 4C-seq

ATAC-seq data was collected from phase 3 ENCODE project [35]. For 4C-seq, gastric mucosa cells from bariatric surgeries were scraped and enzymatically dissociated. 4C-seq libraries were generated, as previously described [36, 37], digested with DnpII (New England Biolabs) and Csp6I (New England Biolabs) and re-circularized with T4 DNA Ligase (Thermo Fisher Scientific). 4C-seq libraries were PCR-amplified (supplementary Table 2) and sequenced on Illumina HiSeqX technology.

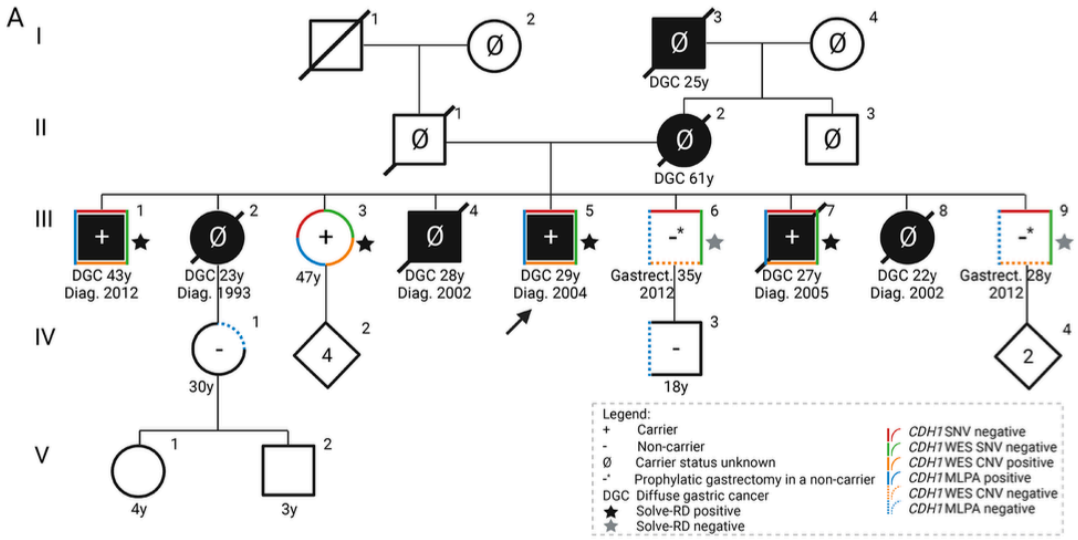
CDH1 interactions on a genome scale were mapped, based on Pipe4C [37] and PeakC [38] with window size 2, alpha fdr 0.1 and minimal distance 500.

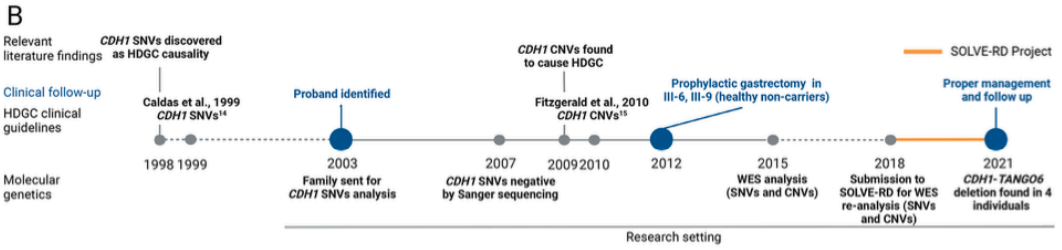
Cell culture

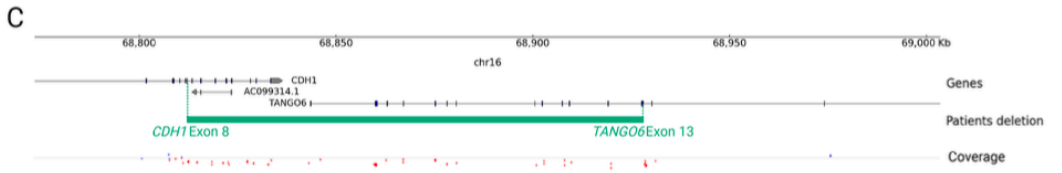
MKN74 human gastric cancer cell line was purchased from the Japanese collection of research bioresources cell bank. MKN74 and isogenic clones were cultured in RPMI medium (Gibco), supplemented with 10% foetal bovine serum (Biowest) and 1% penicillin streptomycin (Gibco) and maintained at 37 °C under 5% CO_2 humidified atmosphere.

CRISPR-Cas9 editing

sgRNAs were designed to target the family CNV or the *CDH1* portion of that CNV using Benchling online platform (supplementary table 3). Individual sgRNAs (Invitrogen) were cloned in LentiCRISPRv2GFP (addgene 82416) or LentiCRISPRv2-mCherry (addgene 99154) vectors using BsmBIv2 (New England Biolabs). Plasmids were transformed into Stbl3 competent cells and colonies were sequenced (supplementary Table 4). Lentiviral particles were produced resourcing to HEK293T cell line with pMD2.G (addgene 12259) and pCMV-dR8.91 (addgene) vectors, following Lipofectamine 3000 manufacture's protocol (Invitrogen) and collected at 48 h. MKN74 was infected with pairs of lentivirus particles in medium supplemented with 10 µg/µl hexadimethrine bromide (Merck Life Science S.L.U.) for 48 h. Transduced cells were selected for GFP and mCherry positive expression at 7 days post-infection using FACS ARIA (BD Biosciences).

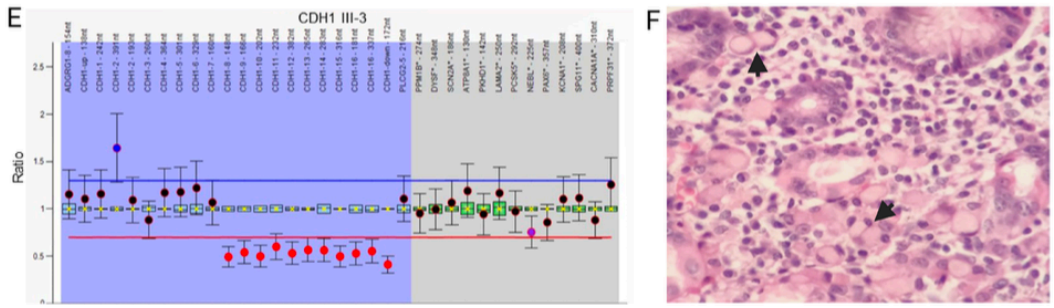






D

ID	ClinCNV	ExomeDepth	Manta
III-1	chr16:68812132-68930295	chr16:68812133-68930295	chr16:68811891-68965337
III-3	chr16:68811681-68928082	chr16:68812133-68930295	No call
III-5	chr16:68812132-68928082	chr16:68812133-68928083	chr16:68811891-68965337
III-6	No call	No call	No call
III-7	chr16:68812132-68930295	chr16:68812133-68930295	chr16:68811891-68965337
III-9	No call	No call	No call
IV-I	No cal	No cal	No call
IV-3	No call	No call	No call



◂**Fig. 1** Diagnostic odyssey, pedigree, variant calling details and cancer histology of a HDGC family carrying a causative *CDH1* CNV. **A** Current family pedigree. Full black symbols: individuals with confirmed DGC; Red outline: negative for SNVs (Sanger sequencing); green outline: negative for SNVs (WES); orange outline: positive for CNVs (WES); blue outline: MLPA-positive for the *CDH1* deletion; traced orange outline: negative for CNVs (WES); traced blue outline: MLPA-negative for the CDH1 deletion. All family members, submitted to multiple genetic tests herein described, signed informed consents. The research work has been approved by the Ethics Committee of Centro Hospitalar Universitário São João, in Porto, Portugal, in the frame of the Solve-RD project with the reference CHUSJ_445/2020; **B** Timeline of events and diagnostic odyssey of the family; **C** *CDH1* CNV found by WES, encompassing part of the *CDH1* and *TANGO6* genes (represented in green) and IGV coverage; Green text represents first and last edited exons; **D** Detected *CDH1* CNV in ClinCNV [31], ExomeDepth [30] and Manta [32]; **E** MLPA analysis performed using SALSA MLPA-Probemix P083 *CDH1* (MRC Holland) in patient III-3; Ratio above blue line indicates increased copy number, whilst ratio below red line indicates reduced copy number; Blue highlight represents *CDH1* probes and 2 neighbour genes; Grey highlight represents reference probes; **F** Haematoxylin and eosin staining of DGC depicting signet-ring cells (arrow heads) (100×magnification) from the proband III-5

Genotyping of edited clones

gDNA was extracted using NZY Tissue gDNA isolation kit (NZYTech), according to the manufacturers' protocol. gDNA was amplified using primers flanking the edition sites (supplementary table 4) and Multiplex PCR kit (Qiagen). PCR products were analysed in gel electrophoresis and Sanger sequenced on an ABI-3130 Genetic Analyzer (Applied Biosystems).

CDH1/E-cadherin expression analysis

CDH1 mRNA expression was assessed by qPCR in triplicates. RNA was extracted using *mir*Vana RNA Isolation Kit (Invitrogen), according to manufacturers' protocol. cDNA was synthesized using SuperScriptII reverse transcriptase (Invitrogen), according to the manufacturers' protocol. *CDH1* mRNA expression was analyzed by qPCR with KAPA PROBE FAST qPCR Master Mix (2X) Kit (Sigma-Aldrich) and probes for *CDH1* (Hs. PT.58.3324071, TaqMan) and 18S (custom assay, IDT). Reactions were sequenced on a 7500 Real-Time PCR System (Applied Biosystems). Relative expression was normalized for the endogenous 18S control and quantified using the $2^{-\Delta\Delta Ct}$ method.

E-cadherin expression was assessed by flow cytometry in triplicates. Cells were detached with Versene (Gibco) and blocked with 3% bovine serum albumin–phosphatase-buffered saline. Cells were incubated with primary mouse monoclonal antibody HECD-1 (Invitrogen), washed and incubated with secondary antibody anti-mouse Alexa Fluor 647 (Invitrogen). Fluorescence was measured using FACS ARIA (BD Biosciences) and Flow Jo version 10 software was used to analyse the data.

Statistical analysis

Statistical analysis was performed using GraphPad Prism version 7.00 software (GraphPad Software Inc.). A Student's t test was used for comparison analysis, assuming equal variance between clones and WT samples. Differences were considered significant when p value < 0.05.

RNA-seq and whole transcriptome sequencing bioinformatics analysis

Duplicated RNA samples collected from control and CRISPR-cas9-edited cell lines were sequenced with TruSeq stranded total RNA with Ribo-Zero Gold library in NovaSeq6000. Unique reads were mapped to GRCh38 human genome using STAR and RSeQC and annotated with USCS GRCh38 (Genecode v36). Batch effect was assessed using a PCA and corrected with ComBat_seq from sba (v.3.44) R package for replicates and samples. Differential expression analysis was performed using Deseq2 (v.1.36.0) R package and canonical transcripts were selected considering $|\log2\text{fold-change}| \geq 1$ and corrected p value < 0.05. ClusterProfiler (v.4.4.4) R package was used for enrichment GO terms analysis (q value < 0.05). Statistics were performed using R (v.4.2.0).

Results

The genetic odyssey of a HDGC-suspected family ends with the finding of a *CDH1-TANGO6* deletion

A family of Mexican origin, meeting the 1999 HDGC clinical criteria [14], was identified in 2003 showing five DGC in three consecutive generations, three occurring in the same generation at ages 23 (III-2), 28 (III-4) and 22 (III-8) (Fig. 1A). In addition, individuals III-5 and III-7 also developed DGC at age 29 and 27, in 2004 and 2005, respectively. Given the highly suggestive HDGC pedigree, samples from family members that remained alive (III-1, III-3, III-5, III-6, III-7 and III-9), were all sent out for *CDH1* Sanger sequencing, under a research protocol (Fig. 1A, B). The genetic test was negative for the presence of *CDH1* SNVs in 2007. Meanwhile, all asymptomatic family members were integrated into a gastric surveillance protocol, similar to that offered to genetically unsolved HDGC families [14]. In 2012, at age 43, individual III-1 developed DGC and underwent gastric surgery with

curative intent. Due to the absence of a genetic diagnosis to find individuals at higher risk of DGC, the clinical team decided to offer (presumably) prophylactic total gastrectomy to asymptomatic individuals III-6 and III-9 (Fig. 1A).

Although, *CDH1* CNVs were first described as a cause of HDGC in 2009 [2], and integrated in HDGC clinical guidelines in 2010 [15], samples from this family were not tested by MLPA. Samples were rather submitted to WES in 2015, again as part of a research protocol, and analysed for SNVs and CNVs which result turned out negative (Fig. 1B), likely due to technical limitations in WES analysis or sample/sequencing quality [39].

WES re-analysis led to the identification of a heterozygous deletion involving half of the *CDH1* gene (starting downstream of *CDH1* exon 7) and the downstream *TANGO6* gene (ending upstream of *TANGO6* exon 14) (Fig. 1B–D). This CNV was identified by several variant callers, namely 'ClinCNV' [31], 'Manta' [32] and 'ExomeDepth' [30] (Fig. 1C,D), and supported by Integrative Genomics Viewer (IGV) visual inspection [40], despite the low quality of the sequencing data (Fig. 1C). The CNV was after validated by MLPA (Fig. 1E) in individuals III-1, III-3, III-5 and III-7 (Fig. 1A). Individuals III-6 and III-9, who had opted for risk-reducing gastrectomy, were actually non-carriers of the CNV (Fig. 1A).

As the pedigree evidences (Fig. 1A), three out of four carriers in generation III, developed DGC with signet-ring cells at ages 27, 29 and 43 (Fig. 1F). Individual III-3 is the only carrier to our knowledge that remains asymptomatic at the age of 47. Besides these family members, there are five not-tested DGC patients, two (I-3 and II-2) died from DGC at ages 25 and 61, being in principle obligated carriers. There are three other relatives (III-2, III-4 and III-8) who died of DGC at age 23, 28 and 22 and were likely also carriers of this predisposing CNV (Fig. 1A). After the genetic diagnosis, a blood sample was obtained from individuals IV-1 and IV-3, who were old enough to be tested, and the CNV screening performed by MLPA returned a negative result for both (Fig. 1A).

Chromatin conformation analysis highlights physical interactions between the *CDH1* promoter and sequences within the deleted region in normal stomach

Although highly penetrant, HDGC caused by *CDH1* pathogenic or likely pathogenic variants is rarely as penetrant as in this family, which presents an uncommonly high number of individuals in a single generation affected by DGC (six out of seven carriers/presumed carriers). Additionally, five out of six developed DGC before age 29, and average age of cancer onset was 29 ± 7.55 years in six carriers/presumed carriers from generation III. We therefore hypothesised that sequences within the deleted region, and extending into the downstream intergenic region, and the neighbouring *TANGO6* gene sequence, would affect *CDH1* expression and the *locus* 3D chromatin structure more strongly, than deletion of the *CDH1* alone.

To address this hypothesis, we mined the GeneHancer database [41] and found that several enhancers within the deleted region, interact with either *TANGO6* or *CDH1* promoters (Fig. 2A). Additionally, we used normal epithelial cells from bariatric surgeries to profile *CDH1* promoter interactions using 4C-seq, and analysed accessible chromatin regions using ATAC-seq from the ENCODE project [35]. We found that the *CDH1* promoter interacts not only with downstream sequences within *CDH1*, but also *TANGO6* and the intergenic region, all deleted by the family CNV, and displaying regions of accessible chromatin (Fig. 2A). These results suggest a role of *CDH1*/*TANGO6* co-regulation in stomach, besides gene physical proximity.

Deletion of *CDH1* and downstream regulatory sequences, present within the HDGC causal CNV, strongly impact *CDH1* expression and specific signalling pathways

To understand the impact of deleting several *CDH1* promoter-interacting sequences, beyond the *CDH1 locus*, we CRISPR-Cas9 deleted either the *CDH1* portion of the CNV, or mimicked the family CNV. This was performed in a human gastric cancer cell line bearing normal *CDH1*/E-cadherin expression and function [42]. We used sgRNAs flanking target regions (supplementary Table 3), and successful editing was confirmed by Sanger sequencing and qRT-PCR (Fig. 2B,C). A clone bearing a 23.705 bp homozygous deletion spanning from *CDH1* intron 7 and extending till exon 16 (chr16:68,811,911–68,835,617, hg38) and a clone bearing a 116.245 bp homozygous deletion starting in *CDH1* intron 7 and ending after *TANGO6* intron 13 (chr16:68,811,915–68,928,161, hg38) were produced (Fig. 2B).

We found that both *CDH1* and *CDH1-TANGO6* deletions led to a significant decrease in *CDH1* mRNA compared to the wild-type (WT) control (p value = 0.045 and p value = 0.0032, respectively), being this downregulation stronger for the *CDH1-TANGO6* deletion as compared to the *CDH1* deletion alone (p value = 0.012) (Fig. 2D). The level of protein downregulation in relation to the WT control was equivalent in both deletion clones, as assessed by flow cytometry (p value < 0.0001 and p value = 0.0002, respectively) (Fig. 2E).

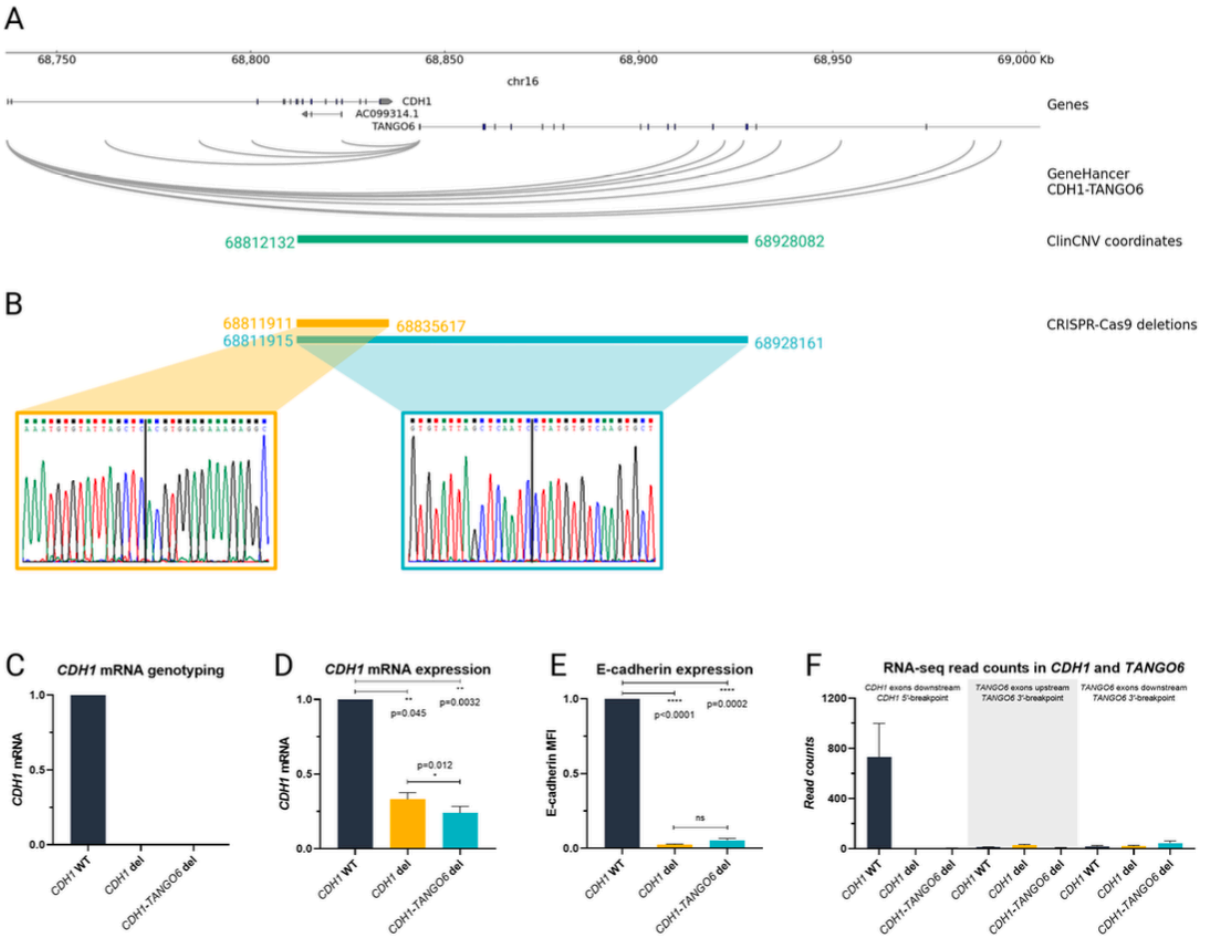
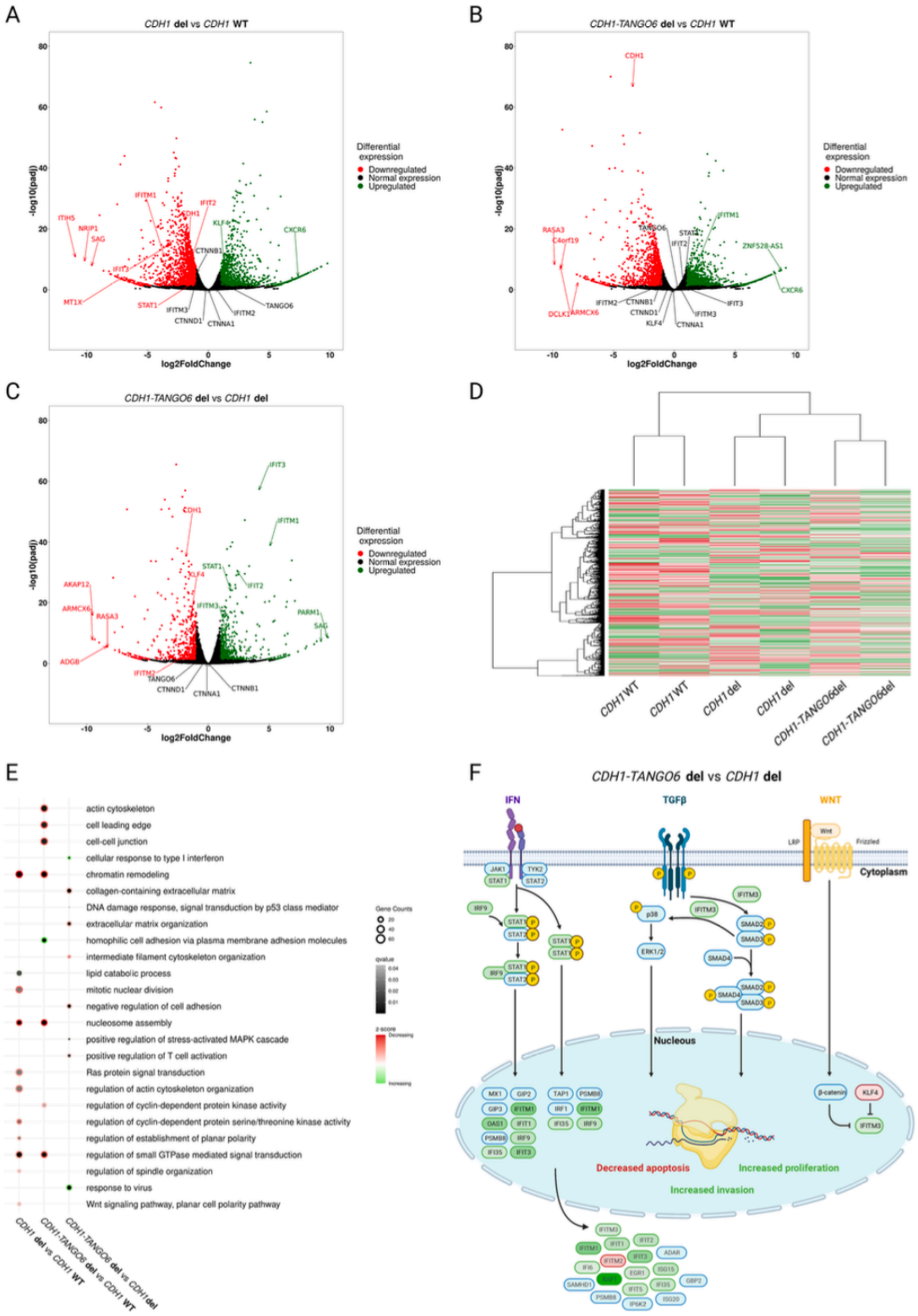


Fig. 2 *CDH1* and *TANGO6* regulatory network and characterization of CRISPR-cas9-edited clones. **A** *CDH1*/*TANGO6* potential enhancers depicted in GeneHancer, accessible chromatin and *CDH1* promotor interactors in normal stomach tissue; **B** Sanger sequencing genotyping of the CNV breakpoints, and deletion coordinates in *CDH1* and *CDH1-TANGO6* CRISPR-cas9-edited clones; **C** mRNA genotyping of *CDH1* and *CDH1-TANGO6*-edited clones resourcing to a probe located in the deleted *CDH1* region; **D** *CDH1* mRNA expression of *CDH1*- and *CDH1-TANGO6*-edited clones resourcing to a probe located in exons 6–7 (Hs.PT.58.3324071, TaqMan); **E** E-cadherin protein expression of *CDH1*- and *CDH1-TANGO6*-edited clones measured by flow cytometry (monoclonal antibody HECD-1, Invitrogen, 1:100 dilution; and mouse Alexa Fluor 647, Invitrogen, 1:500 dilution); **F** RNA-seq expression read counts of *CDH1* and *TANGO6* genes. Data are represented as the mean ± SEM, MFI: median fluorescence intensity. Experiments depicted in panels C–E were performed in triplicates and differences considered statistically significant if p value < 0.05 in a t-test (details in Materials and Methods)

We next explored the effect of both deletions in the cells' transcriptomes. For this, we evaluated the differential transcriptome between *CDH1* or *CDH1-TANGO6* deletion clones and WT cells, and between *CDH1-TANGO6* and *CDH1* deletion clones. We confirmed that *CDH1* mRNA expression was more strongly downregulated (3.4-fold) in the *CDH1-TANGO6* deletion than in the *CDH1* deletion alone (1.6-fold), when each was compared to *CDH1* mRNA expression in WT cells (p value < 0.0001) (Figs. 2F, 3A,B). We also verified that *TANGO6* is not expressed in WT cells, and its expression remains, as expected, unchanged after the 5'-deletion of the *TANGO6* gene (Fig. 2F).

We also observed that the global patterns of differentially expressed genes clearly differ between clones bearing the *CDH1* deletion alone, the *CDH1-TANGO6* deletion and the *CDH1* WT cells (Fig. 3D). There were 3,272 differentially expressed genes between the *CDH1*-deleted clone and *CDH1* WT cells; 2 492 differentially expressed genes between the *CDH1-TANGO6*-deleted clone and *CDH1* WT cells; and 1,102 differentially expressed genes between the *CDH1-TANGO6* and the *CDH1*-deleted clones (supplementary Tables 5,6,7). Gene ontology analysis further revealed that 'nucleosome assembly', 'chromatin remodelling', 'regulation of small GTPase-mediated signal transduction' as well


A
CDH1 del vs CDH1 WT
-log10(padj)
log2FoldChange
Differential expression
Downregulated
Normal expression
Upregulated
B
CDH1-TANGO6 del vs CDH1 WT
C
CDH1-TANGO6 del vs CDH1 del
D
CDH1WT
CDH1WT
CDH1del
CDH1del
CDH1-TANGO6del
CDH1-TANGO6del
E
actin cytoskeleton
cell leading edge
cell-cell junction
cellular response to type I interferon
chromatin remodeling
collagen-containing extracellular matrix
DNA damage response, signal transduction by p53 class mediator
extracellular matrix organization
homophilic cell adhesion via plasma membrane adhesion molecules
intermediate filament cytoskeleton organization
lipid catabolic process
mitotic nuclear division
negative regulation of cell adhesion
nucleosome assembly
positive regulation of stress-activated MAPK cascade
positive regulation of T cell activation
Ras protein signal transduction
regulation of actin cytoskeleton organization
regulation of cyclin-dependent protein kinase activity
regulation of cyclin-dependent protein serine/threonine kinase activity
regulation of establishment of planar polarity
regulation of small GTPase mediated signal transduction
regulation of spindle organization
response to virus
Wnt signaling pathway, planar cell polarity pathway
CDH1 del vs CDH1 WT
CDH1-TANGO6 del vs CDH1 WT
CDH1-TANGO6 del vs CDH1 del
Gene Counts
qvalue
z-score
F
CDH1-TANGO6 del vs CDH1 del
IFN
TGFβ
WNT
Cytoplasm
Nucleus
Decreased apoptosis
Increased proliferation
Increased invasion

◂**Fig. 3** Genome-wide transcriptome regulation of *CDH1* CNV and *CDH1-TANGO6* CNV. A) Volcano-plot illustrating *CDH1* CNV differentially expressed genes with *CDH1* WT; **B** Volcano-plot illustrating *CDH1-TANGO6* CNV differentially expressed genes with *CDH1* WT; **C** Volcano-plot illustrating *CDH1-TANGO6* CNV differentially expressed genes with *CDH1* CNV; **D** Heatmap illustrating genome-wide differentially expressed genes; **E** Selected genome-wide transcriptomic misregulated pathways of *CDH1-TANGO6* CNV and *CDH1* CNV. Grey scale represents q value and red–green scale represents z-score (red for pathways with mainly downregulated genes, white for equally downregulated and upregulated genes and green for pathways with mainly upregulated genes); **F** Response to interferon, TGF-β and Wnt signalling pathways misregulated in *CDH1-TANGO6* CNV vs *CDH1* CNV, created with BioRender.com. Red–green scale represents down- to upregulated genes and blue represents normal expressed genes

as pathways related to cyclin-dependent protein kinase activity and actin cytoskeleton were impacted in both in *CDH1* alone and in the *CDH1-TANGO6* deletion clones (Fig. 3E, supplementary Table 8).

CDH1 deletion exclusive events encompass downregulation of 'mitotic nuclear division', 'regulation of spindle organization' and planar polarity-related pathways, as well as upregulation of 'lipid catabolic processes', amongst others (Fig. 3E). Events exclusively altered in the *CDH1-TANGO6* deletion clone include mainly impairment of adhesion- and cytoskeleton-associated pathways (Fig. 3E).

The direct comparison between the transcriptional profile of clones bearing the *CDH1-TANGO6* deletion and the *CDH1* deletion alone emphasises the negative impact in adhesion-, p53-mediated DNA damage response and cytoskeleton-associated pathways and the positive impact in MAPK cascade-associated proliferation, presented by the *CDH1-TANGO6* deletion clone (Fig. 3E). Additional positive and negative impacts were found in several immune-associated pathways, namely related to 'cellular response to type I interferon' and 'response to virus' in the *CDH1-TANGO6* deletion clone (Fig. 3E,F). The expression of *IFITM* genes (*IFITM1*, *IFITM3* and *IFITM2*) and some of their upstream regulators (*STAT1* and *KLF4*) was particularly affected in the *CDH1-TANGO6* deletion clone (Fig. 3C,F).

Altogether, and upon data integration, cancer cells arising in patients bearing the *CDH1-TANGO6* deletion are likely to present decreased apoptosis, increased proliferation and invasion, and no cell cycle arrest (Fig. 3F).

Discussion

In this work, we demonstrated that regulatory elements, lying within the *CDH1-TANGO6* deletion, are part of a chromatin network that, not only control *CDH1* expression, but also influence the cell transcriptome and associated signalling pathways. We believe this finding explains the high penetrance and extremely young age of cancer-onset in patients from this family. In addition, we describe the diagnostic odyssey of a HDGC-suspected family, that remained genetically unsolved for 18 years, to highlight the importance of incorporating scientific–technological updates and clinical guidelines in routine diagnosis. This is expected to ultimately improve clinical management and care of hereditary cancer patients and families.

A distinctive feature of this family, in comparison with other *CDH1*-positive HDGC families, is the high disease penetrance reflected in generation III, with six out of seven individuals being CNV carriers or presumed carriers and developing DGC. The other is the very early age of cancer-onset (29 ± 7.55 years), compared to the average age reported for DGC in most HDGC families (46,7 ± 16,6 years) [3, 11, 43]. In accordance with our findings, an additional HDGC family has been reported in the literature, bearing a 275 kb *CDH1* (exon 7) and *TANGO6* (full gene) deletion, with four patients in four consecutive generations affected by gastric cancer, from which two were confirmed with signet ring cell carcinoma at ages 30 and 34 years old [44]. These features prompted the next question of this study, which was related to a potential role for the downstream intergenic region and the neighbouring *TANGO6* gene sequence, in modulating *CDH1* expression, the *CDH1 locus* 3D chromatin structure and the transcriptome, in aggravating the clinical presentation in this family.

CRISPR-Cas9 clones mimicking the *CDH1* deletion alone or the family full CNV, demonstrated that the full CNV was more efficient in promoting downregulation of *CDH1* mRNA than the *CDH1* deletion alone. This finding supports the existence of regulatory DNA elements controlling *CDH1* expression, downstream of the *CDH1 locus*, and enclosed in the CNV, which we have also proved to interact with the *CDH1* promoter. Similar findings have been previously reported, as for instance those related to neighbour murine *Shh* and *Lmbr1* genes, in which the ZRS enhancer located within the *Lmbr1* gene, regulates *Shh* gene expression [45].

Our data on differential expression and gene ontology analysis of RNA-seq data from clones mimicking the *CDH1* deletion alone or the family full CNV in comparison to *CDH1* WT, highlighted shared biological terms between these clones related to cell division and spindle orientation, and involving chromatin remodelling. These terms have been previously associated with E-cadherin function [46–49], and likely represent the contribution of the *CDH1* locus deletion itself to the differential transcriptome. The fact that adhesion-related ('negative regulation of cell adhesion') and actin-related pathways ('actin cytoskeleton') were found specifically downregulated upon *CDH1-TANGO6* deletion, and in comparison, to *CDH1*-deleted cells, suggests that the

perturbations induced by the *CDH1-TANGO6* deletion are greater than those induced by the *CDH1* deletion alone in these pathways. Although it is widely known that E-cadherin deficiency leads to weaker adhesion and abnormal microtubule organization [50, 51], which may be involved in migration upon malignant transformation [52], our data specifically link homophilic cell adhesion to the longer *CDH1-TANGO6* deletion. Indeed, this clone presents upregulation of the pathway 'homophilic cell adhesion via plasma membrane adhesion molecules', which may reflect a yet to prove adhesion-compensatory mechanism, promoted by upregulation of protocadherins in the absence of E-cadherin, seen in our data. A similar switch has been previously demonstrated in lobular breast cancer, whereby E-cadherin is replaced by P-cadherin in cancer cells [53].

Immune-associated genes and pathways were particularly altered in the *CDH1-TANGO6* deletion clone, comparing to the clone bearing the *CDH1* deletion alone. *IFITM1* gene was particularly overexpressed in the *CDH1-TANGO6* deletion, a feature previously observed in several aggressive tumours, and shown to enhance tumour proliferation and invasion [42, 54]. In gastroesophageal tumours, for instance, *IFITM1* was found overexpressed comparing to normal adjacent tissues [55–57]. Additionally, *IFITM3* overexpression has been shown to promote gastric cancer growth, through the activation of several pathways, resulting in increased proliferation, invasion, and metastasis [58, 59]. IFITM3 has been shown to promote SMAD2/3/4 phosphorylation and interact with STAT1/2 and PIP3/PI3K. This results in activation of downstream proliferation and oncogenic signalling pathways, such as p38/MAPK [60], JAK/STAT [61–63] and PI3K [64]. IFITM3 plays an additional role in the Wnt signalling pathway, being regulated by β-catenin [58] and KLF4 [65]. Indeed, KLF4, known to promote *IFITM3* transcriptional inhibition [65], is downregulated in the *CDH1-TANGO6* deletion clone providing a rationale for the *IFITM3* upregulation seen in this clone. Finally, *IFITM2,* a pro-apoptotic gene, encoding a protein which has been linked to protection against tumour proliferation [66], was found downregulated in the *CDH1-TANGO6* deletion clone. Most abovementioned observations support a relationship between *IFITM* proteins misregulation and the severe phenotype, observed in this HDGC family. Increased proliferation associated with MAPK cascade [67, 68] and decreased DNA damage response mediated by p53 signalling [69–71], also found in cells bearing the family CNV, may hint to more aggressive DGC, than the commonly described for HDGC patients bearing *CDH1*-coding variants, who generally die by dissemination to the peritoneum [43]. Supporting this, is the fact that individual III-4 died at age 28 with DGC metastases in the lung and bone, a very unusual event in patients with coding *CDH1* variants.

Our data also seem to indicate that whilst a role has not been found for the *TANGO6* mRNA itself, the genomic sequences downstream of the *CDH1 locus*, and within the sequence deleted in this family, host positive regulators of *CDH1* expression. This would explain not only the lower *CDH1* mRNA expression, but also the wider transcriptional and signalling impact, which likely contribute to the clinical presentation in this family. Only two HDGC families have been reported to carry large deletions involving specifically *CDH1* and *TANGO6*, both presenting early-onset DGC and gastric cancer in consecutive generations [44]. Therefore, evidence for the role of regulatory *CDH1* downstream regions in early-onset DGC is still limited. Yet, this study provides the first evidence on the role of *CDH1* downstream sequences in *CDH1* expression control, which may apply to other HDGC families harbouring similar deletions.

This HDGC-suspected family with a severe clinical presentation was somehow lost in genetic follow-up, as clinical standardized diagnostic tests were never used in their clinical path. Negative tests for *CDH1*-coding SNVs and CNVs, both were obtained from research projects, although management guidelines proposed formal genetic testing with Sanger sequencing and MLPA since 2010 in families fulfilling HDGC clinical criteria [2, 15]. This highlights the risk associated with molecular diagnosis being performed in the frame of research studies, rather than in proper diagnostic and clinically approved laboratories, as part of patients' and families' pathways of care. In this case, the most obvious disease gene/causing mechanism was overlooked, as for research purposes, these should have been, in principle, excluded. The result of a WES data re-analysis 18 years after clinical diagnosis, revealed a CNV involving the most obvious HDGC-causing gene, *CDH1* [1, 3, 6, 17]. This case highlights the importance of revisiting, re-contacting and retesting families fulfilling clinical criteria but lacking molecular diagnosis, in diagnostic and clinically approved laboratories, as novel and consolidated genetic susceptibility causes of hereditary syndromes arise, and both technologies and management guidelines evolve.

This family odyssey also illustrates the importance of timely genetic diagnosis. For this family, the greater consequences included two carriers developing cancer, two and nine years after the family was recognized clinically as an HDGC-suspected family; and two non-carrier siblings undergoing unnecessary and presumably prophylactic total gastrectomy, nine years after clinical diagnosis, based on family history only. The Solve-RD project (https://solve-rd.eu/) emerged in this case, as a solution to overcome an unfortunate diagnostic odyssey. The finding of a causal *CDH1* CNV was reported to the clinical team at the hospital of origin in 2021, which promoted re-engagement with the family and effective genetic counselling and disease risk management.

Taken together, our data suggest that patients bearing the *CDH1-TANGO6* deletion develop high penetrance, earlier-onset and possibly more aggressive DGC, than patients with pathogenic variants confined to the *CDH1*, likely due to the impairment of regulatory elements able to activate oncogenic signalling pathways. It also illustrates the inequities and complexities of genetic diagnosis and integrated care in rare diseases, in a rapidly evolving field, which can impact timely genetic diagnosis and disease prevention.

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1007/s10120-023-01395-0.

Acknowledgements This work was supported by: (1) the European Reference Network on Genetic Tumour Risk Syndromes (ERN GENTURIS) – Project ID No. 739547. ERN GENTURIS is partly co-funded by the European Union within the framework of the Third Health Programme "ERN-2016 – Framework Partnership Agreement 2017–2021"; (2) Doctoral Programme in Biomedicine—Faculty of Medicine, University of Porto, to JG-P and CSJ; (3) Doctoral Programme in Computational Sciences, Faculty of Sciences, University of Porto, to MF; (4) Masters in Molecular Oncology, Faculty of Medicine, University of Porto, to NM. The team would also like to thank Nuno Alves from i3S for fruitful discussions on immuno-oncology; Sónia Sousa, Cátia Pereira and Luis Cirnes from Ipatimup Diagnostics, Porto-Portugal for MLPA analysis.

Solve-RD DITF-GENTURIS: Stefan Aretz1,2, Gabriel Capella3, Sérgio Castedo4-7, Richarda M. de Voer8,9, Gareth Evans10, Susana Fernandes11, José Garcia-Pelaez4,5,12, Luzia Garrido7, Elke Holinski-Feder13,14, Nicoline Hoogerbrugge8,9, David Huntsman15,16, Arne Jahn17,18, C. Marleen Kets8,9, Andreas Laner14, Marjolijn Ligtenberg8,9,19, Andrea Meinhardt17, Arjen Mensenkamp8,9, Carla Oliveira4-6, Sophia Peters1, Isabel Quintana3, Evelin Schröck17,18, Anna Sommer1, Isabel Spier1,2, Liesbeth Spruijt8,9, Verena Steinke-Lange13,14, Iris te Paske8,9, Marc Tischkowitz20, Laura Valle3, Rachel van der Post9,19, Yasmijn van Herwaarden9,21, Wendy van Zelst-Stams8,9, Doreen William17,18. Affiliations: 1.Institute of Human Genetics, Medical Faculty, University of Bonn, Bonn, Germany. 2.Center for Hereditary Tumor Syndromes, University Hospital Bonn, Bonn, Germany. 3.Bellvitge Biomedical Research Institute (IDIBELL), Barcelona, Spain. 4.IPATIMUP - Institute of Molecular Pathology and Immunology of the University of Porto, Portugal. 5.i3S - Instituto de Investigação e Inovação em Saúde, Universidade do Porto, Portugal. 6.Faculty of Medicine, University of Porto, Portugal. 7.CHUSJ, Centro Hospitalar e Universitário de São João, Porto, Portugal. 8.Department of Human Genetics, Radboud University Medical Center, Nijmegen, The Netherlands. 9.Radboud Institute for Molecular Life Sciences, Nijmegen, The Netherlands. 10.Division of Evolution, Infection and Genomics, School of Biological Sciences, Faculty of Biology, Medicine and Health, University of Manchester, Manchester M13 9WL, UK. 11.Departament of Genetics, Faculty of Medicine, University of Porto, Portugal. 12.Doctoral Programme in Biomedicine, Faculty of Medicine, University of Porto, Portugal. 13.Medizinische Klinik und Poliklinik IV – Campus Innenstadt, Klinikum der Universität München, Munich, Germany. 14.MGZ - Medical Genetics Center, Munich, Germany. 15.Department of Pathology and Laboratory Medicine, University of British Columbia, Vancouver, British Columbia, Canada. 16.British Columbia Cancer Agency, Vancouver, British Columbia, Canada. 17.Institute of Clinical Genetics, University Hospital Carl Gustav Carus, Technical University Dresden, Dresden, Germany. 18.National Center for Tumor Diseases (NCT), Dresden, Germany. 19.Department of Pathology, Radboud University Medical Center, Nijmegen, the Netherlands. 20.Academic Department of Medical Genetics, Cambridge University, Cambridge, UK. 21.Department of Gastroenterology, Radboud University Medical Center, Nijmegen, the Netherlands.

Author contributions CSJ and JGP contributed equally to the work. CO: supervision, project administration, funding acquisition; CSJ, JGP and CO: conceptualization, formal analysis; MTT, OA, SS, BM-B, MLO-S and MR-T: acquisition of clinical and pathological data; CSJ, JGP, AA and NM: experimental work; MF, GD and SL: bioinformatic analysis; Solve-RD DITF-GENTURIS, AS, IteP, CC, MT, NH and RMdeV: integration of genetics and clinical data; all authors: writing, review, editing and approval of the manuscript.

Funding This work received funding from: (1) the European Union's Horizon 2020 research and innovation programme under grant agreement No 779257 (Solve-RD: H2020-SC1-2017-Single-Stage-RTD), which provided Ph.D. salaries to JG-P, AS, ItP; (2) The European Regional Development Fund (ERDF) through the COMPETE 2020-Operacional Programme for Competitiveness and Internationalisation (POCI), Portugal 2020, and by Portuguese funds through Fundação para a Ciência e Tecnologia (FCT)/Ministério da Ciência, Tecnologia e Inovação (LEGOH: PTDC/BTM-TEC/6706/2020); (3) FCT with salary support for Ph.D. studentships to CSJ (Ref. SFRH/BD/140796/2018), JG-P (Ref. 2022.11952.BD) and MF (Ref. 2020.05763.BD); (4) MT is supported by the NIHR Cambridge Biomedical Research Centre (BRC-1215-20014). Open access funding provided by FCT|FCCN (b-on).

Data availability Raw data supporting the findings of this study are available upon request from the corresponding author. Processed data are available in supplementary materials.

Declarations

Conflcit of interest The authors disclose no conflict of interest.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit http://creativecommons.org/licenses/by/4.0/.

References

1. Guilford P, Hopkins J, Harraway J, McLeod M, McLeod N, Harawira P, et al. E-cadherin germline mutations in familial gastric cancer. Nature. 1998;392(6674):402–5. https://doi.org/10.1038/32918.
2. Oliveira C, Senz J, Kaurah P, Pinheiro H, Sanges R, Haegert A, et al. Germline CDH1 deletions in hereditary diffuse gastric cancer families. Hum Mol Genet. 2009;18(9):1545–55. https://doi.org/10.1093/hmg/ddp046.
3. Garcia-Pelaez J, Barbosa-Matos R, Lobo S, Dias A, Garrido L, Castedo S, et al. Genotype-first approach to identify associations between CDH1 germline variants and cancer phenotypes: a multicentre study by the European reference network on genetic

TGF-β. Mol Biol Rep. 2020;47(1):433–41. https://doi.org/10.1007/s11033-019-05146-2.
64. Lee J, Robinson ME, Ma N, Artadji D, Ahmed MA, Xiao G, et al. Nature. 2020;588(7838):491–7. https://doi.org/10.1038/s41586-020-2884-6.
65. Li D, Peng Z, Tang H, Wei P, Kong X, Yan D, et al. KLF4-mediated negative regulation of IFITM3 expression plays a critical role in colon cancer pathogenesis. Clin Cancer Res. 2011;17(11):3558–68. https://doi.org/10.1158/1078-0432.Ccr-10-2729.
66. Daniel-Carmi V, Makovitzki-Avraham E, Reuven EM, Goldstein I, Zilkha N, Rotter V, et al. The human 1–8D gene (IFITM2) is a novel p53 independent pro-apoptotic gene. Int J Cancer. 2009;125(12):2810–9. https://doi.org/10.1002/ijc.24669.
67. Bonni A, Brunet A, West AE, Datta SR, Takasu MA, Greenberg ME. Cell survival promoted by the Ras-MAPK signaling pathway by transcription-dependent and -independent mechanisms. Science. 1999;286(5443):1358–62. https://doi.org/10.1126/science.286.5443.1358.
68. Hou F, Shi DB, Chen YQ, Gao P. Human epidermal growth factor receptor-2 promotes invasion and metastasis in gastric cancer by activating mitogen-activated protein kinase signaling. Appl Immunohistochem Mol Morphol. 2019;27(7):529–34. https://doi.org/10.1097/pai.0000000000000672.
69. Ohtsuka J, Oshima H, Ezawa I, Abe R, Oshima M, Ohki R. Functional loss of p53 cooperates with the in vivo microenvironment to promote malignant progression of gastric cancers. Sci Rep. 2018;8(1):2291. https://doi.org/10.1038/s41598-018-20572-1.
70. Busuttil RA, Zapparoli GV, Haupt S, Fennell C, Wong SQ, Pang JM, et al. Role of p53 in the progression of gastric cancer. Oncotarget. 2014;5(23):12016–26. https://doi.org/10.18632/oncotarget.2434.
71. Kastan MB, Onyekwere O, Sidransky D, Vogelstein B, Craig RW. Participation of p53 protein in the cellular response to DNA damage. Cancer Res. 1991;51(23 Pt 1):6304–11.

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Authors and Affiliations

Celina São José[1,2,3] · José Garcia-Pelaez[1,2,3] · Marta Ferreira[1,2,4] · Oscar Arrieta[5] · Ana André[1,2] · Nelson Martins[1,2,6] · Samantha Solís[7] · Braulio Martínez-Benítez[8] · María Luisa Ordóñez-Sánchez[7] · Maribel Rodríguez-Torres[7] · Anna K. Sommer[9] · Iris B. A. W. te Paske[10] · Carlos Caldas[11,12,13] · Marc Tischkowitz[14] · Maria Teresa Tusié[7] · Solve-RD DITF-GENTURIS · Nicoline Hoogerbrugge[10] · German Demidov[15] · Richarda M. de Voer[10] · Steve Laurie[16] · Carla Oliveira[1,2,17]

✉ Carla Oliveira
carlaol@i3s.up.pt

[1] i3S–Instituto de Investigação e Inovação em Saúde, Rua Alfredo Allen, 208, 4200-135 Porto, Portugal

[2] IPATIMUP–Instituto de Patologia e Imunologia Molecular da Universidade do Porto, Porto, Portugal

[3] Doctoral Programme in Biomedicine, Faculty of Medicine, University of Porto, Porto, Portugal

[4] Department Computer Science Faculty of Science, University of Porto, Porto, Portugal

[5] Thoracic Oncology Unit, Department of Thoracic Oncology, Instituto Nacional de Cancerología, Mexico City, Mexico

[6] Master Programme in Molecular Medicine and Oncology, Faculty of Medicine, University of Porto, Porto, Portugal

[7] INCMNSZ/Instituto de Investigaciones Biomédicas, Unidad de Biología Molecular y Medicina Genómica Instituto Nacional de Ciencias Médicas y Nutrición Salvador Zubirán, UNAM Mexico City, Mexico

[8] Pathology Department, Instituto Nacional de Ciencias Médicas y Nutrición Salvador Zubirán, INCMNSZ Mexico City, Mexico

[9] Institute of Human Genetics, Medical Faculty, University of Bonn, Bonn, Germany

[10] Department of Human Genetics, Radboud University Medical Center, Radboud Institute for Molecular Life Sciences, Nijmegen, The Netherlands

[11] Cancer Research UK Cambridge Institute, University of Cambridge, Li Ka Shing Centre, Cambridge, UK

[12] Department of Oncology, University of Cambridge, Cambridge, UK

[13] Cambridge Experimental Cancer Medicine Centre (ECMC), CRUK Cambridge Centre, NIHR Cambridge Biomedical Research Centre, University of Cambridge and Cambridge University Hospitals NHS Foundation Trust, Cambridge, UK

[14] Department of Medical Genetics, National Institute for Health Research Cambridge Biomedical Research Centre, University of Cambridge, Cambridge, UK

[15] Institute of Medical Genetics and Applied Genomics, Tübingen, Germany

[16] The Barcelona Institute of Science and Technology, CNAG-CRG, Centre for Genomic Regulation (CRG), Barcelona, Spain

[17] FMUP–Faculty of Medicine of the University of Porto, Porto, Portugal

Springer

Appendix C

Genetic and Epigenetic Alterations of *CDH1* Regulatory Regions in Hereditary and Sporadic Gastric Cancer

Gianluca Tedaldi, Chiara Molinari, Celina São José, Rita Barbosa-Matos, Ana André, Rita Danesi, Valentina Arcangeli, Mila Ravegnani, Luca Saragoni, Paolo Morgagni, Francesca Rebuzzi, Matteo Canale, Sara Pignatta, Elisa Ferracci, Giovanni Martinelli, Guglielmina Nadia Ranzani, Carla Oliveira, Daniele Calistri, Paola Ulivi

Pharmaceuticals

2021

DOI: 10.3390/ph14050457

pharmaceuticals

MDPI

Article

Genetic and Epigenetic Alterations of *CDH1* Regulatory Regions in Hereditary and Sporadic Gastric Cancer

Gianluca Tedaldi [1,*], Chiara Molinari [1], Celina São José [2,3,4], Rita Barbosa-Matos [2,3,5], Ana André [2,3], Rita Danesi [6], Valentina Arcangeli [6], Mila Ravegnani [6], Luca Saragoni [7], Paolo Morgagni [8], Francesca Rebuzzi [1], Matteo Canale [1], Sara Pignatta [1], Elisa Ferracci [9], Giovanni Martinelli [10], Guglielmina Nadia Ranzani [11], Carla Oliveira [2,3,12,*], Daniele Calistri [1,†] and Paola Ulivi [1,†]

1 Biosciences Laboratory, IRCCS Istituto Romagnolo per lo Studio dei Tumori (IRST) "Dino Amadori", 47014 Meldola, Italy; chiara.molinari@irst.emr.it (C.M.); francesca.rebuzzi@irst.emr.it (F.R.); matteo.canale@irst.emr.it (M.C.); sara.pignatta@irst.emr.it (S.P.); daniele.calistri@irst.emr.it (D.C.); paola.ulivi@irst.emr.it (P.U.)
2 i3S, Institute for Research and Innovation in Health, University of Porto, 4200-135 Porto, Portugal; cjose@ipatimup.pt (C.S.J.); amatos@ipatimup.pt (R.B.-M.); aandre@ipatimup.pt (A.A.)
3 IPATIMUP, Institute of Molecular Pathology and Immunology, University of Porto, 4200-135 Porto, Portugal
4 Doctoral Programme in Biomedicine, Faculty of Medicine, University of Porto, 4200-319 Porto, Portugal
5 Doctoral Programme on Cellular and Molecular Biotechnology Applied to Health Sciences (Biotech Health), ICBAS-Institute of Biomedical Sciences Abel Salazar, University of Porto, 4050-313 Porto, Portugal
6 Romagna Cancer Registry, IRCCS Istituto Romagnolo per lo Studio dei Tumori (IRST) "Dino Amadori", 47014 Meldola, Italy; rita.danesi@irst.emr.it (R.D.); valentina.arcangeli@irst.emr.it (V.A.); mila.ravegnani@irst.emr.it (M.R.)
7 Pathology Unit, Morgagni-Pierantoni Hospital, 47121 Forlì, Italy; luca.saragoni@auslromagna.it
8 Department of General Surgery, Morgagni-Pierantoni Hospital, 47121 Forlì, Italy; paolo.morgagni@auslromagna.it
9 Laboratory of Stem Cells and Cancer Genomics, University of Trento, 38123 Trento, Italy; elisa.ferracci@unitn.it
10 Department of Medical Oncology, IRCCS Istituto Romagnolo per lo Studio dei Tumori (IRST) "Dino Amadori", 47014 Meldola, Italy; giovanni.martinelli@irst.emr.it
11 Department of Biology and Biotechnology, University of Pavia, 27100 Pavia, Italy; guglielmina.ranzani@unipv.it
12 Faculty of Medicine of the University of Porto (FMUP), 4200-319 Porto, Portugal
* Correspondence: gianluca.tedaldi@irst.emr.it (G.T.); carlaol@ipatimup.pt (C.O.); Tel.: +39-0543-739232 (G.T.); +351-225570785 (C.O.)
† These authors contributed equally to this work.

Abstract: E-cadherin is a key player in gastric cancer (GC) and germline alterations of *CDH1*, its encoding gene, are responsible for Hereditary Diffuse Gastric Cancer (HDGC) syndrome. This study aimed at elucidating the role of genetic variants and DNA methylation of *CDH1* promoter and enhancers in the regulation of gene expression. For this purpose, we analyzed genetic variants of the *CDH1* gene through Next-Generation Sequencing (NGS) in a series of GC cell lines (NCI-N87, KATO-III, SNU-1, SNU-5, GK2, AKG, KKP) and the corresponding *CDH1* expression levels. By bisulfite genomic sequencing, we analyzed the methylation status of *CDH1* regulatory regions in 8 GC cell lines, in a series of 13 sporadic GC tissues and in a group of 20 HDGC *CDH1*-negative patients and 6 healthy controls. The NGS analysis on *CDH1* coding and regulatory regions detected genetic alterations in 3 out of 5 GC cell lines lacking functional E-cadherin. *CDH1* regulatory regions showed different methylation patterns in patients and controls, GC cell lines and GC tissues, expressing different E-cadherin levels. Our results showed that alterations in terms of genetic variants and DNA methylation patterns of both promoter and enhancers are associated with *CDH1* expression levels and have a role in its regulation.

Keywords: gastric cancer; *CDH1* gene; DNA methylation; regulatory regions; genetic predisposition; Next-Generation Sequencing

check for updates

Citation: Tedaldi, G.; Molinari, C.; São José, C.; Barbosa-Matos, R.; André, A.; Danesi, R.; Arcangeli, V.; Ravegnani, M.; Saragoni, L.; Morgagni, P.; et al. Genetic and Epigenetic Alterations of *CDH1* Regulatory Regions in Hereditary and Sporadic Gastric Cancer. *Pharmaceuticals* **2021**, *14*, 457. https://doi.org/10.3390/ph14050457

Academic Editor: Lawrence Marc Pfeffer

Received: 31 March 2021
Accepted: 7 May 2021
Published: 12 May 2021

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Copyright: © 2021 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (https://creativecommons.org/licenses/by/4.0/).

1. Introduction

In 2020, gastric cancer (GC) ranked fifth for incidence and fourth for mortality, worldwide [1]. Although GC is commonly sporadic, about 10–20% of cases show familial clustering and 1–3% of cases can be considered hereditary [2]. Sporadic GC occurs at an older age, is mainly associated with intestinal histotype (IGC) [3] and is related to environmental risk factors, such as microbial infection and diet [4]. Hereditary GC is characterized by early-onset, is mainly of the diffuse histotype (DGC) and is associated with alterations in different predisposition genes [5]. Nevertheless, with the exception of the inherited familial syndromes, the observed familial clustering of cancer is not well explained and may occur due to an inherited genetic susceptibility, together with shared lifestyle and environmental factors [6].

The main gene involved in hereditary GC is *CDH1*, encoding the E-cadherin protein [7], whose germline mutations are responsible for Hereditary Diffuse Gastric Cancer (HDGC) [8], a genetic syndrome linked to a lifetime GC risk of 42–70% in male carriers and 33–56% in female carriers, who also have a 39–55% risk of lobular breast cancer (LBC) [9,10]. The oncogenesis in HDGC syndrome follows the "two-hit model" [11], in which the presence of a germline mutation of the *CDH1* gene (first hit) is followed by the inactivation of the other allele (second hit) by mutations or deletions (loss of heterozygosity-LOH). However, it has been shown that *CDH1* promoter hypermethylation represents the main cause of inactivation of the wild-type allele in HDGC tumors [12–14].

The *CDH1* gene is a driver of carcinogenesis also in sporadic GC of both diffuse and intestinal histotypes [15,16] and *CDH1* epigenetic alterations, such as promoter hypermethylation and histone modifications, have also been detected in most sporadic GCs as a cause of E-cadherin impairment [15,17–20].

Given the importance of *CDH1* in the oncogenesis of hereditary and sporadic GC, different studies were carried out to understand the regulatory mechanisms of *CDH1* gene expression. In particular, the studies focused on the 65kb-region of *CDH1* intron 2, which is characterized by the presence of cis-regulatory sequences necessary for transcription and, therefore, for the regulation of gene expression [21–23].

Moreover, not only a large body of data supports the notion that DNA methylation is crucial in regulating gene expression, but recent findings also point to a link between this mechanism and cancer predisposition. Accordingly, studies aimed at linking genomic regions associated with cancer susceptibility and expression of genes involved in cancer development have already shown that promoters' and enhancers' sequence polymorphisms and methylation status can contribute to cancer risk [24,25].

The present study aimed to perform a molecular characterization of genetic and epigenetic alterations of *CDH1* promoter and enhancer sequences to deeply understand the mechanisms that regulate gene expression and to provide the rationale for investigating *CDH1* methylation in *CDH1*-negative patients with suspected genetic predisposition to GC.

2. Results

2.1. CDH1 Variant Analysis on GC Cell Lines

The *CDH1* variant analysis was performed by Next-Generation Sequencing (NGS) on seven GC cell lines (NCI-N87, KATO-III, SNU-5, SNU-1, AKG, GK2, KKP).

Results showed the presence of a *CDH1* pathogenic variant in three cell lines (KATO-III, SNU-5 and KKP), whereas in the other four cell lines no *CDH1* variants with relevant significance were detected (Supplementary Table S1).

2.2. Identification of CDH1 Regulatory Regions

The promoter and the enhancers of the *CDH1* gene were identified through the UCSC Genome Browser. The *CDH1* promoter is located in a CpG island of 1310 bp covering exons 1 and 2 (Figure 1). The FANTOM5 tool of the UCSC Genome Browser identified seven enhancers in the *CDH1* locus, one located 4724 bp upstream from the *CDH1* transcription start site (TSS), and six located within intron 2 (Figure 1).

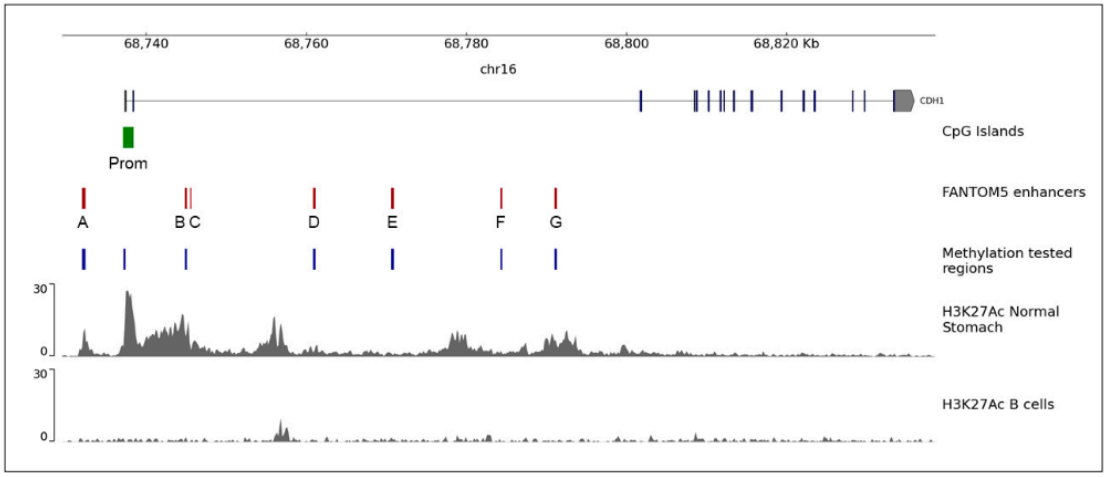


Figure 1. Scheme of the *CDH1* gene (chr16) with the CpG island represented by the green bar and enhancers by red bars. Blue bars represent the regions tested for methylation in the present work. The lower panels represent H3K27Ac, a chromatin mark associated with transcription activation [26], in normal stomach and B cells, respectively. The figure was generated with pyGenomeTracks [27].

2.3. CDH1 Methylation Analysis on DNA from GC Cell Lines, Peripheral Blood and Gastric Tissues

The methylation analysis of *CDH1* promoter and enhancers was performed on eight GC cell lines (NCI-N87, KATO-III, SNU-5, SNU-1, AKG, GK2, KKP and MKN-74), on germline DNA from peripheral blood of 6 healthy individuals and 20 HDGC *CDH1*-negative patients (15 with DGC and 5 with LBC) and on gastric normal and tumor tissues from 13 sporadic GC patients (10 with IGC, 1 with DGC, 1 with mucinous GC and 1 with mixed GC). We analyzed the methylation status of CpG sites located in the core of the *CDH1* promoter and in six enhancers of *CDH1* gene (Figure 1 and Supplementary Table S2).

The methylation analysis revealed different methylation patterns in the eight GC cell lines, with different CpG sites being methylated along promoter and enhancer sequences, in particular enhancers A, B and D (Figure 2).

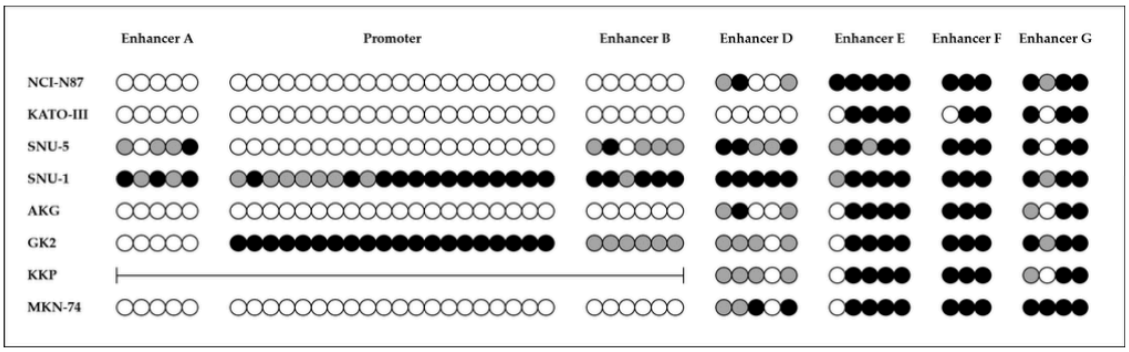


Figure 2. Methylation status of *CDH1* promoter and enhancers of eight GC cell lines expressing different levels of *CDH1*. ○ non-methylated CpG site; ◍ partially methylated CpG site; ● fully methylated CpG site; ⊢ deletion.

The methylation analysis performed on the germline DNA derived from blood of 6 healthy controls and of 20 HDGC *CDH1*-negative patients (15 with DGC and 5 with LBC) revealed a strongly conserved methylation pattern along promoter and enhancer

sequences, with absence of methylation in the promoter and almost full methylation in enhancer B in all patients and controls (Figure 3 and Supplementary Figures S1 and S2).

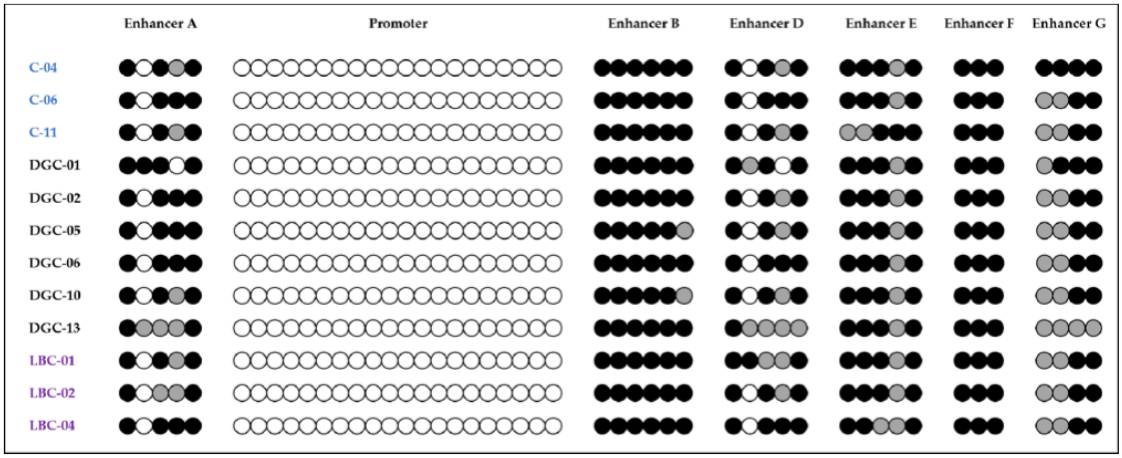


Figure 3. Methylation analysis of *CDH1* promoter and enhancers on a selection of DNA samples from peripheral blood of three healthy individuals (in blue) and nine HDGC *CDH1*-negative patients (six DGC patients in black and three LBC patients in violet). The complete case series is reported in Supplementary Figures S1 and S2. ○ non-methylated CpG site; ◍ partially methylated CpG site; ● fully methylated CpG site.

The methylation analysis performed on DNA from gastric normal and tumor tissues of 13 patients with sporadic GC revealed differences in the methylation pattern of normal vs. tumor tissues, with an increased methylation in particular in enhancers A, B and E of tumor tissues (Figure 4 and Supplementary Figure S3).

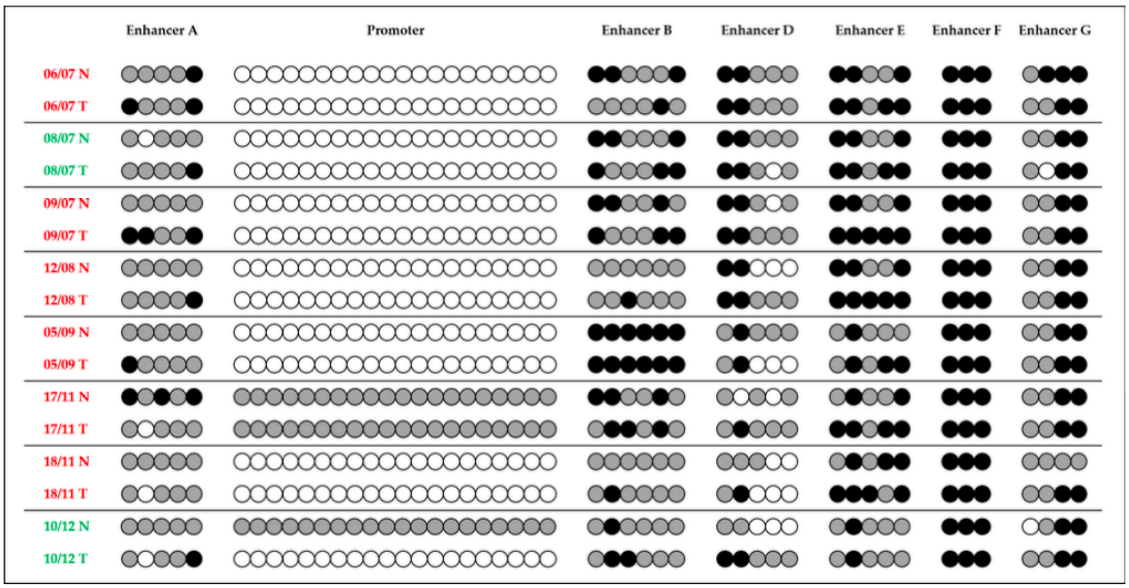


Figure 4. Methylation analysis of *CDH1* promoter and enhancers on DNA samples from gastric normal (N) and tumor (T) tissues from eight IGC patients (GC tissues with *CDH1* downregulation in red and GC tissues with *CDH1* upregulation in green). ○ non-methylated CpG site; ◍ partially methylated CpG site; ● fully methylated CpG site.

2.4. CDH1 Expression Analysis in GC Cell Lines and Gastric Tissues

The *CDH1* expression analysis was performed by quantitative PCR (qPCR) on seven GC cell lines (NCI-N87, KATO-III, SNU-5, SNU-1, AKG, GK2, KKP) and on gastric normal and tumor tissues of eight sporadic IGC patients.

Results from GC cells revealed a reduced *CDH1* expression in four of them (SNU-1, SNU-5, GK2 and KKP) and a "normal" *CDH1* expression in the other three cell lines (NCI-N87, KATO-III and AKG) (Supplementary Table S3).

Results from gastric normal vs. tumor tissues revealed *CDH1* downregulation in six IGC tissues (06/07, 09/07, 12/08, 05/09, 17/11 and 18/11) and *CDH1* upregulation in two IGC tissues (08/07 and 10/12) (Table 1).

Table 1. Results of *CDH1* expression on eight gastric normal (N) and tumor (T) tissues.

Gastric Tissue	$2^{-\Delta Ct}$	Fold Change *CDH1*
06/07 N	2.412	0.056
06/07 T	0.135	
08/07 N	0.157	2.085
08/07 T	0.328	
09/07 N	0.210	0.279
09/07 T	0.059	
12/08 N	2.603	0.053
12/08 T	0.138	
05/09 N	1.338	0.028
05/09 T	0.037	
17/11 N	1.591	0.055
17/11 T	0.087	
18/11 N	2.297	0.042
18/11 T	0.097	
10/12 N	0.056	14.026
10/12 T	0.779	

3. Discussion

GC is one of the most aggressive cancers, with high mortality rates worldwide. For this reason, the prevention and treatment of GC represent the main topics of interest in the research on this disease. One of the most important players in carcinogenesis of both hereditary and sporadic GC is the *CDH1* gene, encoding E-cadherin, a component of the adherent junctions between epithelial cells with a tumor suppressor activity [28].

In the present study, we evaluated the role of genetic and epigenetic alterations of the *CDH1* gene, in particular the methylation of promoter and enhancers, in the regulation of gene expression and in the predisposition to the development of GC.

Indeed, studies aimed at linking genomic regions associated with cancer susceptibility and expression of genes involved in cancer development have already shown that promoters' and enhancers' sequence polymorphisms and methylation status can contribute to cancer risk [24,25].

This evidence prompted us to test this model in the *CDH1* gene on GC cell lines and different tissues from patients with hereditary and sporadic GC.

Initially, we identified regulatory regions of the *CDH1* gene using bioinformatics tools for the identification of genomic regions with enhancer properties. This analysis revealed that, besides the promoter, there are seven regions in the *CDH1* gene with regulatory functions and six of them are located in intron 2, which has been previously identified as a cis-regulatory element with a pivotal role in controlling the gene expression [22].

Consequently, we explored the landscape of the *CDH1* gene in terms of DNA mutation and methylation on a series of eight GC cell lines, correlating these results to the gene

expression. In particular, we identified *CDH1* pathogenic variants in three GC cell lines (KATO-III, SNU-5, KKP). Interestingly, SNU-5 and KKP cell lines also had a significant downregulation of *CDH1* expression. KKP cell lines showed a homozygous deletion of a large region including the promoter (with exons 1 and 2) and enhancers A and B, located upstream and downstream, respectively, while the methylation status of the other enhancers (D–G) did not show significant differences compared to GC cell lines expressing *CDH1*. The SNU-5 cell line showed a *CDH1* variant immediately after exon 5 affecting splicing and partial methylation of enhancers A, B and D (absent in the promoter). Of note, these two cell lines are derived from ascitic effusion of patients with a poorly differentiated GC. On the other hand, the third *CDH1*-mutated GC cell line, KATO-III, despite the presence of a variant in exon 7 with a deleterious effect on splicing [29], showed neither decreased *CDH1* expression nor increased promoter/enhancer methylation, suggesting that this variant does not alter the transcription, but rather the protein function. Indeed, KATO-III cell line is derived from a DGC with signet ring cells, a typical characteristic of GCs of HDGC patients.

The methylation analysis on the *CDH1* promoter and enhancers revealed increased methylation levels also in SNU-1 and GK2 cell lines, expressing low *CDH1* levels but lacking *CDH1* variants. Intriguingly, GK2 cells show full methylation of the *CDH1* promoter and partial methylation of enhancers B and D (but absent in enhancer A). On the other hand, SNU-1 cells show medium-high levels of methylation in promoter and enhancers A, B and D, compared to GC cell lines expressing *CDH1*. Of note, these two cell lines are both derived from poorly differentiated GCs.

The last three cell lines (NCI-N87, AKG and MKN-74) showed methylation neither in *CDH1* promoter nor in enhancers A and B. Indeed, these GC cell lines have a "normal" expression of *CDH1* gene, as demonstrated in the present work (for NCI-N87 and AKG) or previously reported (for MKN-74) [30,31]. Of note, these three GC cell lines are all derived from IGCs well or moderately differentiated.

Overall, the methylation results obtained in GC cell lines suggest that, besides the promoter, the methylation of CpG sites located in the *CDH1* enhancers correlates with the gene expression (Table 2).

Table 2. Summary of the results obtained on GC cell lines from *CDH1* sequencing, methylation and expression analyses.

Cell Line	*CDH1* Status [1]	*CDH1* Promoter Methylation [2]	*CDH1* Enhancer Methylation [2]				*CDH1* Expression
			A	B	D	E/F/G	
NCI-N87	wt	–	–	–	+	+	normal
KATO-III	mut	–	–	–	–	+	normal
SNU-5	mut	–	+	+	+	+	low
SNU-1	wt	+	+	+	+	+	low
AKG	wt	–	–	–	+	+	normal
GK2	wt	+	–	+	+	+	low
KKP [3]	mut	ND	ND	ND	+	+	low
MKN-74 [4]	wt	–	–	–	+	+	normal

[1] wt: *CDH1* wild-type; mut: *CDH1* pathogenic variant. [2] –: absence of methylation; +: presence of partial or full methylation. [3] The *CDH1* promoter and enhancers A/B are deleted in KKP cell line consequently the methylation analysis for these regions was not detectable (ND). [4] *CDH1* sequencing and expression analyses for MKN-74 cell line have been previously reported [30,31].

These results prompted us to analyze the *CDH1* promoter/enhancer methylation in the germline DNA of HDGC *CDH1*-negative patients, in order to identify a signature predisposing GC development in these patients. Indeed, an increased methylation of *CDH1* promoter was previously identified in an HDGC *CDH1*-negative family [32].

To study the methylation of *CDH1* promoter and enhancers in the germline DNA, we analyzed the methylation on DNA extracted from peripheral blood on a series of 6 healthy individuals and 20 HDGC *CDH1*-negative patients (15 with DGC and 5 with LBC). Unexpectedly, this analysis showed a strongly conserved methylation pattern in *CDH1* promoter/enhancers both in the healthy controls and the HDGC patients. In particular,

enhancer B showed a strong methylation pattern, together with enhancers A and D that were characterized by the presence of CpG sites with partial or full methylation.

E-cadherin, being a component of cell junctions, is not expressed in blood cells but the absence of methylation in the promoter suggests that another mechanism is responsible for *CDH1* downregulation. Our results highlight that, in blood cells, the methylation of *CDH1* enhancers could be responsible for the lack of expression of the gene. On the other hand, the methylation analysis of promoter/enhancers did not prove to be a valid method to identify individuals with a genetic predisposition to GC. Although DNA methylation is generally not heritable, alterations of the mechanisms underlying DNA methylation can, instead, be inherited, for example, in the form of germline mutations in genes involved in the DNA methylation/demethylation processes. However, considering our results, this event seems not to be a frequent cause of HDGC syndrome.

Finally, we tested the methylation levels of *CDH1* promoter/enhancers on a series of sporadic GCs. Results obtained from the analysis of eight IGCs expressing different levels of *CDH1* gene showed the presence of some differences in the methylation of *CDH1* promoter/enhancers between normal and tumor tissues. One of the two tumors (10/12) with upregulation of *CDH1* gene showed a decrease in the promoter methylation compared to the normal tissue, whereas the other (08/07) was characterized by a methylation status substantially comparable with the normal tissue, with some CpG sites of enhancers D and G less methylated in the tumor tissue.

Regarding the six tumors with downregulation of *CDH1* gene, they were characterized by the absence of promoter methylation in all normal and tumor tissues, with the exception of one case (17/11) that showed partial methylation of all the CpG sites analyzed in both normal and tumor tissues. On the other hand, the methylation status of *CDH1* enhancers showed slight differences between normal and tumor tissues, with an increased methylation especially in the CpG sites of enhancers A, B and E in almost all the cases. An additional series of five patients with sporadic GC was studied for *CDH1* promoter, enhancer A and B methylation (Supplementary Figure S3) showing a similar pattern.

This result suggests that *CDH1* enhancer methylation could have a role in gene expression regulation also in GC of the intestinal histotype, in which E-cadherin impairment is a less frequent event [15]. Moreover, in IGC, the *CDH1* gene has been demonstrated to be regulated by other epigenetic mechanisms such as microRNA expression [16,33].

Taking into account all these results, we can conclude that the methylation status of *CDH1* enhancers correlates with the gene expression levels and seems to have an important role in the gene regulation. Particularly in blood cells, where promoter methylation is absent, the presence of a specific pattern of methylation in the enhancers suggests that it could be a repression mechanism stronger than promoter methylation. This provides the rationale to investigate *CDH1* promoter/enhancer methylation in the gastric tissue of patients with suspected genetic predisposition to GC, since it could be an early event of the carcinogenesis process in HDGC *CDH1*-negative patients.

Moreover, *CDH1* hypermethylation could be a potential novel drug target for developing personalized therapies [34,35]. However, further studies on a larger number of patients with GC of different histotypes are necessary to confirm our results and to deeply understand the role of enhancer methylation in the regulation of *CDH1* gene.

4. Materials and Methods

4.1. Cell Lines

The study was performed on four commercial GC cell lines (SNU-1 ATCC Cat# CRL-5971, RRID:CVCL_0099, SNU-5 ATCC Cat# CRL-5973, RRID:CVCL_0078, KATO-III ATCC Cat# HTB-103, RRID:CVCL_0371, NCI-N87 ATCC Cat# CRL-5822, RRID:CVCL_1603) purchased from ATCC (American Type Culture Collection, Manassas, VA, USA), three GC cell lines (GK2, AKG, KKP) established and characterized at IRST laboratory, and one GC cell line (MKN-74) provided by the team of C. Oliveira, a co-author from the Institute of Molecular Pathology and Immunology of University of Porto (IPATIMUP, Portugal).

The characteristics of the GC cell lines used in the present work are summarized in the Supplementary Table S4.

All the cell lines were maintained as a monolayer, incubated at 37 °C with 5% CO_2 and subcultured weekly. For SNU-1, NCI-N87, and MKN-74, the culture medium was composed of RPMI-1640 Medium (ATCC) supplemented with 10% fetal bovine serum (FBS) (Euroclone, Milan, Italy). For SNU-5 and KATO-III, the culture medium was Iscove's Modified Dulbecco's Medium (ATCC) supplemented with 20% FBS. For AKG, GK2 and KKP, the culture medium was composed of DMEM/Ham's F12 (1:1) supplemented with FBS (10%), glutamine (2 mM), non-essential amino acids (1%) and insulin (10 mg/mL) (Sigma-Aldrich, St. Louis, MO, USA). Cell lines were routinely tested for Mycoplasma using MycoAlert Mycoplasma Detection Kit (Lonza, Basel, Switzerland).

4.2. Blood Samples

The patients included in the study were selected by the Genetic Counseling service of the Area Vasta Romagna from 2010 to 2018 because they had a family and personal history of DGC and/or LBC.

In particular, the selection of patients was performed according to the first four criteria established by the International Gastric Cancer Linkage Consortium (IGCLC) in the 2015 HDGC guidelines, which must be valid among the first- and second-degree relatives: 2 GC cases regardless of age with at least 1 confirmed DGC (I); 1 case of DGC < 40 years (II); personal or family history of DGC and LBC with 1 diagnosed < 50 years (III); bilateral LBC or family history of 2 or more cases of LBC < 50 years (IV) [36].

All 20 patients underwent molecular testing for point mutations and copy number variations (CNVs) of the *CDH1* gene using Next-Generation Sequencing (NGS) and Multiplex Ligation-dependent Probe Amplification (MLPA), as reported in our previous work [37]. Of the 20 patients, 15 had developed DGC (range 22–39 years) and 5 had LBC (range 35–48 years). Patients' characteristics are reported in Supplementary Table S5.

After the subscription of the informed consent, we collected the peripheral blood of these 20 patients and of 6 healthy controls.

4.3. Gastric Tissue Samples

Thirteen patients with sporadic GC submitted to surgical resection and classified by an expert pathologist, according to Lauren's classification [3], were included in this study. Surgical samples were immediately cryopreserved after resection and fresh-frozen (FF) tumor tissue and matched normal gastric epithelium samples were stored as previously described [33]. The patients' characteristics are summarized in Supplementary Table S6.

4.4. Nucleic Acid Extraction

The DNA was extracted from GC cell lines, peripheral blood samples and FF tumor and normal tissues using the QIAamp DNA Mini Kit (Qiagen, Hilden, Germany), according to manufacturer's instructions. DNA was quantified by Qubit 4 Fluorometer (Thermo Fisher Scientific, Waltham, MA, USA).

Total RNA extraction was performed from GC cell lines and from FF tumor and normal tissues using TRIzol (Invitrogen, Carlsbad, CA, USA) in accordance with manufacturer's instructions. Four micrograms of extracted RNA were treated with DNase and purified by using RNeasy MinElute Cleanup kit (Qiagen, Hilden, Germany), RNA was quantified by spectrophotometer Nanodrop-ND-1000 (Thermo Fisher Scientific).

4.5. Next-Generation Sequencing

The DNA libraries for the NGS analysis were prepared with Nextera Flex for Enrichment (Illumina, San Diego, CA, USA) and a custom probe panel (Integrated DNA Technologies, Coralville, IA, USA), and run on the Miseq sequencer (Illumina). Results were analyzed with Miseq Reporter software (Illumina) and further processed with a customized bioinformatics pipeline.

4.6. Identification of CDH1 Regulatory Regions

The promoter and the enhancers of the *CDH1* gene were identified through the UCSC Genome Browser [38] with the FANTOM5 tool [39,40] in the 2018 version. The enhancers were named from A to G starting from the most upstream one. The coordinates of *CDH1* promoter and enhancers are reported in the Supplementary Table S2.

4.7. DNA Methylation Analysis

DNA samples were treated with bisulfite using the Epitect Bisulfite kit (Qiagen, Hilden, Germany),

CDH1 promoter and enhancers were amplified by standard PCR with the Multiplex PCR kit (Qiagen, Hilden, Germany), and the PCR products were sequenced by the BigDye Terminator v.3.1 cycle sequencing kit (Thermo Fisher Scientific) on an ABI-3130 Genetic analyzer (Applied Biosystems, Foster City, CA, USA), with the exception of enhancer C that was excluded from the analysis because of the absence of CpG sites. The primers used for amplification and sequencing were designed on the in silico bisulfite converted sequence and are reported in Supplementary Table S7.

4.8. Expression Analysis

The retrotranscription of RNA to cDNA was performed on 1000 ng of total RNA using the SuperScript II Reverse Transcriptase, random primers, dNTP mix and RNaseOUT (Invitrogen, Carlsbad, CA, USA) according to manufacturer's instructions.

The *CDH1* expression analysis was performed by real-time PCR with TaqMan probes for *CDH1* gene as target (Hs01023894_m1, Thermo Fisher Scientific) and *B2M* gene as endogenous control for normalization (Hs00984230_m1, Thermo Fisher Scientific). The runs were performed on the 7500 Real-Time PCR System (Applied Biosystems).

Expression levels of the target genes were obtained by normalizing the results using the endogenous control *B2M*. For GC cell lines, the relative expression was quantified using the comparative $2^{-\Delta Ct}$ method. For the gastric tissues, the Fold Change (FC) values in gene expression were calculated using the $2^{-\Delta\Delta Ct}$ method [41,42]. An FC ≤ 0.50 and ≥ 1.5 were used as cut-off for downregulation and upregulation, respectively.

5. Conclusions

In conclusion, the results we obtained from the methylation analysis of *CDH1* promoter and enhancers demonstrate the existence of different methylation patterns within the regulatory regions of the gene and highlight that the methylation in *CDH1* enhancers could have an important role in the regulation of gene expression.

Supplementary Materials: The following are available online at https://www.mdpi.com/article/10.3390/ph14050457/s1, Figure S1: Methylation analysis of *CDH1* promoter and enhancers from peripheral blood of 6 healthy individuals, Figure S2: Methylation analysis of *CDH1* promoter and enhancers from peripheral blood of 20 HDGC *CDH1*-negative patients, Figure S3: Methylation analysis of *CDH1* promoter and enhancers A and B on DNA samples from gastric normal (N) and tumor (T) tissues from 5 patients with sporadic GC, Table S1: Results of the *CDH1* variant analysis on 7 GC cell lines, Table S2: Genomic coordinates of *CDH1* regulatory elements identified by FANTOM5, Table S3: Results of *CDH1* expression on 7 different GC cell lines, Table S4: Characteristics of the 8 GC cell lines used in the study, Table S5: Clinical characteristics of the 20 HDGC *CDH1*-negative patients whose peripheral blood was analyzed, Table S6: Clinical characteristics of the 13 patients from whose gastric normal and tumor tissues were analyzed, Table S7: Sequences of the primers used for amplification and sequencing in the methylation analysis of *CDH1* regulatory regions on bisulfite-converted DNA.

Author Contributions: Conceptualization, D.C., G.N.R., and C.O.; methodology, C.M., C.S.J., R.B.-M., A.A., M.C., S.P.; validation, G.T., F.R. and E.F.; investigation, G.T., C.M., C.S.J., R.B.-M., A.A., F.R. and E.F.; resources, P.U. and C.O.; data curation, R.D., V.A., M.R., P.M., L.S.; writing—original draft preparation, G.T.; writing—review and editing, P.U., D.C., C.O, G.N.R.; visualization, G.T., C.S.J., R.B.-M.; supervision, P.U., D.C., C.O, G.N.R., G.M.; project administration, P.U. and D.C. All authors have read and agreed to the published version of the manuscript.

Funding: This research and its authors were funded by IRCCS IRST (G.T., C.M., R.D. V.A., M.R., F.R., M.C., S.P., G.M., D.C., P.U.) and by FEDER-Fundo Europeu de Desenvolvimento Regional funds through the COMPETE 2020–Operacional Programme for Competitiveness and Internationalization (POCI), Portugal 2020, and by Portuguese funds through FCT–Fundação para a Ciência e a Tecnologia/Ministério da Ciência, Tecnologia e Inovação in the framework of the project "Institute for Research and Innovation in Health Sciences" (POCI-01-0145-FEDER-007274) (C.S.J., R.B.-M., A.A., C.O.). This work was also financed by the project NORTE-01-0145-FEDER-000029 (CANCER)-supported by Norte Portugal Regional Programme (NORTE 2020), under the PORTUGAL 2020 Partnership Agreement, through the European Regional Development Fund (ERDF)–project POCI-01-0145-FEDER-016390 (CancelStem) and PTDC/BTM-TEC/30164/2017 (3DChroMe), funded by ERDF, POCI and FCT.

Institutional Review Board Statement: The study was conducted according to the guidelines of the Declaration of Helsinki, and approved by the Institutional Ethics Committee of IRST IRCCS (IRSTB069, approved on 20 September 2017).

Informed Consent Statement: Informed consent was obtained from all subjects involved in the study.

Data Availability Statement: The data presented in this study are available on request from the corresponding authors.

Acknowledgments: C.S.J. acknowledges the support of the Faculty of Medicine from the University of Porto, specifically by the Doctoral Programme in Biomedicine and the FCT–Fundação para a Ciência e a Tecnologia for her Ph.D. fellowship (ref. SFRH/BD/140796/2018). R.B.-M. acknowledges the support of the Institute of Biomedical Sciences Abel Salazar from the University of Porto, specifically by the Doctoral Programme BiotechHealth and the FCT–Fundação para a Ciência e a Tecnologia for her Ph.D. fellowship (ref. SFRH/BD/145132/2019).

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Sung, H.; Ferlay, J.; Siegel, R.L.; Laversanne, M.; Soerjomataram, I.; Jemal, A.; Bray, F. Global cancer statistics 2020: Globocan estimates of incidence and mortality worldwide for 36 cancers in 185 countries. CA. *Cancer J. Clin.* **2021**. [CrossRef]
2. Oliveira, C.; Pinheiro, H.; Figueiredo, J.; Seruca, R.; Carneiro, F. Familial gastric cancer: Genetic susceptibility, pathology, and implications for management. *Lancet Oncol.* **2015**, *16*, e60–e70. [CrossRef]
3. Lauren, P. The two histological main types of gastric carcinoma: Diffuse and so-called intestinal-type carcinoma. An attempt at a histo-clinical classification. *Acta Pathol. Microbiol. Scand.* **1965**, *64*, 31–49. [CrossRef]
4. Van Cutsem, E.; Sagaert, X.; Topal, B.; Haustermans, K.; Prenen, H. Gastric cancer. *Lancet* **2016**, *388*, 2654–2664. [CrossRef]
5. van der Post, R.S.; Oliveira, C.; Guilford, P.; Carneiro, F. Hereditary gastric cancer: What's new? Update 2013–2018. *Fam. Cancer* **2019**, *18*, 363–367. [CrossRef] [PubMed]
6. Choi, Y.J.; Kim, N.; Jang, W.; Seo, B.; Oh, S.; Shin, C.M.; Lee, D.H.; Jung, H.C. Familial Clustering of Gastric Cancer: A Retrospective Study Based on the Number of First-Degree Relatives. *Medicine* **2016**, *95*, e3606. [CrossRef]
7. Berx, G.; Staes, K.; van Hengel, J.; Molemans, F.; Bussemakers, M.J.; van Bokhoven, A.; van Roy, F. Cloning and characterization of the human invasion suppressor gene E-cadherin (CDH1). *Genomics* **1995**, *26*, 281–289. [CrossRef]
8. Blair, V.R.; McLeod, M.; Carneiro, F.; Coit, D.G.; D'Addario, J.L.; van Dieren, J.M.; Harris, K.L.; Hoogerbrugge, N.; Oliveira, C.; van der Post, R.S.; et al. Hereditary diffuse gastric cancer: Updated clinical practice guidelines. *Lancet Oncol.* **2020**, *21*, e386–e397. [CrossRef]
9. Hansford, S.; Kaurah, P.; Li-Chang, H.; Woo, M.; Senz, J.; Pinheiro, H.; Schrader, K.A.; Schaeffer, D.F.; Shumansky, K.; Zogopoulos, G.; et al. Hereditary Diffuse Gastric Cancer Syndrome: CDH1 Mutations and Beyond. *JAMA Oncol.* **2015**, *1*, 23–32. [CrossRef]
10. Roberts, M.E.; Ranola, J.M.O.; Marshall, M.L.; Susswein, L.R.; Graceffo, S.; Bohnert, K.; Tsai, G.; Klein, R.T.; Hruska, K.S.; Shirts, B.H. Comparison of CDH1 Penetrance Estimates in Clinically Ascertained Families vs Families Ascertained for Multiple Gastric Cancers. *JAMA Oncol.* **2019**, *5*, 1325–1331. [CrossRef]

11. Guilford, P.; Hopkins, J.; Harraway, J.; McLeod, M.; McLeod, N.; Harawira, P.; Taite, H.; Scoular, R.; Miller, A.; Reeve, A.E. E-cadherin germline mutations in familial gastric cancer. *Nature* **1998**, *392*, 402–405. [CrossRef] [PubMed]
12. Grady, W.M.; Willis, J.; Guilford, P.J.; Dunbier, A.K.; Toro, T.T.; Lynch, H.; Wiesner, G.; Ferguson, K.; Eng, C.; Park, J.G.; et al. Methylation of the CDH1 promoter as the second genetic hit in hereditary diffuse gastric cancer. *Nat. Genet.* **2000**, *26*, 16–17. [CrossRef] [PubMed]
13. Machado, J.C.; Oliveira, C.; Carvalho, R.; Soares, P.; Berx, G.; Caldas, C.; Seruca, R.; Carneiro, F.; Sobrinho-Simões, M. E-cadherin gene (CDH1) promoter methylation as the second hit in sporadic diffuse gastric carcinoma. *Oncogene* **2001**, *20*, 1525–1528. [CrossRef] [PubMed]
14. Graziano, F.; Humar, B.; Guilford, P. The role of the E-cadherin gene (CDH1) in diffuse gastric cancer susceptibility: From the laboratory to clinical practice. *Ann. Oncol. Off. J. Eur. Soc. Med. Oncol.* **2003**, *14*, 1705–1713. [CrossRef]
15. Corso, G.; Carvalho, J.; Marrelli, D.; Vindigni, C.; Carvalho, B.; Seruca, R.; Roviello, F.; Oliveira, C. Somatic mutations and deletions of the E-cadherin gene predict poor survival of patients with gastric cancer. *J. Clin. Oncol.* **2013**, *31*, 868–875. [CrossRef]
16. Carvalho, J.; van Grieken, N.C.; Pereira, P.M.; Sousa, S.; Tijssen, M.; Buffart, T.E.; Diosdado, B.; Grabsch, H.; Santos, M.A.S.; Meijer, G.; et al. Lack of microRNA-101 causes E-cadherin functional deregulation through EZH2 up-regulation in intestinal gastric cancer. *J. Pathol.* **2012**, *228*, 31–44. [CrossRef]
17. Tamura, G.; Yin, J.; Wang, S.; Fleisher, A.S.; Zou, T.; Abraham, J.M.; Kong, D.; Smolinski, K.N.; Wilson, K.T.; James, S.P.; et al. E-Cadherin gene promoter hypermethylation in primary human gastric carcinomas. *J. Natl. Cancer Inst.* **2000**, *92*, 569–573. [CrossRef]
18. To, K.-F.; Leung, W.K.; Lee, T.-L.; Yu, J.; Tong, J.H.M.; Chan, M.W.Y.; Ng, E.K.W.; Chung, S.C.S.; Sung, J.J.Y. Promoter hypermethylation of tumor-related genes in gastric intestinal metaplasia of patients with and without gastric cancer. *Int. J. Cancer* **2002**, *102*, 623–628. [CrossRef]
19. Shokraii, F.; Moharrami, M.; Motamed, N.; Shahhoseini, M.; Totonchi, M.; Ezzatizadeh, V.; Firouzi, J.; Khosravani, P.; Ebrahimi, M. Histone Modification Marks Strongly Regulate CDH1 Promoter in Prostospheres as A Model of Prostate Cancer Stem Like Cells. *Cell J.* **2019**, *21*, 124–134.
20. Oikawa, T.; Otsuka, Y.; Onodera, Y.; Horikawa, M.; Handa, H.; Hashimoto, S.; Suzuki, Y.; Sabe, H. Necessity of p53-binding to the CDH1 locus for its expression defines two epithelial cell types differing in their integrity. *Sci. Rep.* **2018**, *8*, 1595. [CrossRef]
21. Stemmler, M.P.; Hecht, A.; Kinzel, B.; Kemler, R. Analysis of regulatory elements of E-cadherin with reporter gene constructs in transgenic mouse embryos. *Dev. Dyn.* **2003**, *227*, 238–245. [CrossRef] [PubMed]
22. Stemmler, M.P.; Hecht, A.; Kemler, R. E-cadherin intron 2 contains cis-regulatory elements essential for gene expression. *Development* **2005**, *132*, 965–976. [CrossRef]
23. Oliveira, P.; Sanges, R.; Huntsman, D.; Stupka, E.; Oliveira, C. Characterization of the intronic portion of cadherin superfamily members, common cancer orchestrators. *Eur. J. Hum. Genet.* **2012**, *20*, 878–883. [CrossRef]
24. Aran, D.; Hellman, A. DNA methylation of transcriptional enhancers and cancer predisposition. *Cell* **2013**, *154*, 11–13. [CrossRef] [PubMed]
25. Aran, D.; Sabato, S.; Hellman, A. DNA methylation of distal regulatory sites characterizes dysregulation of cancer genes. *Genome Biol.* **2013**, *14*, R21. [CrossRef] [PubMed]
26. Heintzman, N.D.; Hon, G.C.; Hawkins, R.D.; Kheradpour, P.; Stark, A.; Harp, L.F.; Ye, Z.; Lee, L.K.; Stuart, R.K.; Ching, C.W.; et al. Histone modifications at human enhancers reflect global cell-type-specific gene expression. *Nature* **2009**, *459*, 108–112. [CrossRef]
27. Lopez-Delisle, L.; Rabbani, L.; Wolff, J.; Bhardwaj, V.; Backofen, R.; Grüning, B.; Ramírez, F.; Manke, T. pyGenomeTracks: Reproducible plots for multivariate genomic data sets. *Bioinformatics* **2020**, *37*, 422–423. [CrossRef]
28. Christofori, G.; Semb, H. The role of the cell-adhesion molecule E-cadherin as a tumour-suppressor gene. *Trends Biochem. Sci.* **1999**, *24*, 73–76. [CrossRef]
29. Oda, T.; Kanai, Y.; Oyama, T.; Yoshiura, K.; Shimoyama, Y.; Birchmeier, W.; Sugimura, T.; Hirohashi, S. E-cadherin gene mutations in human gastric carcinoma cell lines. *Proc. Natl. Acad. Sci. USA* **1994**, *91*, 1858–1862. [CrossRef] [PubMed]
30. Shimada, S.; Akiyama, Y.; Mogushi, K.; Ishigami-Yuasa, M.; Kagechika, H.; Nagasaki, H.; Fukamachi, H.; Yuasa, Y.; Tanaka, S. Identification of selective inhibitors for diffuse-type gastric cancer cells by screening of annotated compounds in preclinical models. *Br. J. Cancer* **2018**, *118*, 972–984. [CrossRef]
31. Carneiro, P.; Moreira, A.M.; Figueiredo, J.; Barros, R.; Oliveira, P.; Fernandes, M.S.; Ferro, A.; Almeida, R.; Oliveira, C.; Carneiro, F.; et al. S100P is a molecular determinant of E-cadherin function in gastric cancer. *Cell Commun. Signal.* **2019**, *17*, 155. [CrossRef] [PubMed]
32. Pinheiro, H.; Bordeira-Carriço, R.; Seixas, S.; Carvalho, J.; Senz, J.; Oliveira, P.; Inácio, P.; Gusmão, L.; Rocha, J.; Huntsman, D.; et al. Allele-specific CDH1 downregulation and hereditary diffuse gastric cancer. *Hum. Mol. Genet.* **2010**, *19*, 943–952. [CrossRef] [PubMed]
33. Rossi, T.; Tedaldi, G.; Petracci, E.; Abou Khouzam, R.; Ranzani, G.N.; Morgagni, P.; Saragoni, L.; Monti, M.; Calistri, D.; Ulivi, P.; et al. E-cadherin Downregulation and microRNAs in Sporadic Intestinal-Type Gastric Cancer. *Int. J. Mol. Sci.* **2019**, *20*, 4452. [CrossRef]
34. Huang, R.; Ding, P.; Yang, F. Clinicopathological significance and potential drug target of CDH1 in breast cancer: A meta-analysis and literature review. *Drug Des. Devel. Ther.* **2015**, *9*, 5277–5285. [PubMed]

35. Canale, M.; Casadei-Gardini, A.; Ulivi, P.; Arechederra, M.; Berasain, C.; Lollini, P.-L.; Fernández-Barrena, M.G.; Avila, M.A. Epigenetic Mechanisms in Gastric Cancer: Potential New Therapeutic Opportunities. *Int. J. Mol. Sci.* **2020**, *21*, 5500. [CrossRef] [PubMed]
36. van der Post, R.S.; Vogelaar, I.P.; Carneiro, F.; Guilford, P.; Huntsman, D.; Hoogerbrugge, N.; Caldas, C.; Schreiber, K.E.C.; Hardwick, R.H.; Ausems, M.G.E.M.; et al. Hereditary diffuse gastric cancer: Updated clinical guidelines with an emphasis on germline CDH1 mutation carriers. *J. Med. Genet.* **2015**, *52*, 361–374. [CrossRef]
37. Tedaldi, G.; Pirini, F.; Tebaldi, M.; Zampiga, V.; Cangini, I.; Danesi, R.; Arcangeli, V.; Ravegnani, M.; Abou Khouzam, R.; Molinari, C.; et al. Multigene Panel Testing Increases the Number of Loci Associated with Gastric Cancer Predisposition. *Cancers* **2019**, *11*, 1340. [CrossRef]
38. UCSC Genome Browser. Available online: https://genome.ucsc.edu/ (accessed on 8 May 2021).
39. FANTOM Consortium and the RIKEN PMI and CLST (DGT); Forrest, A.R.R.; Kawaji, H.; Rehli, M.; Baillie, J.K.; de Hoon, M.J.L.; Haberle, V.; Lassmann, T.; Kulakovskiy, I.V.; Lizio, M.; et al. A promoter-level mammalian expression atlas. *Nature* **2014**, *507*, 462–470.
40. Fishilevich, S.; Nudel, R.; Rappaport, N.; Hadar, R.; Plaschkes, I.; Iny Stein, T.; Rosen, N.; Kohn, A.; Twik, M.; Safran, M.; et al. GeneHancer: Genome-wide integration of enhancers and target genes in GeneCards. *Database* **2017**, *2017*. [CrossRef]
41. Livak, K.J.; Schmittgen, T.D. Analysis of relative gene expression data using real-time quantitative PCR and the 2(-Delta Delta C(T)) Method. *Methods* **2001**, *25*, 402–408. [CrossRef]
42. Schmittgen, T.D.; Livak, K.J. Analyzing real-time PCR data by the comparative C(T) method. *Nat. Protoc.* **2008**, *3*, 1101–1108. [CrossRef] [PubMed]

Appendix D

Other syndromes and genes associated with gastric cancer predisposition

Gianluca Tedaldi, Celina São José, Carla Oliveira

In: Corso, G., Veronesi, P., Roviello, F. (eds) Hereditary Gastric and Breast Cancer Syndrome.

Springer, Cham

2023

DOI: 10.1007/978-3-031-21317-5_7

Other Syndromes and Genes Associated with Gastric Cancer Predisposition

7

Gianluca Tedaldi, Celina São José, and Carla Oliveira

Abstract

Gastric cancer is the fifth most common type of cancer and the fourth leading cause of cancer-related death; nevertheless, genetic predisposition to this malignancy is still widely unexplored.

Besides hereditary diffuse gastric cancer (HDGC), associated with germline *CDH1* and *CTNNA1* pathogenic variants, other genetic syndromes characterized by high risk to develop gastric cancer have been described, encompassing gastric adenocarcinoma and proximal polyposis of the stomach (GAPPS), associated with germline genetic variants in the *APC* promoter, and familial intestinal gastric cancer (FIGC), still lacking a clear genetic cause.

Moreover, gastric cancer risk is associated with pathogenic variants in genes involved in DNA mismatch repair, such as *MLH1* and *MSH2* (Lynch syndrome), apoptosis, including *TP53* (Li-Fraumeni syndrome) and double-strand break

G. Tedaldi (✉)
Biosciences Laboratory, IRCCS Istituto Romagnolo per lo Studio dei Tumori (IRST) "Dino Amadori", Meldola, Italy
e-mail: gianluca.tedaldi@irst.emr.it

C. São José
i3S, Instituto de Investigação e Inovação em Saúde, University of Porto, Porto, Portugal

IPATIMUP, Institute of Molecular Pathology and Immunology, University of Porto, Porto, Portugal

Doctoral Programme in Biomedicine, Faculty of Medicine, University of Porto, Porto, Portugal

C. Oliveira
i3S, Instituto de Investigação e Inovação em Saúde, University of Porto, Porto, Portugal

IPATIMUP, Institute of Molecular Pathology and Immunology, University of Porto, Porto, Portugal

FMUP - Faculty of Medicine of the University of Porto, Porto, Portugal

© The Author(s), under exclusive license to Springer Nature Switzerland AG 2023

G. Corso et al. (eds.), *Hereditary Gastric and Breast Cancer Syndrome*,
https://doi.org/10.1007/978-3-031-21317-5_7

repair, such as *BRCA1*/*BRCA2* and *PALB2* (hereditary breast and ovarian cancer syndrome).

Furthermore, gastric cancer can be a manifestation of gastrointestinal polyposis syndromes, such as those associated with *APC* (familial adenomatous polyposis), *MUTYH* (MUTYH-associated polyposis), *BMPR1A*/*SMAD4* (juvenile polyposis syndrome), *STK11* (Peutz-Jeghers syndrome), and *PTEN* (Cowden syndrome) genes.

Recent advances in molecular techniques, such as next-generation sequencing, led to the identification of many new genes involved in the predisposition to gastric cancer, some of which are low or moderate penetrant that predispose to other syndromes.

Consequently, in patients with early onset gastric cancer and/or strong gastric cancer family history, the use of multigene panel testing should be considered in cancer risk assessment, including different surveillance recommendations for each syndrome.

7.1 Introduction

Familial predisposition to gastric cancer (GC) has been categorized into three main syndromes with primary predisposition to the stomach: (1) hereditary diffuse gastric cancer (HDGC), (2) familial intestinal gastric cancer (FIGC), and (3) gastric adenocarcinoma and proximal polyposis of the stomach (GAPPS). While tumor burden and main genetic causality are established for HDGC (*CDH1*, *CTNNA1*) and GAPPS (*APC*), FIGC remains genetically unexplained and understudied. Nevertheless, other genes that predispose for other cancer syndromes encompass GC within their tumor spectrum: (1) *MLH1*, *MSH2*, *MSH6*, and *PMS2* (Lynch syndrome, LS, or hereditary nonpolyposis colorectal cancer, HNPCC), (2) *TP53* (Li-Fraumeni syndrome, LFS), (3) *BRCA1*, *BRCA2*, and *PALB2* (hereditary breast and ovarian cancer, HBOC), (4) *APC* (familial adenomatous polyposis, FAP), (5) *MUTYH* (MUTYH-associated polyposis, MAP), (6) *BMPR1A* and *SMAD4* (juvenile polyposis syndrome, JPS), (7) *STK11* (Peutz-Jeghers syndrome, PJS), and (8) *PTEN* (PTEN Hamartoma Tumor syndrome, PHTS) (Table 7.1) [1].

7.2 HDGC

Pathogenic or likely pathogenic variants in *CDH1* predispose to HDGC, an autosomal dominant syndrome characterized by diffuse gastric cancer (DGC) and lobular breast cancer (LBC) [2].

In recent years, next-generation sequencing (NGS) approaches have evolved exponentially, leading to the identification of new genes in HDGC. In 2013, the first germline truncating variant in *CTNNA1*, encoding the α-E-catenin protein, was described in an HDGC family [3]. To date, and after multiple HDGC families being

Table 7.1 Hereditary syndromes associated with GC

Syndrome	Gene	GC risk (%)	References
GAPPS	*APC (promoter 1B)*	IGC 13%	[18]
FIGC	Probably polygenic	Variable	[19]
LS	*MLH1*	IGC 5–10%	[20–24]
	MSH2	IGC 9%	[23, 24]
	MSH6	IGC ≤1%–7.9%	[23–26]
	PMS2	Low	[23, 24]
	EPCAM	Low	[24, 27]
LFS	*TP53*	IGC or DGC 2–5%	[28, 29]
HBOC	*BRCA1/BRCA2*	IGC 2%	[29, 30]
FAP	*APC*	IGC 4–7% (Asian population), low (Western population)	[24, 31]
MAP	*MUTYH*	IGC 2–5%	[24, 32]
PJS	*STK11*	IGC 29%	[24, 33]
JPS	*SMAD4/BMPR1A*	IGC or DGC 10–30%	[24, 34]
CS	*PTEN*	Low	[29, 35]

IGC: intestinal-type gastric cancer
DGC: diffuse-type gastric cancer

identified to carry *CTNNA1* truncating variants, *CTNNA1* remains the only gene, besides *CDH1*, clearly associated with the HDGC syndrome [4, 5]. Germline mutations in *MAP3K6* and *MYD88* have also been reported in HDGC families [6, 7]; however, the specific role of these genes remains unclear and their involvement in GC predisposition is still questionable [4]. In 2015, a targeted analysis with a panel of 55 cancer-related genes performed on 144 *CDH1*-negative cases found candidate mutations in 16 probands (11%), including high and moderate penetrance mutations in *CTNNA1*, *BRCA2*, *STK11*, *SDHB*, *PRSS1*, *ATM*, *MSR1*, and *PALB2* [8]. Very recently, a whole exome analysis on 54 *CDH1*-negative GC patients did not identify obvious candidates for GC predisposition [9], while, a gene panel-based analysis of 333 HDGC and non-HDGC cases identified 11 mutation carriers of *PALB2*, *BRCA1*, and *RAD51C*, which are genes involved in DNA homologous recombination (HR) [10]. A recent meta-analysis, performed on NGS published data, identified a list of genes carrying deleterious variants in families meeting the 2020 HDGC clinical criteria [11]. Pathogenic or likely pathogenic variants were found in candidate genes involved in DNA damage response pathways [11], encompassing *ATM* [12, 13], *BRCA1* [13], *BRCA2* [8, 13], *PALB2* [8, 10, 13, 14], *RAD51C* [10], and *ATR* [14]. In fact, *PALB2* and *ATM* were the most frequently mutated genes in the HDGC setting [11]. The former has been extensively associated with breast cancer predisposition [15], while the latter has been associated with both breast and gastric cancer susceptibility [16, 17]. Interestingly, *PALB2* loss of function variants have been shown to be enriched in the HDGC setting, compared to the general population [14]. While *PALB2* association with HDGC holds promise, *ATM* pleiotropy prevents a clear association with this disease.

7.3 GAPPS

In 2012, GAPPS, was described as an autosomal dominant syndrome [18]. The key clinical features of GAPPS include fundic gland polyposis (FGP) of the stomach with occasional hyperplastic and adenomatous polyps, focal foveolar-type dysplasia, hyperproliferative aberrant pits and development of adenomas with gastric type dysplasia or intestinal-/mixed-type gastric adenocarcinoma [18, 36, 37]. Current diagnostic criteria are depicted in Table 7.2 [36, 38].

In 2016, linkage analysis on six selected families mapped the gene to the 5q22 chromosomal region. Through Sanger sequencing, point mutations in *APC* promoter 1B, that co-segregated with the disease in all three families, were identified [38, 39]. Therefore, GAPPS is considered a part of a broad phenotypic spectrum of inherited polyposis associated with *APC* germline defects, but with tropism to the stomach (see paragraph "Familial Adenomatous Polyposis"). Since then, 12 additional families were found to harbor *APC* promoter 1B single nucleotide variants (SNVs) [40–43]. Two SNVs were found co-segregating within a family with severer phenotype, but their individual contribution remains unclear [38].

GAPPS phenotypes are diverse among individuals, in the number of polyps, from 30 to hundreds and GC age of onset ranging from 23 to 75 years of age [18, 43]. In fact, third-generation individuals display a much severer phenotype than first-generation obligated carriers [18]. Altogether, these observations suggest incomplete penetrance of *APC* promoter 1B SNVs that may be aggravated by environmental factors and moderate/low penetrance variants. Risk to develop intestinal- or mixed-type GC is 13% (Table 7.1) [18].

Surveillance of GAPPS families includes endoscopic surveillance with biopsies and prophylactic gastrectomy, due to a rapid malignant progression of FGP [18, 40, 43].

Table 7.2 GAPPS clinical criteria for genetic testing

Clinical criteria		Genetic screening
Essential criteria	Body and fundus gastric polyps	*APC* promoter 1B SNVs
	No evidence of colorectal or duodenal polyposis	
	>100 proximal stomach polyps or >30 polyps in a first degree relative GAPPS diagnosed patient	
	Predominantly fundic gland polyposis, which may have dysplasia	
	Relative with dysplastic FGPs or GC	
Supportive criteria	Autosomal dominant inheritance pattern	
	Presence of hyperproliferative aberrant pits, hyperplastic polyps, and gastric-type adenomas	

7.4 FIGC

FIGC is the HDGC counterpart that predisposes to intestinal-type gastric cancer (IGC). Current clinical criteria have been defined by the international gastric cancer linkage consortium (IGCLC) in 1999, depending on the GC incidence in the population and are depicted in Table 7.3 [44, 45]. Countries with a high GC incidence, such as Japan and Portugal, should use criteria analogous to those proposed for Lynch syndrome [46], while in countries with a low GC incidence, including USA and UK, FIGC selection criteria are more restrictive.

To date, no germline defects have been found to be recurrently associated with FIGC predisposition, which currently has unknown age of onset, tumor spectrum, and penetrance. Thus, clinical criteria have not been updated or validated since firstly described in 1999 [44]. Recently, the average IGC age of onset in FIGC families was found to be 10 years earlier than observed for the sporadic setting [19]. At the somatic level, *TP53*, *BRCA2*, *ATM*, *FOXF1*, *FHIT*, *SDHB*, *MSH6*, *CTNNA1*, and *PXN* were found mutated at higher frequencies in tumors from FIGC patients than in sporadic IGC, which also correlates with increased MSI frequency. The FIGC tumor spectrum is broad and predisposes to IGC, but also to colorectal and breast cancer, at lower frequencies [19]. A recent meta-analysis found *BRCA2* as the most frequently mutated gene in the germline DNA of FIGC probands, reaching 17% [11], a frequency that was similar to that of *BRCA2* somatic variants in sporadic IGC (9%) and higher than that of sporadic DGC (5%) [47].

Carvalho and colleagues [19] hint toward FIGC as a polygenic syndrome, since germline defects in major genes were not found in a large FIGC cohort. These authors also proposed redefinition of clinical criteria for FIGC to at least 2 GC cases diagnosed at any age, with one histologically confirmed as IGC [19].

Considering the number of genes that can be involved in this disease, the lifetime GC risk is not easy to determine due to the high genetic variability (Table 7.1).

Current surveillance is evaluated and applied on a case-by-case basis, yet recommendations include endoscopy in first-degree relatives, 10 years earlier than the earliest IGC age of onset [48], or gastroduodenoscopy at 40 years of age or 5 years earlier than the youngest IGC diagnosed in the family [49]. Eradication of *H. pylori* infection is recommended in FIGC families, due to its high frequency in this setting [49].

Table 7.3 FIGC clinical criteria

Clinical criteria		Genetic screening
High GC incidence countries	At least three relatives with IGC, one first-degree of the other two	Unknown germline cause
	At least two successive generations affected	
	GC diagnosed <50 years of age in at least one relative	
Low GC incidence countries	At least two first/second-degree relatives with IGC, one diagnosed <50 years of age	
	At least three relatives with IGC at any age	

7.5 Non-polyposis Syndromes

7.5.1 Lynch Syndrome

Lynch syndrome (LS) predisposes to colorectal and endometrial cancers and follow an autosomal dominant inheritance pattern [50]. LS is caused by pathogenic variants in *MLH1*, *MSH2*, *MSH6*, and *PMS2*, that encode the DNA mismatch repair (MMR) proteins [51], or by large deletions of the *EPCAM* gene, located upstream of *MSH2* [52]. MMR proteins work in a coordinated mode to repair the DNA mismatches that arise during DNA replication and recombination [53].

LS patients also have an increased risk of developing other tumors [54, 55], encompassing a lifetime risk to develop gastric cancer, estimated to be 1–10%, according to the altered gene (Table 7.1).

Regarding GC surveillance, LS patients with an *MLH1*/*MSH2* pathogenic variant, a family history of GC, and other risk factors should undergo upper endoscopy every 3–5 years beginning at age 40 [24].

Moreover, patients with LS, who have a deficiency of the MMR system (dMMR), can benefit from chemoprevention based on the daily use of aspirin [56] and, in case MSI cancers develop, may be treated with anti-PD-1/PD-L1 therapy [57, 58].

7.5.2 Li-Fraumeni Syndrome

The *TP53* gene is located on chromosome 17p13.1 and encodes the p53 protein, a tumor suppressor that responds to different cellular stresses to regulate expression of target genes, thereby inducing cell cycle arrest, apoptosis, senescence, DNA repair, or metabolism changes [59]. Due to its crucial function in maintaining the genomic stability, p53 has been defined as "the guardian of the genome" and, indeed, *TP53* somatic alterations are present in approximately 50% of sporadic tumors [60], conferring to p53 an important role as a biomarker for the diagnosis, tumor progression, poor prognosis, and reduced sensitivity for anticancer drugs [61].

Germline pathogenic variants in the *TP53* gene are associated with Li-Fraumeni syndrome (LFS), a rare autosomal dominant disorder characterized by a high predisposition to several types of cancer, such as brain tumors, breast cancer, sarcomas, acute leukemia, and adrenocortical tumors [28, 62–71].

The lifetime risk of GC for patients with LFS, although not consensual, has been estimated to be 2–5% (Table 7.1) [28, 72, 73].

Given the risk of developing gastrointestinal cancers, the guidelines suggest that LFS patients should undergo upper endoscopy and colonoscopy every 2–5 years starting from age 25 years [29]. Moreover, in children, the recommendations are to perform clinical examination and abdominal ultrasound every 6 months, annual whole-body MRI, and brain MRI from the first year of life, if the *TP53* variant is known to be associated with childhood cancers. In adults, the surveillance should include every year clinical examination, whole-body MRI, breast MRI in females from 20 until 65 years, and brain MRI until 50 years [63].

7.5.3 BRCA1- and BRCA2-Associated Hereditary Breast and Ovarian Cancer

The *BRCA1* gene, located on chromosome 17q21.31, encodes a nuclear protein involved in DNA repair, cell cycle checkpoint control, and maintenance of genomic stability forming a large multi-subunit protein complex known as BRCA1-associated genome surveillance complex (BASC) [74–77].

The *BRCA2* gene is located on chromosome 13q13.1 and encodes a nuclear protein involved in repairing damaged DNA, recruiting the recombinase RAD51 to the DNA double-strand breaks (DSBs) through the formation of a BRCA1-PALB2-BRCA2 complex [74, 75, 78, 79].

Germline pathogenic variants in *BRCA1* and *BRCA2* genes are associated with the hereditary breast and ovarian cancer (HBOC) syndrome [80], characterized by a high risk of developing breast and ovarian cancer in females [81–83], breast and prostate cancer in males [84–87] and pancreatic cancer in both sexes [88–90].

Further, *BRCA1* pathogenic variants have been associated with an increased risk of colon cancer [91] and *BRCA2* pathogenic variants have been associated with uveal melanoma [92, 93].

Recently, pathogenic variants in *BRCA1/2* and other genes involved in breast/ovarian cancer predisposition have been associated with an increased GC risk [8, 10, 12–14]. The IGC risk is estimated to be 2% in *BRCA1/2* pathogenic variant carriers (Table 7.1) [30], however prevention should be evaluated on the basis of family history [24].

Moreover, the discovery of the therapeutic potential of inhibitors of the poly adenosine-diphosphate ribose polymerase (PARP) in carriers of germline/somatic *BRCA1/2* pathogenic variants with ovarian, breast, prostate, and pancreatic cancers led to a revolution in the treatment of these tumors [94–100]. PARP inhibitors have shown their efficacy also in patients with pathogenic variants in genes involved in the HR pathway [101–104]. These results pave the way for the future use of PARP inhibitors in all tumors with a deficiency of the HR system, independently of the germline or somatic nature of the alteration, including GC [105].

7.6 Polyposis Syndromes

7.6.1 Familial Adenomatous Polyposis

The APC protein is a tumor suppressor that acts as a Wnt signaling antagonist, and regulates transcriptional activation, cell migration and apoptosis [106]. Pathogenic or likely pathogenic alterations in the *APC* gene (chromosome 5q22.2) predispose to familial adenomatous polyposis (FAP) [107, 108]. This autosomal dominant syndrome is characterized by polyposis and carcinomas in the gastrointestinal tract, as well as, extra-gastrointestinal carcinomas, such as thyroid [34]. While classical FAP predisposes to hundreds to thousands of colonic and rectal polyps that may develop into colorectal carcinoma, attenuated FAP (AFAP) displays a much milder

phenotype [34, 109]. Families with AFAP present fewer and latter-onset of both polyps and carcinomas, as well as cancer-decreased risk [110]. The phenotype severity is dependent on the mutation location within the *APC* gene [111], as above mentioned for GAPPS with unique predisposition to the stomach [38].

FAP and AFAP also predispose to gastric polyps in >60% and 93% of patients, respectively [112]. However, gastric adenocarcinoma risk ranges between 4% and 7% in the Asian population, with no increased risk for the western population (Table 7.1) [24, 31]. In fact, FGP and focal low-grade dysplasia in the stomach commonly do not undergo malignant transformation [113, 114]. Nevertheless, increased risk is observed in the presence of FGP stomach carpeting, polyps larger than 20 mm, tubular adenomas, high-grade dysplasia polyps, pyloric gland adenomas, and in specific geographical areas [31, 115, 116]. According to these high-risk features and family history, specialized surveillance or gastrectomy may be recommended [24].

7.6.2 MUTYH-Associated Polyposis

The *MUTYH* gene is located on chromosome 1p34.1 and encodes the MutY DNA glycosylase, involved in oxidative DNA damage repair and, if unrepaired, apoptosis signaling [117].

MUTYH-associated polyposis (MAP) distinguishes from (A)FAP by presenting a recessive inheritance pattern with reduced risk for colonic and duodenal adenomas (fewer than 100) and carcinomas (5%). Thus, biallelic pathogenic or likely pathogenic variants in *MUTYH* (chromosome 1p34.1) predispose to MAP [118]. Risk to develop IGC ranges from 2% for females to 4% for males (Table 7.1) [32].

Current surveillance measurements include upper endoscopy and side viewing duodenoscopy every 3 months to 4 years beginning at age 30–35 years with subsequent follow-up based on initial findings [24, 119, 120].

7.6.3 Juvenile Polyposis Syndrome

The *BMPR1A* gene, located on chromosome 10q23.2, encodes the bone morphogenetic protein receptor type IA, a transmembrane serine/threonine kinase that binds members of the TGF-β superfamily and plays a role in signal transduction, apoptosis and cell differentiation [121].

The *SMAD4* gene (chromosome 18q21.2) encodes a member of the Smad family of signal transduction proteins that are activated by transmembrane serine-threonine receptor kinases in response to TGF-β and bone morphogenetic protein signaling pathways. *SMAD4* is a transcription factor that acts as a tumor suppressor and inhibits epithelial cell proliferation [122].

Germline pathogenic variants in *BMPR1A* and *SMAD4* genes are associated with juvenile polyposis syndrome (JPS), an autosomal dominant disorder, that

predisposes to hamartomatous polyps in the gastrointestinal tract, specifically in the stomach, small intestine, colon, and rectum [123].

The majority of juvenile polyps are benign, however can undergo malignant transformation. Lifetime estimates of developing gastrointestinal cancers in families with JPS range from 11% to 86%, with variability by region, time period included, and associated gene [124–128]. In fact, approximately 15% of JPS individuals develop cancer [127, 129]. While, the GC incidence is approximately around 10–30% in JPS patients with gastric polyps (Table 7.1) [130, 131], the risk of colorectal cancer ranges between 17% and 22% by 35 years of age and approaches 68% by 60 years of age [132]. In JPS context, small bowel and pancreatic cancers have also been reported [133–137]. Individuals with *SMAD4*-related JPS are more likely to have a personal or family history of upper gastrointestinal polyps than individuals with a *BMPR1A* pathogenic variant. The gastric phenotype in individuals with a *SMAD4* pathogenic variant tends to be more aggressive with significant polyposis, anemia, and a higher GC risk [125, 127, 128].

According to the clinical practice guidelines for JPS, the gastric surveillance recommended for individuals with a *BMPR1A* or *SMAD4* pathogenic variant includes colonoscopy and upper endoscopy every 3 years beginning at age 15 or earlier if symptomatic. If polyps are found, after polyp treatment an annual screening is recommended until no polyps are found, followed by a screening every 3 years [24, 138, 139].

7.6.4 Peutz-Jeghers Syndrome

The *STK11* gene (formerly *LKB1*) is located on chromosome 19p13.3 and encodes a serine/threonine kinase that acts as a tumor suppressor, regulating energy metabolism and cell polarity [140].

Germline pathogenic variants in the *STK11* gene are associated with Peutz-Jeghers syndrome (PJS), an autosomal dominant syndrome. PJS is characterized by melanocytic macules of the lips, buccal mucosa and digits, multiple gastrointestinal hamartomatous polyps, and an increased risk for different tumors, encompassing colorectal, gastric, pancreatic, breast, and ovarian cancers [141].

In *STK11* pathogenic variant carriers, the lifetime GC risk is estimated to be 29% (Table 7.1) [33, 34, 142, 143]. For this reason, the clinical guidelines suggest that PJS patients should undergo upper endoscopy with polypectomy every 2–3 years, starting at the age of 18; shorter intervals may be indicated based on polyp size, number, and pathology [24].

7.6.5 PTEN Hamartoma Tumor Syndrome

The *PTEN* gene (chromosome 10q23.31) encodes a phosphatase which antagonizes the PI3K signaling pathway and negatively regulates the MAPK pathway [144].

Germline pathogenic variants in *PTEN* are associated with the PTEN hamartoma tumor syndrome (PHTS) that includes Cowden syndrome, Bannayan-Riley-Ruvalcaba syndrome, PTEN-related Proteus syndrome, and PTEN-related proteus-like syndrome [145].

Cowden syndrome (CS) is an autosomal dominant disorder that predisposes to benign hamartomas and increased lifetime risk of breast, thyroid, uterine, colorectal, and other cancers, including stomach [145–147]. Upper or lower gastrointestinal polyps occur in more than 90% of individuals with a *PTEN* pathogenic variant [148]. In the stomach, the most common findings are hyperplastic polyps, hamartomas, and ganglioneuromas [149–151].

Cowden syndrome does not have increased risk of gastric malignancy (Table 7.1); however, complications of benign neoplasm can occur [35]. Indeed, some CS patients have symptoms including hemorrhage, obstruction, and pain [35]. According to the guidelines, PHTS patients should undergo upper and lower endoscopy with removal of polyps beginning at age 35 years with frequency dependent on degree of polyposis identified [145].

7.7 Conclusions

GC is one of the most common and deadly tumors and, among risk factors for the development of this cancer, genetic predisposition plays an important role.

Besides HDGC, associated with *CDH1* and *CTNNA1* pathogenic variants, other genetic syndromes characterized by high risk to develop GC have been described: GAPPS, associated with genetic variants in the *APC* promoter, and FIGC, still lacking a clear genetic cause.

In addition to these three syndromes, genes including *TP53*, *BRCA1/2*, and MMR genes, whose variants are associated with other cancer genetic syndromes, also include an increased risk for GC (Table 7.1).

Moreover, genes associated with the development of gastrointestinal polyps, such as *APC*, *MUTYH*, *BMPR1A*, *SMAD4*, *STK1*, and *PTEN* may also evolve in GC (Table 7.1).

The evidence of GC risk associated with these syndromes and the availability of recommendations for the management of variant carriers suggest that these genes should be included in a gene panel for the identification of patients at risk of developing GC.

In summary, new genes are constantly emerging from NGS studies, showing that GC predisposition is distributed over several genes, with only a small portion of genes being recurrently mutated.

These findings address the choice of wide panels, including the genes involved in the main cancer syndromes. This creates new diagnostic opportunities but also increases the risk of an incorrect genetic diagnosis [152]. Importantly, the identification of a pathogenic germline variant can not only guide the choice of the best chemoprevention and prophylactic surgeries but also the choice of novel targeted therapies, toward personalized medicine based on the genetic characteristics of each patient.

References

1. van der Post RS, Vogelaar IP, Carneiro F et al (2015) Hereditary diffuse gastric cancer: updated clinical guidelines with an emphasis on germline CDH1 mutation carriers. J Med Genet 52:361–374. https://doi.org/10.1136/jmedgenet-2015-103094
2. Guilford P, Hopkins J, Harraway J et al (1998) E-cadherin germline mutations in familial gastric cancer. Nature 392:402–405. https://doi.org/10.1038/32918
3. Majewski IJ, Kluijt I, Cats A et al (2013) An α-E-catenin (CTNNA1) mutation in hereditary diffuse gastric cancer. J Pathol 229:621–629. https://doi.org/10.1002/path.4152
4. Weren RDA, van der Post RS, Vogelaar IP et al (2018) Role of germline aberrations affecting CTNNA1, MAP3K6 and MYD88 in gastric cancer susceptibility. J Med Genet 55:669–674. https://doi.org/10.1136/jmedgenet-2017-104962
5. Lobo S, Benusiglio PR, Coulet F et al (2021) Cancer predisposition and germline CTNNA1 variants. Eur J Med Genet 64:104316. https://doi.org/10.1016/j.ejmg.2021.104316
6. Gaston D, Hansford S, Oliveira C et al (2014) Germline mutations in MAP3K6 are associated with familial gastric cancer. PLoS Genet 10:e1004669. https://doi.org/10.1371/journal.pgen.1004669
7. Vogelaar IP, Ligtenberg MJL, van der Post RS et al (2016) Recurrent candidiasis and early-onset gastric cancer in a patient with a genetically defined partial MYD88 defect. Familial Cancer 15:289–296. https://doi.org/10.1007/s10689-015-9859-z
8. Hansford S, Kaurah P, Li-Chang H et al (2015) Hereditary diffuse gastric cancer syndrome: CDH1 mutations and beyond. JAMA Oncol 1:23–32. https://doi.org/10.1001/jamaoncol.2014.168
9. Vogelaar IP, van der Post RS, van Krieken JHJ et al (2017) Unraveling genetic predisposition to familial or early onset gastric cancer using germline whole-exome sequencing. Eur J Hum Genet 25:1246–1252. https://doi.org/10.1038/ejhg.2017.138
10. Sahasrabudhe R, Lott P, Bohorquez M et al (2017) Germline mutations in PALB2, BRCA1, and RAD51C, which regulate DNA recombination repair, in patients with gastric cancer. Gastroenterology 152:983–986.e6. https://doi.org/10.1053/j.gastro.2016.12.010
11. Garcia-Pelaez J, Barbosa-Matos R, São José C et al (2022) Gastric cancer genetic predisposition and clinical presentations: established heritable causes and potential candidate genes. Eur J Med Genet 65:104401. https://doi.org/10.1016/j.ejmg.2021.104401
12. Slavin T, Neuhausen SL, Rybak C et al (2017) Genetic gastric cancer susceptibility in the international clinical cancer genomics community research network. Cancer Genet 216–217:111–119. https://doi.org/10.1016/j.cancergen.2017.08.001
13. Tedaldi G, Pirini F, Tebaldi M et al (2019) Multigene panel testing increases the number of loci associated with gastric cancer predisposition. Cancers (Basel) 11. https://doi.org/10.3390/cancers11091340
14. Fewings E, Larionov A, Redman J et al (2018) Germline pathogenic variants in PALB2 and other cancer-predisposing genes in families with hereditary diffuse gastric cancer without CDH1 mutation: a whole-exome sequencing study. Lancet Gastroenterol Hepatol 3:489–498. https://doi.org/10.1016/S2468-1253(18)30079-7
15. Antoniou AC, Casadei S, Heikkinen T et al (2014) Breast-cancer risk in families with mutations in PALB2. N Engl J Med 371:497–506. https://doi.org/10.1056/NEJMoa1400382
16. Helgason H, Rafnar T, Olafsdottir HS et al (2015) Loss-of-function variants in ATM confer risk of gastric cancer. Nat Genet 47:906–910. https://doi.org/10.1038/ng.3342
17. Moslemi M, Moradi Y, Dehghanbanadaki H et al (2021) The association between ATM variants and risk of breast cancer: a systematic review and meta-analysis. BMC Cancer 21:27. https://doi.org/10.1186/s12885-020-07749-6
18. Worthley DL, Phillips KD, Wayte N et al (2012) Gastric adenocarcinoma and proximal polyposis of the stomach (GAPPS): a new autosomal dominant syndrome. Gut 61:774–779. https://doi.org/10.1136/gutjnl-2011-300348

19. Carvalho J, Oliveira P, Senz J et al (2020) Redefinition of familial intestinal gastric cancer: clinical and genetic perspectives. J Med Genet 58:1. https://doi.org/10.1136/jmedgenet-2019-106346
20. Cheng L, Eng C, Nieman LZ et al (2011) Trends in colorectal cancer incidence by anatomic site and disease stage in the United States from 1976 to 2005. Am J Clin Oncol 34:573–580. https://doi.org/10.1097/COC.0b013e3181fe41ed
21. Siegel RL, Medhanie GA, Fedewa SA, Jemal A (2019) State variation in early-onset colorectal cancer in the United States, 1995-2015. J Natl Cancer Inst 111:1104–1106. https://doi.org/10.1093/jnci/djz098
22. Giardiello FM, Offerhaus JG Phenotype and cancer risk of various polyposis syndromes. Eur J Cancer 31A:1085–7. https://doi.org/10.1016/0959-8049(95)00139-a
23. Møller P, Seppälä TT, Bernstein I et al (2018) Cancer risk and survival in path_MMR carriers by gene and gender up to 75 years of age: a report from the prospective Lynch syndrome database. Gut 67:1306–1316. https://doi.org/10.1136/gutjnl-2017-314057
24. National Comprehensive Cancer Network (NCCN) (2022) Genetic/Familial High-Risk Assessment: Colorectal v 2.2021. https://www.nccn.org/professionals/physician_gls/pdf/genetics_colon.pdf
25. Siegel RL, Miller KD, Fuchs HE, Jemal A (2022) Cancer statistics, 2022. CA Cancer J Clin 72:7–33. https://doi.org/10.3322/caac.21708
26. Bailey CE, Hu C-Y, You YN et al (2015) Increasing disparities in the age-related incidences of colon and rectal cancers in the United States, 1975-2010. JAMA Surg 150:17–22. https://doi.org/10.1001/jamasurg.2014.1756
27. Rumilla K, Schowalter KV, Lindor NM et al (2011) Frequency of deletions of EPCAM (TACSTD1) in MSH2-associated Lynch syndrome cases. J Mol Diagn 13:93–99. https://doi.org/10.1016/j.jmoldx.2010.11.011
28. Masciari S, Dewanwala A, Stoffel EM et al (2011) Gastric cancer in individuals with Li-Fraumeni syndrome. Genet Med 13:651–657. https://doi.org/10.1097/GIM.0b013e31821628b6
29. National Comprehensive Cancer Network (NCCN) (2022) Genetic/familial high-risk assessment: breast, ovarian, and pancreatic v. 2.2022. https://www.nccn.org/professionals/physician_gls/pdf/genetics_bop.pdf
30. Friedenson B (2005) BRCA1 and BRCA2 pathways and the risk of cancers other than breast or ovarian. MedGenMed 7:60
31. Shibata C, Ogawa H, Miura K et al (2013) Clinical characteristics of gastric cancer in patients with familial adenomatous polyposis. Tohoku J Exp Med 229:143–146. https://doi.org/10.1620/tjem.229.143
32. Win AK, Reece JC, Dowty JG et al (2016) Risk of extracolonic cancers for people with biallelic and monoallelic mutations in MUTYH. Int J Cancer 139:1557–1563. https://doi.org/10.1002/ijc.30197
33. Hearle N, Schumacher V, Menko FH et al (2006) Frequency and spectrum of cancers in the Peutz-Jeghers syndrome. Clin Cancer Res 12:3209–3215. https://doi.org/10.1158/1078-0432.CCR-06-0083
34. Syngal S, Brand RE, Church JM et al (2015) ACG clinical guideline: genetic testing and management of hereditary gastrointestinal cancer syndromes. Am J Gastroenterol 110:223–262. https://doi.org/10.1038/ajg.2014.435
35. Correia TF, Mesquita I, Marcos M, et al (2021) Surgical approach to gastric polyposis in Cowen syndrome-case report. J Surg Case Rep 2021:rjab258. https://doi.org/10.1093/jscr/rjab258
36. Rudloff U (2018) Gastric adenocarcinoma and proximal polyposis of the stomach: diagnosis and clinical perspectives. Clin Exp Gastroenterol 11:447–459. https://doi.org/10.2147/CEG.S163227
37. Gullo I, van der Post RS, Carneiro F (2021) Recent advances in the pathology of heritable gastric cancer syndromes. Histopathology 78:125–147. https://doi.org/10.1111/his.14228

38. Li J, Woods SL, Healey S et al (2016) Point mutations in exon 1B of APC reveal gastric adenocarcinoma and proximal polyposis of the stomach as a familial adenomatous polyposis variant. Am J Hum Genet 98:830–842. https://doi.org/10.1016/j.ajhg.2016.03.001
39. Beer A, Streubel B, Asari R et al (2017) Gastric adenocarcinoma and proximal polyposis of the stomach (GAPPS) - a rare recently described gastric polyposis syndrome - report of a case. Z Gastroenterol 55:1131–1134. https://doi.org/10.1055/s-0043-117182
40. Foretová L, Navrátilová M, Svoboda M et al (2019) GAPPS - gastric adenocarcinoma and proximal polyposis of the stomach syndrome in 8 families tested at Masaryk Memorial Cancer Institute - prevention and prophylactic gastrectomies. Klin Onkol 32:109–117. https://doi.org/10.14735/amko2019S109
41. Kanemitsu K, Iwatsuki M, Yamashita K et al (2021) Two Asian families with gastric adenocarcinoma and proximal polyposis of the stomach successfully treated via laparoscopic total gastrectomy. Clin J Gastroenterol 14:92–97. https://doi.org/10.1007/s12328-020-01290-6
42. Mitsui Y, Yokoyama R, Fujimoto S et al (2018) First report of an Asian family with gastric adenocarcinoma and proximal polyposis of the stomach (GAPPS) revealed with the germline mutation of the APC exon 1B promoter region. Gastric Cancer 21:1058–1063. https://doi.org/10.1007/s10120-018-0855-5
43. Repak R, Kohoutova D, Podhola M et al (2016) The first European family with gastric adenocarcinoma and proximal polyposis of the stomach: case report and review of the literature. Gastrointest Endosc 84:718–725. https://doi.org/10.1016/j.gie.2016.06.023
44. Caldas C, Carneiro F, Lynch HT et al (1999) Familial gastric cancer: overview and guidelines for management. J Med Genet 36:873–880
45. Oliveira C, Suriano G, Ferreira P et al (2004) Genetic screening for familial gastric cancer. Hered Cancer Clin Pract 2:51–64. https://doi.org/10.1186/1897-4287-2-2-51
46. Shinmura K, Kohno T, Takahashi M et al (1999) Familial gastric cancer: clinicopathological characteristics, RER phenotype and germline p53 and E-cadherin mutations. Carcinogenesis 20:1127–1131. https://doi.org/10.1093/carcin/20.6.1127
47. Garcia-Pelaez J, Barbosa-Matos R, Gullo I et al (2021) Histological and mutational profile of diffuse gastric cancer: current knowledge and future challenges. Mol Oncol 1–27:2841. https://doi.org/10.1002/1878-0261.12948
48. Sereno M, Aguayo C, Guillén Ponce C et al (2011) Gastric tumours in hereditary cancer syndromes: clinical features, molecular biology and strategies for prevention. Clin Transl Oncol 13:599–610. https://doi.org/10.1007/s12094-011-0705-y
49. Corso G, Roncalli F, Marrelli D et al (2013) History, pathogenesis, and management of familial gastric cancer: original study of John XXIII's family. Biomed Res Int 2013:385132. https://doi.org/10.1155/2013/385132
50. Kohlmann W, Gruber SB (2004) Lynch syndrome. GeneReviews® (updated 2021)
51. Chung DC, Rustgi AK (1995) DNA mismatch repair and cancer. Gastroenterology 109:1685–1699. https://doi.org/10.1016/0016-5085(95)90660-6
52. Tutlewska K, Lubinski J, Kurzawski G (2013) Germline deletions in the EPCAM gene as a cause of Lynch syndrome - literature review. Hered Cancer Clin Pract 11:9. https://doi.org/10.1186/1897-4287-11-9
53. Iyer RR, Pluciennik A, Burdett V, Modrich PL (2006) DNA mismatch repair: functions and mechanisms. Chem Rev 106:302–323. https://doi.org/10.1021/cr0404794
54. Bonadona V, Bonaïti B, Olschwang S et al (2011) Cancer risks associated with germline mutations in MLH1, MSH2, and MSH6 genes in Lynch syndrome. JAMA 305:2304–2310. https://doi.org/10.1001/jama.2011.743
55. Walsh MD, Buchanan DD, Cummings MC et al (2010) Lynch syndrome-associated breast cancers: clinicopathologic characteristics of a case series from the colon cancer family registry. Clin Cancer Res 16:2214–2224. https://doi.org/10.1158/1078-0432.CCR-09-3058
56. Burn J, Bishop DT, Mecklin J-P et al (2008) Effect of aspirin or resistant starch on colorectal neoplasia in the Lynch syndrome. N Engl J Med 359:2567–2578. https://doi.org/10.1056/NEJMoa0801297

57. Le DT, Uram JN, Wang H et al (2015) PD-1 blockade in tumors with mismatch-repair deficiency. N Engl J Med 372:2509–2520. https://doi.org/10.1056/NEJMoa1500596
58. Zhao P, Li L, Jiang X, Li Q (2019) Mismatch repair deficiency/microsatellite instability-high as a predictor for anti-PD-1/PD-L1 immunotherapy efficacy. J Hematol Oncol 12:54. https://doi.org/10.1186/s13045-019-0738-1
59. Harris CC (1996) Structure and function of the p53 tumor suppressor gene: clues for rational cancer therapeutic strategies. J Natl Cancer Inst 88:1442–1455. https://doi.org/10.1093/jnci/88.20.1442
60. Ozaki T, Nakagawara A (2011) p53: the attractive tumor suppressor in the cancer research field. J Biomed Biotechnol 2011:603925. https://doi.org/10.1155/2011/603925
61. Boyd MT, Vlatkovic N (2008) p53: a molecular marker for the detection of cancer. Expert Opin Med Diagn 2:1013–1024. https://doi.org/10.1517/17530059.2.9.1013
62. Schneider K, Zelley K, Nichols KE, Garber J (1999) Li-Fraumeni Syndrome. GeneReviews® (updated 2019)
63. Frebourg T, Bajalica Lagercrantz S, Oliveira C et al (2020) Guidelines for the Li-Fraumeni and heritable TP53-related cancer syndromes. Eur J Hum Genet 28:1379–1386. https://doi.org/10.1038/s41431-020-0638-4
64. Li FP, Fraumeni JF (1969) Soft-tissue sarcomas, breast cancer, and other neoplasms. A familial syndrome? Ann Intern Med 71:747–752. https://doi.org/10.7326/0003-4819-71-4-747
65. Chompret A, Brugières L, Ronsin M et al (2000) P53 germline mutations in childhood cancers and cancer risk for carrier individuals. Br J Cancer 82:1932–1937. https://doi.org/10.1054/bjoc.2000.1167
66. Nichols KE, Malkin D, Garber JE et al (2001) Germ-line p53 mutations predispose to a wide spectrum of early-onset cancers. Cancer Epidemiol Biomark Prev 10:83–87
67. Olivier M, Goldgar DE, Sodha N et al (2003) Li-Fraumeni and related syndromes: correlation between tumor type, family structure, and TP53 genotype. Cancer Res 63:6643–6650
68. Varley JM (2003) Germline TP53 mutations and Li-Fraumeni syndrome. Hum Mutat 21:313–320. https://doi.org/10.1002/humu.10185
69. Gonzalez KD, Noltner KA, Buzin CH et al (2009) Beyond Li Fraumeni syndrome: clinical characteristics of families with p53 germline mutations. J Clin Oncol 27:1250–1256. https://doi.org/10.1200/JCO.2008.16.6959
70. Wong P, Verselis SJ, Garber JE et al (2006) Prevalence of early onset colorectal cancer in 397 patients with classic Li-Fraumeni syndrome. Gastroenterology 130:73–79. https://doi.org/10.1053/j.gastro.2005.10.014
71. Birch JM, Alston RD, McNally RJ et al (2001) Relative frequency and morphology of cancers in carriers of germline TP53 mutations. Oncogene 20:4621–4628. https://doi.org/10.1038/sj.onc.1204621
72. Bougeard G, Renaux-Petel M, Flaman JM et al (2015) Revisiting Li-Fraumeni syndrome from TP53 mutation carriers. J Clin Oncol 33:2345–2352. https://doi.org/10.1200/JCO.2014.59.5728
73. Mai PL, Best AF, Peters JA et al (2016) Risks of first and subsequent cancers among TP53 mutation carriers in the National Cancer Institute Li-Fraumeni syndrome cohort. Cancer 122:3673–3681. https://doi.org/10.1002/cncr.30248
74. Roy R, Chun J, Powell SN (2011) BRCA1 and BRCA2: different roles in a common pathway of genome protection. Nat Rev Cancer 12:68–78. https://doi.org/10.1038/nrc3181
75. Foulkes WD, Shuen AY (2013) In brief: BRCA1 and BRCA2. J Pathol 230:347–349. https://doi.org/10.1002/path.4205
76. Wang Y, Cortez D, Yazdi P et al (2000) BASC, a super complex of BRCA1-associated proteins involved in the recognition and repair of aberrant DNA structures. Genes Dev 14:927–939
77. Hedenfalk IA, Ringnér M, Trent JM, Borg A (2002) Gene expression in inherited breast cancer. Adv Cancer Res 84:1–34

78. Xia B, Sheng Q, Nakanishi K et al (2006) Control of BRCA2 cellular and clinical functions by a nuclear partner, PALB2. Mol Cell 22:719–729. https://doi.org/10.1016/j.molcel.2006.05.022
79. Buisson R, Dion-Côté A-M, Coulombe Y et al (2010) Cooperation of breast cancer proteins PALB2 and piccolo BRCA2 in stimulating homologous recombination. Nat Struct Mol Biol 17:1247–1254. https://doi.org/10.1038/nsmb.1915
80. Petrucelli N, Daly MB, Pal T (1998) BRCA1- and BRCA2-associated hereditary breast and ovarian cancer. GeneReviews® (updated 2022)
81. Antoniou A, Pharoah PDP, Narod S et al (2003) Average risks of breast and ovarian cancer associated with BRCA1 or BRCA2 mutations detected in case series unselected for family history: a combined analysis of 22 studies. Am J Hum Genet 72:1117–1130. https://doi.org/10.1086/375033
82. Chen S, Parmigiani G (2007) Meta-analysis of BRCA1 and BRCA2 penetrance. J Clin Oncol 25:1329–1333. https://doi.org/10.1200/JCO.2006.09.1066
83. Mavaddat N, Peock S, Frost D et al (2013) Cancer risks for BRCA1 and BRCA2 mutation carriers: results from prospective analysis of EMBRACE. J Natl Cancer Inst 105:812–822. https://doi.org/10.1093/jnci/djt095
84. Tai YC, Domchek S, Parmigiani G, Chen S (2007) Breast cancer risk among male BRCA1 and BRCA2 mutation carriers. J Natl Cancer Inst 99:1811–1814. https://doi.org/10.1093/jnci/djm203
85. Leongamornlert D, Mahmud N, Tymrakiewicz M et al (2012) Germline BRCA1 mutations increase prostate cancer risk. Br J Cancer 106:1697–1701. https://doi.org/10.1038/bjc.2012.146
86. Kote-Jarai Z, Leongamornlert D, Saunders E et al (2011) BRCA2 is a moderate penetrance gene contributing to young-onset prostate cancer: implications for genetic testing in prostate cancer patients. Br J Cancer 105:1230–1234. https://doi.org/10.1038/bjc.2011.383
87. Tedaldi G, Tebaldi M, Zampiga V et al (2020) Male breast cancer: results of the application of multigene panel testing to an Italian cohort of patients. Diagnostics (Basel, Switzerland) 10. https://doi.org/10.3390/diagnostics10050269
88. Thompson D, Easton DF, Breast Cancer Linkage Consortium (2002) Cancer incidence in BRCA1 mutation carriers. J Natl Cancer Inst 94:1358–1365. https://doi.org/10.1093/jnci/94.18.1358
89. Ghiorzo P, Pensotti V, Fornarini G et al (2012) Contribution of germline mutations in the BRCA and PALB2 genes to pancreatic cancer in Italy. Familial Cancer 11:41–47. https://doi.org/10.1007/s10689-011-9483-5
90. Iqbal J, Ragone A, Lubinski J et al (2012) The incidence of pancreatic cancer in BRCA1 and BRCA2 mutation carriers. Br J Cancer 107:2005–2009. https://doi.org/10.1038/bjc.2012.483
91. Ford D, Easton DF, Bishop DT et al (1994) Risks of cancer in BRCA1-mutation carriers. Breast Cancer Linkage Consortium. Lancet 343:692–695. https://doi.org/10.1016/s0140-6736(94)91578-4
92. Moran A, O'Hara C, Khan S et al (2012) Risk of cancer other than breast or ovarian in individuals with BRCA1 and BRCA2 mutations. Familial Cancer 11:235–242. https://doi.org/10.1007/s10689-011-9506-2
93. Breast Cancer Linkage Consortium (1999) Cancer risks in BRCA2 mutation carriers. J Natl Cancer Inst 91:1310–1316. https://doi.org/10.1093/jnci/91.15.1310
94. Fong PC, Boss DS, Yap TA et al (2009) Inhibition of poly(ADP-ribose) polymerase in tumors from BRCA mutation carriers. N Engl J Med 361:123–134. https://doi.org/10.1056/NEJMoa0900212
95. Sandhu SK, Schelman WR, Wilding G et al (2013) The poly(ADP-ribose) polymerase inhibitor niraparib (MK4827) in BRCA mutation carriers and patients with sporadic cancer: a phase 1 dose-escalation trial. Lancet Oncol 14:882–892. https://doi.org/10.1016/S1470-2045(13)70240-7

96. Cortesi L, Rugo HS, Jackisch C (2021) An overview of PARP inhibitors for the treatment of breast cancer. Target Oncol 16:255. https://doi.org/10.1007/s11523-021-00796-4
97. Audeh MW, Carmichael J, Penson RT et al (2010) Oral poly(ADP-ribose) polymerase inhibitor olaparib in patients with BRCA1 or BRCA2 mutations and recurrent ovarian cancer: a proof-of-concept trial. Lancet 376:245–251. https://doi.org/10.1016/S0140-6736(10)60893-8
98. Tutt A, Robson M, Garber JE et al (2010) Oral poly(ADP-ribose) polymerase inhibitor olaparib in patients with BRCA1 or BRCA2 mutations and advanced breast cancer: a proof-of-concept trial. Lancet (London, England) 376:235–244. https://doi.org/10.1016/S0140-6736(10)60892-6
99. Sandhu SK, Omlin A, Hylands L et al (2013) Poly (ADP-ribose) polymerase (PARP) inhibitors for the treatment of advanced germline BRCA2 mutant prostate cancer. Ann Oncol 24:1416–1418. https://doi.org/10.1093/annonc/mdt074
100. Golan T, Hammel P, Reni M et al (2019) Maintenance Olaparib for germline BRCA-mutated metastatic pancreatic cancer. N Engl J Med 381:317–327. https://doi.org/10.1056/NEJMoa1903387
101. Hodgson DR, Dougherty BA, Lai Z et al (2018) Candidate biomarkers of PARP inhibitor sensitivity in ovarian cancer beyond the BRCA genes. Br J Cancer 119:1401–1409. https://doi.org/10.1038/s41416-018-0274-8
102. Li A, Geyer FC, Blecua P et al (2019) Homologous recombination DNA repair defects in PALB2-associated breast cancers. NPJ Breast Cancer 5:23. https://doi.org/10.1038/s41523-019-0115-9
103. Konstantinopoulos PA, Spentzos D, Karlan BY et al (2010) Gene expression profile of BRCAness that correlates with responsiveness to chemotherapy and with outcome in patients with epithelial ovarian cancer. J Clin Oncol 28:3555–3561. https://doi.org/10.1200/JCO.2009.27.5719
104. Mateo J, Lord CJ, Serra V et al (2019) A decade of clinical development of PARP inhibitors in perspective. Ann Oncol 30:1437. https://doi.org/10.1093/annonc/mdz192
105. Wang Y, Zheng K, Huang Y et al (2021) PARP inhibitors in gastric cancer: beacon of hope. J Exp Clin Cancer Res 40:211. https://doi.org/10.1186/s13046-021-02005-6
106. Ilyas M, Tomlinson IP (1997) The interactions of APC, E-cadherin and beta-catenin in tumour development and progression. J Pathol 182:128–137. https://doi.org/10.1002/(SICI)1096-9896(199706)182:2<128::AID-PATH839>3.0.CO;2-Q
107. Dinarvand P, Davaro EP, Doan JV et al (2019) Familial adenomatous polyposis syndrome: an update and review of extraintestinal manifestations. Arch Pathol Lab Med 143:1382–1398. https://doi.org/10.5858/arpa.2018-0570-RA
108. Jasperson KW, Patel SG, Ahnen DJ (1998) APC-associated polyposis conditions. GeneReviews® (updated 2022)
109. Ma H, Brosens LAA, Offerhaus GJA et al (2018) Pathology and genetics of hereditary colorectal cancer. Pathology 50:49–59. https://doi.org/10.1016/j.pathol.2017.09.004
110. Knudsen AL, Bülow S, Tomlinson I et al (2010) Attenuated familial adenomatous polyposis: results from an international collaborative study. Color Dis 12:e243–e249. https://doi.org/10.1111/j.1463-1318.2010.02218.x
111. Nieuwenhuis MH, Vasen HFA (2007) Correlations between mutation site in APC and phenotype of familial adenomatous polyposis (FAP): a review of the literature. Crit Rev Oncol Hematol 61:153–161. https://doi.org/10.1016/j.critrevonc.2006.07.004
112. Lynch HT, Smyrk T, McGinn T et al (1995) Attenuated familial adenomatous polyposis (AFAP). A phenotypically and genotypically distinctive variant of FAP. Cancer 76:2427–2433. https://doi.org/10.1002/1097-0142(19951215)76:12<2427::aid-cncr2820761205>3.0.co;2-b
113. Garrean S, Hering J, Saied A et al (2008) Gastric adenocarcinoma arising from fundic gland polyps in a patient with familial adenomatous polyposis syndrome. Am Surg 74:79–83

114. Arnason T, Liang W-Y, Alfaro E et al (2014) Morphology and natural history of familial adenomatous polyposis-associated dysplastic fundic gland polyps. Histopathology 65:353–362. https://doi.org/10.1111/his.12393
115. Leone PJ, Mankaney G, Sarvapelli S et al (2019) Endoscopic and histologic features associated with gastric cancer in familial adenomatous polyposis. Gastrointest Endosc 89:961–968. https://doi.org/10.1016/j.gie.2018.12.018
116. Mankaney G, Leone P, Cruise M et al (2017) Gastric cancer in FAP: a concerning rise in incidence. Familial Cancer 16:371–376. https://doi.org/10.1007/s10689-017-9971-3
117. Slupska MM, Baikalov C, Luther WM et al (1996) Cloning and sequencing a human homolog (hMYH) of the Escherichia coli mutY gene whose function is required for the repair of oxidative DNA damage. J Bacteriol 178:3885–3892. https://doi.org/10.1128/jb.178.13.3885-3892.1996
118. Nielsen M, Infante E, Brand R (2012) MUTYH Polyposis. GeneReviews® (updated 2021)
119. Hurley JJ, Thomas LE, Walton S-J et al (2018) The impact of chromoendoscopy for surveillance of the duodenum in patients with MUTYH-associated polyposis and familial adenomatous polyposis. Gastrointest Endosc 88:665–673. https://doi.org/10.1016/j.gie.2018.04.2347
120. Spigelman AD, Williams CB, Talbot IC et al (1989) Upper gastrointestinal cancer in patients with familial adenomatous polyposis. Lancet (London, England) 2:783–785. https://doi.org/10.1016/s0140-6736(89)90840-4
121. Mishina Y, Starbuck MW, Gentile MA et al (2004) Bone morphogenetic protein type IA receptor signaling regulates postnatal osteoblast function and bone remodeling. J Biol Chem 279:27560–27566. https://doi.org/10.1074/jbc.M404222200
122. Massagué J (2012) TGFβ signalling in context. Nat Rev Mol Cell Biol 13:616–630. https://doi.org/10.1038/nrm3434
123. Larsen Haidle J, Howe JR (2003) Juvenile polyposis syndrome. GeneReviews® (updated 2022)
124. Latchford AR, Neale K, Phillips RKS, Clark SK (2012) Juvenile polyposis syndrome: a study of genotype, phenotype, and long-term outcome. Dis Colon Rectum 55:1038–1043. https://doi.org/10.1097/DCR.0b013e31826278b3
125. Aytac E, Sulu B, Heald B et al (2015) Genotype-defined cancer risk in juvenile polyposis syndrome. Br J Surg 102:114–118. https://doi.org/10.1002/bjs.9693
126. Ishida H, Ishibashi K, Iwama T (2018) Malignant tumors associated with juvenile polyposis syndrome in Japan. Surg Today 48:253–263. https://doi.org/10.1007/s00595-017-1538-2
127. Blatter R, Tschupp B, Aretz S et al (2020) Disease expression in juvenile polyposis syndrome: a retrospective survey on a cohort of 221 European patients and comparison with a literature-derived cohort of 473 SMAD4/BMPR1A pathogenic variant carriers. Genet Med 22:1524–1532. https://doi.org/10.1038/s41436-020-0826-1
128. MacFarland SP, Ebrahimzadeh JE, Zelley K et al (2021) Phenotypic differences in juvenile polyposis syndrome with or without a disease-causing SMAD4/BMPR1A variant. Cancer Prev Res (Phila) 14:215–222. https://doi.org/10.1158/1940-6207.CAPR-20-0348
129. Howe JR, Roth S, Ringold JC et al (1998) Mutations in the SMAD4/DPC4 gene in juvenile polyposis. Science 280:1086–1088. https://doi.org/10.1126/science.280.5366.1086
130. Ma C, Giardiello FM, Montgomery EA (2014) Upper tract juvenile polyps in juvenile polyposis patients: dysplasia and malignancy are associated with foveolar, intestinal, and pyloric differentiation. Am J Surg Pathol 38:1618–1626. https://doi.org/10.1097/PAS.0000000000000283
131. Aretz S, Stienen D, Uhlhaas S et al (2007) High proportion of large genomic deletions and a genotype phenotype update in 80 unrelated families with juvenile polyposis syndrome. J Med Genet 44:702–709. https://doi.org/10.1136/jmg.2007.052506
132. Brosens LAA, van Hattem A, Hylind LM et al (2007) Risk of colorectal cancer in juvenile polyposis. Gut 56:965–967. https://doi.org/10.1136/gut.2006.116913
133. Höfting I, Pott G, Stolte M (1993) The syndrome of juvenile polyposis. Leber Magen Darm 23(107–8):111–112

134. Howe JR, Mitros FA, Summers RW (1998) The risk of gastrointestinal carcinoma in familial juvenile polyposis. Ann Surg Oncol 5:751–756. https://doi.org/10.1007/BF02303487
135. Agnifili A, Verzaro R, Gola P et al (1999) Juvenile polyposis: case report and assessment of the neoplastic risk in 271 patients reported in the literature. Dig Surg 16:161–166. https://doi.org/10.1159/000018711
136. Goff V Transparency and disclosure: the route to accountability. Issue Brief (Inst Health Care Costs Solut) 2:1–12
137. Woodford-Richens K, Bevan S, Churchman M et al (2000) Analysis of genetic and phenotypic heterogeneity in juvenile polyposis. Gut 46:656–660. https://doi.org/10.1136/gut.46.5.656
138. Achatz MI, Porter CC, Brugières L et al (2017) Cancer screening recommendations and clinical Management of Inherited Gastrointestinal Cancer Syndromes in childhood. Clin Cancer Res 23:e107–e114. https://doi.org/10.1158/1078-0432.CCR-17-0790
139. Cohen S, Hyer W, Mas E et al (2019) Management of Juvenile Polyposis Syndrome in children and adolescents: a position paper from the ESPGHAN polyposis working group. J Pediatr Gastroenterol Nutr 68:453–462. https://doi.org/10.1097/MPG.0000000000002246
140. Xu X, Jin D, Durgan J, Hall A (2013) LKB1 controls human bronchial epithelial morphogenesis through p114RhoGEF-dependent RhoA activation. Mol Cell Biol 33:2671–2682. https://doi.org/10.1128/MCB.00154-13
141. McGarrity TJ, Amos CI, Baker MJ (2001) Peutz-Jeghers Syndrome. GeneReviews® (updated 2021)
142. Lim W, Olschwang S, Keller JJ et al (2004) Relative frequency and morphology of cancers in STK11 mutation carriers. Gastroenterology 126:1788–1794. https://doi.org/10.1053/j.gastro.2004.03.014
143. Giardiello FM, Brensinger JD, Tersmette AC et al (2000) Very high risk of cancer in familial Peutz-Jeghers syndrome. Gastroenterology 119:1447–1453. https://doi.org/10.1053/gast.2000.20228
144. Pezzolesi MG, Zbuk KM, Waite KA, Eng C (2007) Comparative genomic and functional analyses reveal a novel cis-acting PTEN regulatory element as a highly conserved functional E-box motif deleted in Cowden syndrome. Hum Mol Genet 16:1058–1071. https://doi.org/10.1093/hmg/ddm053
145. Yehia L, Eng C (2001) PTEN Hamartoma Tumor Syndrome. GeneReviews® (updated 2021)
146. Mester J, Eng C (2015) Cowden syndrome: recognizing and managing a not-so-rare hereditary cancer syndrome. J Surg Oncol 111:125–130. https://doi.org/10.1002/jso.23735
147. Tan M-H, Mester JL, Ngeow J et al (2012) Lifetime cancer risks in individuals with germline PTEN mutations. Clin Cancer Res 18:400–407. https://doi.org/10.1158/1078-0432.CCR-11-2283
148. Heald B, Mester J, Rybicki L et al (2010) Frequent gastrointestinal polyps and colorectal adenocarcinomas in a prospective series of PTEN mutation carriers. Gastroenterology 139:1927–1933. https://doi.org/10.1053/j.gastro.2010.06.061
149. Ngeow J, Eng C (2015) PTEN hamartoma tumor syndrome: clinical risk assessment and management protocol. Methods 77–78:11–19. https://doi.org/10.1016/j.ymeth.2014.10.011
150. Zbuk KM, Eng C (2007) Hamartomatous polyposis syndromes. Nat Clin Pract Gastroenterol Hepatol 4:492–502. https://doi.org/10.1038/ncpgasthep0902
151. Rosty C (2018) The role of the surgical pathologist in the diagnosis of gastrointestinal polyposis syndromes. Adv Anat Pathol 25:1–13. https://doi.org/10.1097/PAP.0000000000000173
152. Adams DR, Eng CM (2018) Next-generation sequencing to diagnose suspected genetic disorders. N Engl J Med 379:1353–1362. https://doi.org/10.1056/NEJMra1711801

Appendix E

3D chromatin architecture re-wiring at the *CDH3*/*CDH1* *loci* contributes to E-cadherin to P-cadherin expression switch in gastric cancer

Celina São José, Carla Pereira, Marta Ferreira, Ana André, Hugo Osório, Irene Gullo, Fátima Carneiro, and Carla Oliveira

Biology

2023

DOI: 10.3390/biology12060803

biology

MDPI

Article

3D Chromatin Architecture Re-Wiring at the *CDH3*/*CDH1* Loci Contributes to E-Cadherin to P-Cadherin Expression Switch in Gastric Cancer

Celina São José [1,2,3], Carla Pereira [1,2], Marta Ferreira [1,2,4], Ana André [1], Hugo Osório [1,2,5], Irene Gullo [1,2,5,6], Fátima Carneiro [1,2,5,6] and Carla Oliveira [1,2,5,*]

1 Instituto de Investigação e Inovação em Saúde, Universidade do Porto, 4200-135 Porto, Portugal; cjose@ipatimup.pt (C.S.J.); hosorio@i3s.up.pt (H.O.); igullo@ipatimup.pt (I.G.)
2 Institute of Molecular Pathology and Immunology of the University of Porto, 4200-135 Porto, Portugal
3 Doctoral Programme in Biomedicine, Faculty of Medicine, University of Porto, 4200-319 Porto, Portugal
4 Doctoral Program in Computer Sciences, Faculty of Sciences, University of Porto, 4169-007 Porto, Portugal
5 Faculty of Medicine, University of Porto, 4200-319 Porto, Portugal
6 Department of Pathology, Centro Hospitalar Universitário São João, 4200-319 Porto, Portugal
* Correspondence: carlaol@i3s.up.pt

Simple Summary: The cell architecture is created and maintained via adhesion molecules, called cadherins, that keep similar cells bound to each other. E-cadherin is a cadherin encoded by the *CDH1* gene, essential in normal epithelia. P-cadherin, encoded by the *CDH3* gene, exerts a similar function but is mainly present in tumours. E- to P-cadherin switch is a common event in several epithelial tumours. We unveiled a mechanism explaining the E- to P-cadherin switch (*CDH1* to *CDH3* switch), which we found to occur in many gastric cancers. We induced this switch by depleting *CDH1* in gastric cancer cells, and, as they started to produce *CDH3*, this rescued cell-adhesion. Despite this rescue, cells bearing the switch showed migratory and proliferative features, commonly observed in aggressive tumours. We observed that *CDH1* depletion leads to novel genomic interactions between the *CDH1* and *CDH3* gene sequences. We identified the sequences needed for these interactions to occur, which seem also responsible for controlling the amount of E- or P-cadherin molecules in the cell. Our data provide evidence that loss of E-cadherin is able to compromise the normal interactions between the *CDH1* to *CDH3* gene sequences, allowing the expression of P-cadherin in gastric cancers.

Abstract: Cadherins are cell–cell adhesion molecules, fundamental for cell architecture and polarity. E-cadherin to P-cadherin switch can rescue adherens junctions in epithelial tumours. Herein, we disclose a mechanism for E-cadherin to P-cadherin switch in gastric cancers. *CDH1* and *CDH3* mRNA expression was obtained from 42 gastric tumours' RNA-seq data. CRISPR-Cas9 was used to knock out *CDH1* and a putative regulatory element. *CDH1*-depleted and parental cells were submitted to proteomics and enrichment GO terms analysis; ATAC-seq/4C-seq with a *CDH1* promoter viewpoint to assess chromatin accessibility and conformation; and RT-PCR/flow cytometry to assess *CDH1*/E-cadherin and *CDH3*/P-cadherin expression. In 42% of gastric tumours analysed, *CDH1* to *CDH3* switch was observed. *CDH1* knockout triggered *CDH1*/E-cadherin complete loss and *CDH3*/P-cadherin expression increase at plasma membrane. This switch, likely rescuing adherens junctions, increased cell migration/proliferation, commonly observed in aggressive tumours. E- to P-cadherin switch accompanied increased *CDH1* promoter interactions with *CDH3*–eQTL, absent in normal stomach and parental cells. *CDH3*–eQTL deletion promotes *CDH3*/*CDH1* reduced expression. These data provide evidence that loss of *CDH1*/E-cadherin expression alters the *CDH3* locus chromatin conformation, allowing a *CDH1* promoter interaction with a *CDH3*-eQTL, and promoting *CDH3*/P-cadherin expression. These data highlight a novel mechanism triggering E- to P-cadherin switch in gastric cancer.

check for updates

Citation: São José, C.; Pereira, C.; Ferreira, M.; André, A.; Osório, H.; Gullo, I.; Carneiro, F.; Oliveira, C. 3D Chromatin Architecture Re-Wiring at the *CDH3/CDH1* Loci Contributes to E-Cadherin to P-Cadherin Expression Switch in Gastric Cancer. *Biology* **2023**, *12*, 803. https://doi.org/10.3390/biology12060803

Academic Editors: Alberto Orfao, Ilda Patrícia Ribeiro and Isabel Marques Carreira

Received: 18 April 2023
Revised: 25 May 2023
Accepted: 29 May 2023
Published: 31 May 2023

Copyright: © 2023 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (https://creativecommons.org/licenses/by/4.0/).

Keywords: gastric cancer; cadherins switch; regulatory elements; 3D chromatin architecture; chromatin interactions; CDH1; CDH3

1. Introduction

Cadherins are tissue-specific transmembrane glycoproteins that play a key role in stable cell–cell adhesion, cellular polarization, maintenance of tissue architecture, and embryogenesis [1,2]. E-cadherin and P-cadherin are encoded by *CDH1* and *CDH3* genes, respectively, and present a great degree of sequence and gene structure homology [3]. Despite their homology, both proteins display tissue-specific expression patterns and functions, being often mutually exclusive at the cellular level [3]. While E-cadherin is widely expressed in epithelial tissues [4], P-cadherin expression is restricted to organ regions, such as myoepithelial cells of the breast or mesothelia [5]. Whereas E-cadherin acts as a tumour suppressor in most epithelial tissues, P-cadherin de novo expression or upregulation has been linked to mammary cells' proliferation and migration in mice [6], tumour-promotion in gastric [7,8] and breast cancer [9,10], and worse prognosis in breast cancer [11–13].

Aberrant cadherin isoform switching occurs in tumours with implications for patients' survival [14,15]. E-cadherin to N-cadherin switch is a mark of epithelial to mesenchymal transition, in which differentiated epithelial cells, expressing E-cadherin, transition to dedifferentiated mesenchymal-like cells, expressing N-cadherin, and are prone to invade and metastasize [16]. Evidence of E-cadherin to P-cadherin switching has been observed in both gastric [17] and breast cancer, with implications for survival in the latter [11].

Mechanisms for E-cadherin to P-cadherin switching have been suggested, mainly through MiR-34a/WNT1 cascade [18], but fail to explain all observed switching phenotypes. As *CDH1* and *CDH3* are juxtaposed genes, derived from gene duplication, and both belong to the same topological-associating domain (TAD) [19], their expression is most likely co-regulated and dependent on the tissue-specific 3D chromatin architecture. The 3D chromatin architecture tightly regulates gene expression in time and space through physical interactions involving promoters and regulatory elements located in non-coding regions in their vicinity [20]. These regulatory networks are generally tissue-specific [21], and their impairment may interfere with the normal 3D chromatin architecture, leading to genes' misexpression patterns, causing or aggravating disease [22,23].

Herein, we aimed to characterize a potential 3D chromatin architecture re-wiring process, occurring at the *CDH3–CDH1* loci, that may contribute to E-cadherin to P-cadherin switch in gastric cancer.

2. Materials and Methods

2.1. Samples

Gastric cancers from 43 patients were selected for the study, displaying equivalent male to female ratio, range of age at diagnosis, distribution of tumour stage, and histological type. Clinicopathological features were collected from clinical records. The present project was approved by the Ethical Committee of Centro Hospitalar Universitário de São João on 16 March 2017, with internal reference CES072017.

2.2. RNA Extraction

Tumour areas with at least 75% tumour cells were selected for RNA extraction using MagMax FFPE DNA/RNA Ultra Kit (Applied Biosystems, Life technologies, Carlsbad, CA, USA), according to the manufacturer's instructions. Briefly, sections of paraffin-embedded tumour samples were de-paraffinized and protease K-digested. RNA was bonded to magnetic beads, pelleted against a magnetic stand, and the supernatant containing RNA washed and eluted. Concentrations were determined using Qubit® 2.0 Fluorometer (Invitrogen, Life Technologies Europe BV, Porto, Portugal) with RNA high sensitivity assay,

and integrity was assessed with a 2100 Bioanalyzer Instrument (Agilent Technologies, Instruments Soquimica, Lisboa, Portugal).

2.3. RNA-Seq and Whole Transcriptome Sequencing Bioinformatics Analysis

RNA was sequenced with Ion AmpliSeq™ Transcriptome Human Gene Expression Panel kit in an Ion Chef™ 550 NGS instrument. We obtained sequencing reads aligned to the GRCh37 human genome using TMAP [24] and annotated with USCS GRCh37. Only canonical transcripts were selected. *CDH1* and *CDH3* mRNA expression was considered positive for a threshold of 10 TPM or higher. Processed total RNA-seq from normal stomach tissue samples ENCSR853WOM, ENCSR471RUK, and ENCSR000AFI was collected from ENCODE [25–27].

2.4. Cell Lines and Culture

MKN74 human cell line was acquired from the JCRB Cell Bank, cultured in RPMI medium (GIBCO, Fisher Scientific, Porto Salvo, Portugal) supplemented with 10% fetal bovine serum (GIBCO, Fisher Scientific, Porto Salvo, Portugal) and maintained at 37 °C and 5% CO_2 in a high-humidity atmosphere. STR analysis was used to confirm cell identification. Cells were free from mycoplasma contamination.

2.5. Generation of CRISPR-Cas9 E-Cadherin Deletion

Single guided RNAs (sgRNAs) were designed to target the *CDH1* exon 2 and impair E-cadherin function, resourced using the Zhang Lab tool, currently available at Benchling online platform. Individual sgRNAs (Sigma-Aldrich, Merck Life Science, Darmstadt, Germany) (Supplementary Table S1) were cloned into pSpCas9(BB)-2A-Puro (PX459) V2.0 plasmid (Addgene 62988), in BbsI restriction site (New England Biolabs), according to the method of Ran and colleagues [28]. Each sgRNA-plasmid was transformed into stbl3 cells (Invitrogen), and colonies were screened by PCR and Sanger sequenced (Supplementary Table S1). sgRNAs were transfected in MKN74 cell line. Briefly, cells were seeded in 12-well plates and grown for 24 h, and pairs of plasmids were transfected with lipofectamine 3000 (Invitrogen), according to the manufacturer's instructions. Puromycin (Merck, Merck Life Science, Darmstadt, Germany) treatment started at 48 h and was renewed every 72 h, until non-transfected cells were dead.

2.6. Genotyping of Edited Clones

gDNA was extracted using NZY Tissue gDNA Isolation kit (NZYTech, Lisboa, Portugal) according to the manufacturer's protocol. gDNA was amplified with multiplex PCR kit (Qiagen, Germantown, MD, USA) and primers flanking the edition sites (Supplementary Table S1, Sigma-Aldrich). Edition was confirmed by Sanger, sequenced using BigDye Terminator v.3.1 cycle sequencing kit (Thermo Fisher Scientific, Waltham, MA, USA, EUA) on an ABI-3130 Genetic Analyzer (Applied Biosystems Europe B.V., Porto, Portugal).

2.7. mRNA and Protein Expression Analysis

CDH1/3 mRNA expression was assessed by qPCR in triplicate. Briefly, RNA was extracted using mirVana RNA Isolation Kit (Invitrogen), according to manufacturer's protocol. cDNA was synthesized using 1 µg of template RNA, and SuperScriptII reverse transcriptase (Invitrogen) was used to synthesize cDNA, according to the manufacturer's instructions. *CDH1* mRNA expression was analysed by qPCR with KAPA PROBE FAST qPCR Master Mix (2X) Kit (Sigma-Aldrich) and probes for *CDH1* (Hs.PT.58.3324071, IDT), *CDH3* (Hs.PT.51.5028751, IDT), and *18S* (custom assay, IDT) as endogenous control. Reactions were sequenced on a 7500 Real-Time PCR System (Applied Biosystems). Relative expression was normalized for the endogenous *18S* control and quantified using the $2^{-\Delta\Delta Ct}$ method [29].

E- and P-cadherin expression was assessed by flow cytometry in triplicate. Cells were blocked for 30 min with 3% bovine serum albumin–phosphatase buffer saline (BSA,

NZYTech) and incubated with primary mouse antibodies HECD-1 or P-cadherin (1:100 dilution; 1 h at 4 °C; Invitrogen), followed by secondary anti-mouse Alexa Fluor 647 antibody (1:500; 45 min at 4 °C; Invitrogen). FACS ARIA (BD Biosciences, Franklin Lakes, NJ, USA) was used to measure mean fluorescence, and Flow Jo version 10 software was used to analyse the data.

2.8. Statistical Analysis

GraphPad Prism version 7.00 software (GraphPad Software Inc, San Diego, CA, USA) was used to perform statistical analysis. Student's *t*-test was used for comparison analysis, assuming equal variance between clones and wild-type samples. Differences were considered significant for p-value <0.05.

2.9. Immunofluorescence

Cells were fixed for 20 min in 4% paraformaldehyde (Merck Life Science, Darmstadt, Germany) and blocked with 5% BSA (NZYTech) for 30 min. Cells were incubated with primary mouse antibody anti-E-cadherin (24E10) or anti-β-catenin (1:50; ON; 4 °C, Cell Signaling Technology, Danvers, MA, EUA) and primary rabbit monoclonal antibody anti-P-cadherin (#2130; 1:50; ON; 4 °C, Cell Signaling Technology), followed by secondary antibodies Alexa Fluor 488 Donkey Anti-Mouse (1:500; 60 min; Life Technologies) or Alexa Fluor 594 anti-rabbit (1:500, 60 min; Life Technologies). Cell nuclei were stained with DAPI (1 µg/mL in PBS; 5 min incubation; Sigma, Merck Life Science, Darmstadt, Germany). All coverslips were mounted using Vectashield mounting media (Vector Laboratories, Newark, CA, USA) and cells were analysed by fluorescence microscopy (Imager.Z1, AxioCam fluorescence microscope or Eclipse TE-2000, both from Zeiss, Oberkochen, Germany) using AxioVision software (Rockville, White Plains, NY, USA).

2.10. Protein Extraction

Cells were lysed with lysis buffer containing 1% (v/v) Triton X-100 and 1% (v/v) IGEPAL in PBS, supplemented with phosphatase (Sigma-Aldrich) and protease inhibitor cocktails (Roche, Mannheim, Germany) at 4 °C. Modified Bradford Assay (Bio-Rad, Hercules, CA, USA) was used to quantify total protein content of the samples.

2.11. Proteomics Sample Processing

Proteomics was performed according to solid-phase-enhanced sample-preparation (SP3) protocol [30]. Briefly, proteins were solubilized for 10 min with 100 mM Tris pH 8.5, 1% sodium deoxycholate, 10 mM tris(2-carboxyethyl) phosphine (TCEP), and 40 mM chloroacetamide at 95 °C at 1000 rpm. Proteins were digested with Trypsin/LysC (2 µg; Thermo Fisher Scientific, Waltham, MA, USA, EUA) overnight at 37 °C and 1000 rpm. Peptides were cleaned up and quantified.

2.12. Protein Identification and Quantification, and Enrichment GO Terms Analysis

Proteome Discoverer 2.1 software was used for identification (UniProt/SwissProt human database using Mascot (Version 2.5.1, Matrix Sciences)) and/or quantification of proteins. Carbamidomethylation of cysteines was set as fixed modification and methionine oxidation as variable modification, with trypsin as selected enzyme. High confidence master proteins with at least two unique peptides and without contaminants were considered. Differentially expressed proteins were selected considering adjusted p-value $\leq$0.05, abundance ratio $\leq$0.67 (downregulated) or $\geq$1.5 (upregulated), and concordance between replicates. ClusterProfiler (v.4.4.4) R package was used for enrichment GO terms analysis (q-value < 0.05). Statistics were performed using R (v.4.2.0).

2.13. ATAC-Seq

ATAC-seq data from normal stomach tissue were collected from phase 3 ENCODE project [31]. ATAC-seq libraries were generated from MKN74 cell line and *CDH1* deletion

clone, as described previously [32]. A total of 5 × 10^4 cells were lysed and chromatin was tagmented using Tn5 transposase and purified with MiniElute PCR purification kit (Qiagen), following the manufacturer's protocol. The appropriate number of PCR cycles was determined using SYBR Green qPCR. Transposed DNA fragments were amplified by PCR and purified using QIAquick PCR purification kit (Qiagen), following the manufacturer's protocol. Quality control of ATAC-seq libraries was accessed by tapestation, and samples were sequenced in Illumina HiSeqX technology according to standard protocols. ATAC-seq peaks were mapped resourcing to an ENCODE-DCC ATAC-seq bioinformatics pipeline with default parameters for adaptors trimming (cutadapt: -e 0.1 -m 5), alignment (bowtie), quality filtering for mapq < 30, removal of duplicates and mitochondrial reads (piccard), peak calling (macs2), and replicate statistical analysis (idr, p-value < 0.05).

2.14. C-Seq

Fresh gastric tissue from bariatric surgeries was collected and washed in HBSS 1x (GIBCO, Fisher Scientific, Porto Salvo, Portugal). Gastric mucosa cells were scraped and enzymatically dissociated with collagenase type I (Merck Life Science, Darmstadt, Germany), dispase (Merck), trypsin inhibitor (Sigma-Aldrish, Merck Life Science, Darmstadt, Germany), BSA (Nzytech, Lisboa, Portugal), dithiothreitol (Invitrogen Life Technologies Europe BV, Porto, Portugal), and HBSS 1× (GIBCO, Fisher Scientific, Porto Salvo, Portugal) at 37 °C and 150 rpm for at least 1 h. From these, 4C-seq libraries were generated, as previously described [33,34], containing 1 × 10^7 cells, cross-linked in 2% paraformaldehyde and lysed. Nuclei suspensions were digested with DnpII (New England Biolabs, EVRY cedex, Évry-Courcouronnes, France) as primary and Csp6I (New England Biolabs) as secondary restriction enzymes and re-circularized with T4 DNA Ligase (Thermo Fisher Scientific). Then, 4C-seq libraries were purified using Amicon Ultra-15 10 kDa (MWCO) (Millipore) and PCR-amplified with 3.2 µg per reaction (Supplementary Table S2). The *CDH1* viewpoint was designed to target chr16: 68737018-68737215, hg38. Samples were paired-end sequenced on Illumina HiSeqX technology (150 bp reads) according to standard protocols. *CDH1* interactions on a genome scale were mapped from sequenced 4C libraries, using a bioinformatics pipeline based on Pipe4C [34] and PeakC [35] with window size 2, alpha fdr 0.1, and minimal distance 500. Collection of bariatric surgeries' material was approved by the Ethical Committee of Centro Hospitalar Universitário de São João on 8 May 2020, with internal reference CE 305-19.

3. Results

3.1. CDH1 to CDH3 mRNA Expression Switch Is a Frequent Event in Gastric Cancer

To understand the frequency of and explore potential mechanisms which underly E- to P-cadherin switching in gastric cancer, we started by analysing RNA-seq data from 43 gastric cancers to evaluate *CDH1* and *CDH3* mRNA expression. We found that *CDH1* downregulation and *CDH3* upregulation are the most frequent combination of events in this cohort, occurring in 42% of all cases, followed by expression of both cadherins (33%) (Figure 1A). Downregulation of both cadherins occurred in 19% of tumours, while only 7% of all tumours expressed *CDH1* and lacked *CDH3* expression (Figure 1A). In this cohort, 75% of the tumours expressed *CDH3* (Figure 1A, upper panels), a feature that is acquired in tumours, as normal stomach epithelial cells do not express *CDH3* (Figure 1B). In 42% of all cases, there was a *CDH1* to *CDH3* expression switch (Figure 1A, upper left panel).

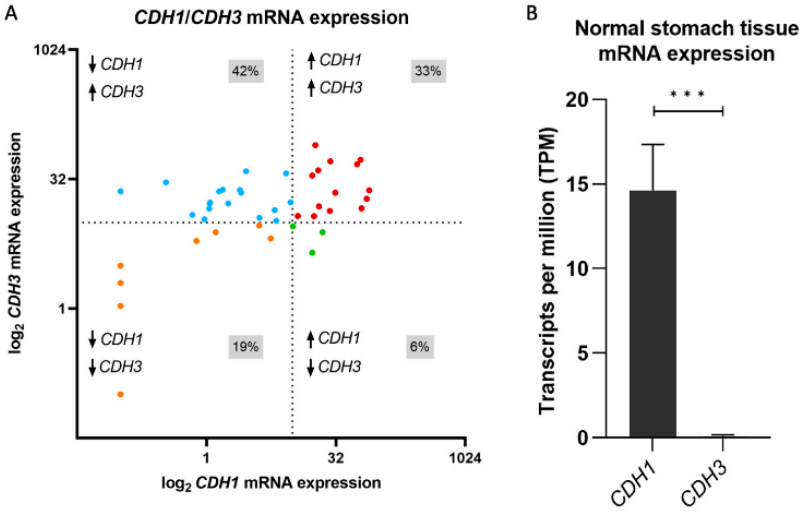


Figure 1. *CDH1* and *CDH3* mRNA expression in gastric tissues. (**A**) Distribution of *CDH1* and *CDH3* mRNA expression in gastric cancer tumours. Blue dots represent gastric cancers with low *CDH1* and high *CDH3* expression, red dots represent gastric cancers with high expression for both *CDH1* and *CDH3*, orange dots represent gastric cancers with low *CDH1* and *CDH3* expression, green dots represent gastric cancers with high *CDH1* and low *CDH3* expression. Dashed lines represent 10 TPM, the threshold for low or high expression. Data shown in transcripts per million (TPM). (**B**) *CDH1* and *CDH3* mRNA expression in stomach normal tissue from ENCODE [25–27]. Data shown in transcripts per million (TPM) (*** p value ≤ 0.001).

3.2. E-Cadherin Loss of Function Triggers E-Cadherin to P-Cadherin Switch in Gastric Cancer

Given the abundance of gastric cancers bearing simultaneous loss of *CDH1* expression and *CDH3* upregulation, we tested whether depleting *CDH1* would trigger *CDH3* upregulation.

We used sgRNAs flanking *CDH1* exon 2 (Supplementary Table S1), and successful removal of exon 2 was confirmed by Sanger sequencing in the MKN74 cell line (Supplementary Table S1). This clone carried a homozygous deletion encompassing the *CDH1* exon 2 (Figure 2A) and presented complete loss of both *CDH1* mRNA and E-cadherin protein expression (Figure 2B). As a consequence of this deletion, both *CDH3* mRNA and P-cadherin protein expression increased by 60% (p-value = 0.05 and p-value = 0.037, respectively) (Figure 2C). The immunofluorescence data highlight the P-cadherin expression increase at the plasma membrane upon E-cadherin expression loss, likely rescuing adherens junctions (Figure 2D). Indeed, P-cadherin co-localized with adhesion partner β-catenin at the cell membrane in the *CDH1 deletion* clone, while the parental cell line displayed lower (and heterogeneous) P-cadherin, as was the case in the cytoplasm (Figure 2D). Altogether, these results support that depletion of E-cadherin, due to gene deletions, can trigger E- to P-cadherin switch.

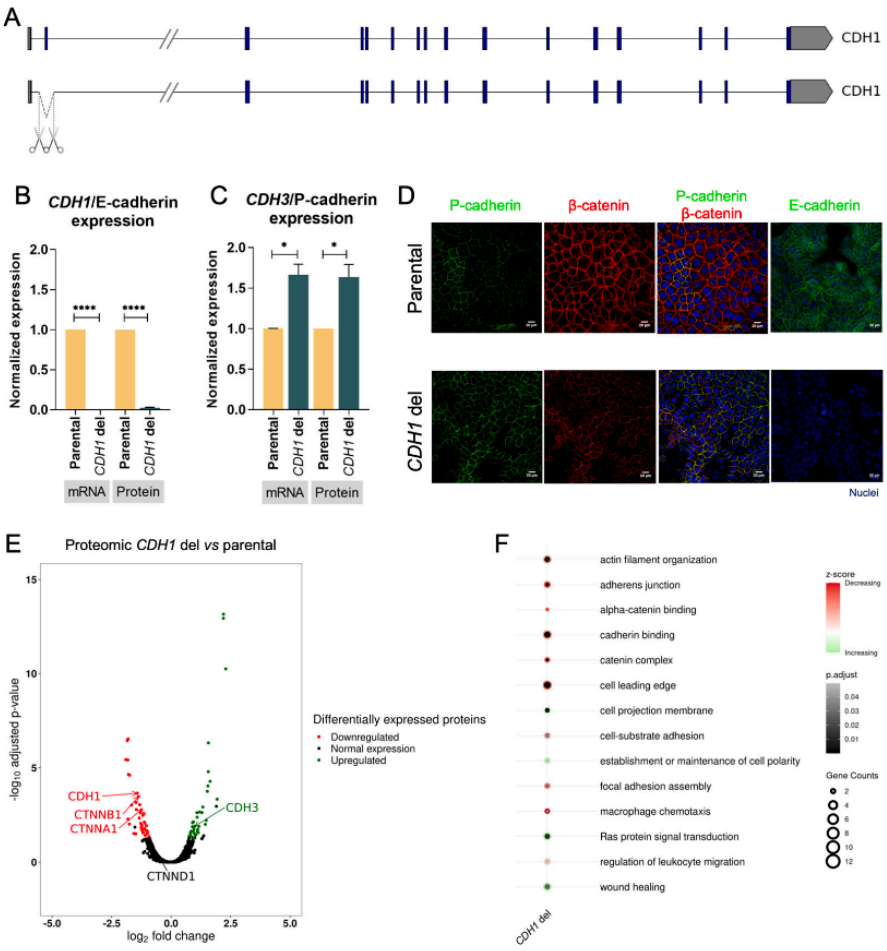


Figure 2. Characterization of E- to P-cadherin switch. (**A**) CRISPR-Cas9 design strategy. (**B**) Characterization of *CDH1* deletion in *CDH1*/E-cadherin mRNA and protein expression. Values represented as mean ± SEM. Protein expression represented as median fluorescence intensity (MFI) (**** p value ≤ 0.0001). (**C**) Characterization of *CDH1* deletion in *CDH3*/P-cadherin mRNA and protein expression. Values represented as mean ± SEM. Protein expression represented as median fluorescence intensity (MFI) (* p value ≤ 0.05). (**D**) Immunofluorescence of parental cells (top) and *CDH1* deletion cells (down) for P-cadherin, β-catenin, and E-cadherin. Nuclei were stained with DAPI (blue), and white scale bars represent a distance of 20 µm. (**E**) Volcano plot illustrating differentially expressed proteins in *CDH1* deletion vs. parental cells. (**F**) Enrichment analysis of biological processes, molecular function, and cellular components in upregulated and downregulated proteins.

3.3. E-Cadherin Expression Loss Triggers Downregulation of Adhesion Complex Partners and Adhesion-Related Pathways

To further dissect the impact of E-cadherin loss, we used proteomics to evaluate the differential proteome between the parental cell line and the *CDH1* deletion clone. A total of 4579 proteins were identified, from which 41 were downregulated and 39 were upregulated in the *CDH1* deletion clone (Supplementary Table S3). The E- to P-cadherin switch was confirmed in the *CDH1* deletion clone, with downregulation of the former (0.37-fold, p-value = 0.0002) and upregulation of the latter (1.99-fold, p-value = 0.018) (Figure 2D). In addition to E- and P-cadherin expression, other adhesion complex partners became impaired in the *CDH1* deletion clone, namely, *CTNNA1* (αE-catenin) and *CTNNB1*

(β-catenin), while others remain unchanged, namely, *CTNND1* (p120) (Figure 2D). Gene ontology analysis revealed several adhesion-associated pathways downregulated upon *CDH1* deletion (Figure 2E, Supplementary Table S4), encompassing 'adherens junction', 'cadherin binding', 'focal adhesion assembly', and catenin- and actin-associated pathways (Figure 2E, Supplementary Table S4). In contrast, the 'establishment and maintenance of cell polarity' pathway became upregulated, likely revealing the rescue by P-cadherin. Events associated with migration, such as 'cell projection membrane' and 'wound healing', were found upregulated in cells with E- to P-cadherin switch (Figure 2E, Supplementary Table S4). Downregulation of immune-associated pathways 'macrophage chemotaxis' and 'regulation of leucocyte migration' and upregulation of oncogenic pathway 'Ras protein signal transduction' were also found upon *CDH1* deletion (Figure 2E, Supplementary Table S4).

3.4. D Chromatin Architecture Re-Wiring at the CDH3–CDH1 Loci Contributes to E-Cadherin to P-Cadherin Switch in Gastric Cancer

In an attempt to identify the mechanism triggering the E- to P-cadherin switch, upon *CDH1* depletion, we analysed regulatory elements within the *CDH3–CDH1* loci that may control expression of cadherins in the stomach. We mined the Ensembl database and found an eQTL located within the *CDH3* intron 2 (chr16: 68661867-68662578 and chr16: 68663306-68670017, hg38) (Figure 3A). Normal epithelial cells from bariatric surgeries, MKN74 parental cells as well as *CDH1* deletion cells, were profiled for *CDH1*-promoter interactions with the eQTL using 4C-seq, and for accessibility of chromatin using ATAC-seq. Normal stomach accessible chromatin was also collected from the ENCODE project and analysed [31]. In normal stomach, which expresses *CDH1* but not *CDH3*, the chromatin was not accessible at the *CDH3* promoter nor at the *CDH3*-eQTL, but was accessible at the *CDH1* promoter (Figure 3A, ATAC-seq normal stomach). Neither the *CDH3* promoter nor the *CDH3*–eQTL interacted with the *CDH1* promoter (Figure 3A, 4C-seq normal stomach). In contrast, upon *CDH1* deletion, availability of the chromatin increased around the eQTL sequence and was maintained as available at the *CDH1* and *CDH3* promoters (Figure 3A, ATAC-seq parental vs. ATAC-seq *CDH1* deletion). Additionally, new interactions were established between the *CDH1* promoter and the *CDH3*–eQTL region (Figure 3A, 4C-seq parental vs. 4C-seq *CDH1* deletion). These chromatin architecture changes seem to indicate that, upon *CDH1*/E-cadherin depletion, the *CDH1* promoter interacts with the *CDH3* eQTL, potentially leading to *CDH3*/P-cadherin expression increase (Figure 2B–D). If this is the case, deleting the *CDH3*–eQTL in a cell line expressing both *CDH3* and *CDH1* could provide information on the expression regulatory role of the *CDH3*-eQTL. To address this, we CRISPR-Cas9 deleted the eQTL in parental MKN74 cells. This led to a 50% decrease in both *CDH3* and *CDH1* mRNA expression (p-value = 0.0196 and p-value = 0.0084, respectively), and a 30% decrease in P-cadherin and E-cadherin protein expression (p-value = 0.035 and p-value = 0.0486, respectively) as compared to parental MKN74 cells (Figure 3B). These results pinpoint *CDH3*–eQTL as a potential enhancer of both *CDH3* and *CDH1* in this cell line. Altogether, these data provide evidence that loss of *CDH1*/E-cadherin expression alters the chromatin conformation at the *CDH3* locus, allowing interaction of the *CDH1* promoter with the *CDH3*–eQTL, which seems to potentiate *CDH3*/P-cadherin expression.

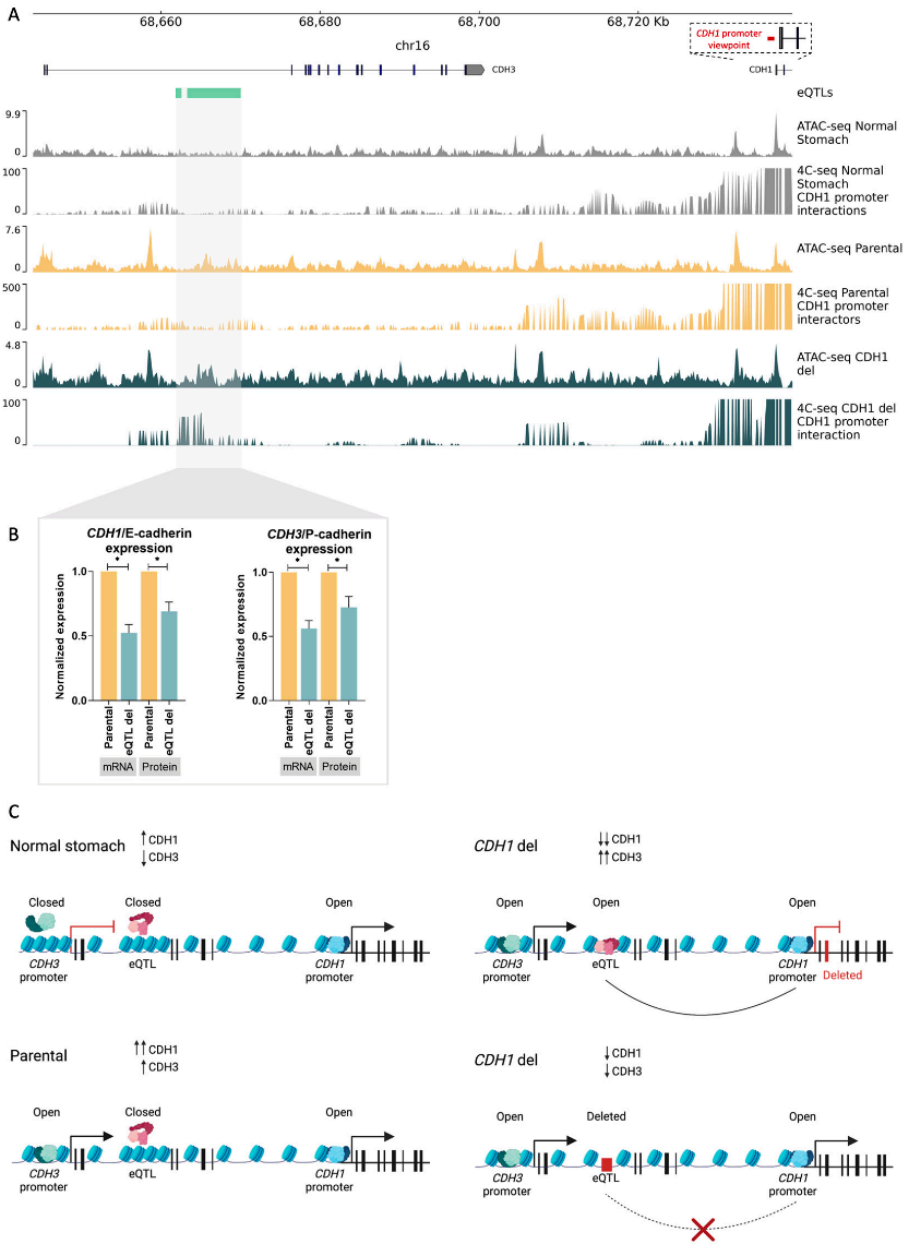


Figure 3. The 3D chromatin architecture in E- to P-cadherin switch. (**A**) Accessible chromatin and *CDH1*-promoter interactors in normal stomach tissue, parental MKN74 cells, and *CDH1* deletion clone. *CDH1* promoter viewpoint is shown in red. (**B**) Characterization of *CDH3*–eQTL deletion in *CDH1*/E-cadherin and *CDH3*/P-cadherin mRNA, and protein expression. Values represented as mean ± SEM. Protein expression represented as median fluorescence intensity (MFI) (* p value ≤ 0.05). (**C**) Model of the role of 3D chromatin architecture in E- to P-cadherin switch. Red rectangles represent deletions.

4. Discussion

Herein, we aimed to characterize a potential 3D chromatin architecture re-wiring process, occurring at the *CDH3*–*CDH1* loci, that may explain the *CDH1*/E-cadherin to

CDH3/P-cadherin switch observed in gastric cancer. First, we demonstrated that 75% of the gastric cancers analysed overexpressed *CDH3*. Additionally, we observed that, in 42% of all cases, *CDH3* overexpression occurred concomitantly with *CDH1* downregulation, representing an E- to P-cadherin switch. We also found that an eQTL within the *CDH3* intron 2, acting as a regulatory element, was able to modulate both *CDH1* and *CDH3* expression, and that it interacted with the *CDH1* promoter upon *CDH1* loss of expression. To the best of our knowledge, this body of results highlights a novel mechanism which triggers E- to P-cadherin switch in gastric cancer.

Our data suggest that E- to P-cadherin switch is a common event in gastric cancers. The impact of P-cadherin overexpression in cancer is still controversial, as this molecule seems to act as tumour-promoting and a marker of a bad prognosis in several cancer types—namely, in breast, glioblastoma, and gastric cancer [36]—but may also act as a marker of a good prognosis—namely, in colon cancer [37]. Breast cancer cells presenting E- to P-cadherin switch, and those expressing both cadherins, reveal increased migration capacity in vitro [38], while breast cancer epithelial cells expressing E-cadherin and lacking P-cadherin expression, resembling the normal breast, and cells lacking both cadherins display reduced migration [38].

Herein, we created a model of E- to P-cadherin switch by using CRISPR-Cas9 editing to deplete *CDH1*. This model presented reduced expression of adhesion complex partners, but maintained β-catenin expression at adherens junctions, which co-localized with P-cadherin at the cell membrane. This likely represents an adherens junctions rescue by P-cadherin in E-cadherin-depleted cells, as previously reported [39]. Evidence for E- to P-cadherin switch has been linked to invasive lobular breast cancer with tubular elements, which show reduced E-cadherin expression, normal β-catenin expression, and P-cadherin upregulation, highlighting rescue of adherens junctions in breast tubular elements when E-cadherin is lost [39].

Our data on differential protein expression and gene ontology analysis in the *CDH1* deletion clone, in comparison to an isogenic parental *CDH1* cell line, highlighted downregulation of adhesion-, catenin- and actin-related pathways, as well as upregulation of cell polarity, all associated with E-cadherin main functions. Despite the fact that E-cadherin deficiency weakens cell-adhesion, cell polarity regulation further supports P-cadherin rescue of adherens junctions (Christgen et al. 2020). Our gene ontology data also support increased migration potential of cells with E- to P-cadherin switch, namely, through activation of the 'wound healing' pathway, also described for breast cancer cells [38]. Our findings also suggest that cells with E- to P-cadherin switch may have increased proliferation through the activation of Ras oncogenic signalling pathways. As in many published reports, the abovementioned findings support that E- to P-cadherin switch is associated with more aggressive tumour phenotypes.

In the present study, we used the abovementioned cell model to identify the mechanism responsible for the E- to P-cadherin switch. We raised the hypothesis that regulatory elements in the *CDH3*/*CDH1* loci could play a role in the observed E- to P-cadherin expression switch, as gene expression is tightly regulated through the physical proximity enabled by the 3D-chromatin architecture (Andrey and Mundlos 2017) [20]. We demonstrated that, upon *CDH1* depletion triggered by a gene deletion, the chromatin re-shuffles at the *CDH3*/*CDH1* loci, and its promoter becomes physically bound to an eQTL overlapping accessible chromatin located in *CDH3* intron 2; this associates with increased *CDH3*/P-cadherin expression. The absence of interactions between the *CDH1* promoter and this eQTL in normal stomach epithelial tissue may explain why *CDH3* is not expressed in this particular tissue. While these data provide a rationale for E- to P-cadherin switch in gastric cancer, they do not inform whether the eQTL promotes *CDH3* expression. To address this question, we deleted the eQTL in a tumour cell line expressing wild-type E- and P-cadherins, preventing it from interacting with the *CDH1* promoter. This resulted in downregulation of both P- and E-cadherin expression and demonstrated that the *CDH3*–eQTL controlled P-cadherin expression. Taken together, these data provide evidence that

loss of *CDH1*/E-cadherin expression alters the chromatin conformation at the *CDH3* locus, allowing interaction of the *CDH1* promoter with the *CDH3*–eQTL, which seems to potentiate *CDH3*/P-cadherin expression. Our findings also suggest that the *CDH3*–eQTL acts as an enhancer regulating both P- and E-cadherin protein expression, thus regulating P-cadherin expression in the switch mechanism. Similar findings have been observed for the ZRS enhancer located within the *Lmbr1* gene, also controlling *Shh* gene expression [40].

Single-cell sorting of CRISPR-Cas9 clones herein described only yielded one homozygous clone for each set of sgRNAs harbouring the intended deletions, which we acknowledge as a study limitation.

5. Conclusions

Overall, this study provides insights into a noncoding regulatory element that controls E- and P-cadherin expression and triggers E- to P-cadherin switch in the presence of a *CDH1* depletion, through a re-shuffle of the *CDH1*/*CDH3* loci.

Supplementary Materials: The following supporting information can be downloaded at https://www.mdpi.com/article/10.3390/biology12060803/s1, Supplementary Table S1: sgRNAs and primers used for CRISPR-Cas9 and genotyping; Supplementary Table S2: Primers used for 4C-seq; Supplementary Table S3: Differentially expressed proteins in *CDH1* exon 2 del vs. parental cells; Supplementary Table S4: Enriched GO terms of differentially expressed proteins in *CDH1* exon 2 del vs. parental cells.

Author Contributions: Conceptualization, C.S.J., C.P., and C.O.; Formal analysis, C.S.J., M.F., H.O., and C.O.; Funding acquisition, F.C. and C.O.; Bioinformatics analysis, M.F.; Methodology, C.S.J., C.P., A.A., H.O., I.G., and F.C.; Project administration, C.O.; Supervision, C.O.; Writing—original draft, C.S.J.; Writing—review and editing, C.P., M.F., and C.O. All authors have read and agreed to the published version of the manuscript.

Funding: This work received funding from (1) FEDER (Fundo Europeu de Desenvolvimento Regional) funds through the COMPETE 2020—Operacional Programme for Competitiveness and Internationalisation (POCI), Portugal 2020 and by Fundação para a Ciência e a Tecnologia (FCT): "Solving the 3D Chromatin Structure of CDH1 Locus to Identify Disease-Associated Mechanisms (3DChrome)", PTDC/BTM-TEC/30164/2017; POCI-01-0145-FEDER-030164 and "Life-time Risk Estimations and Genetic Modifiers Of Hereditary Diffuse Gastric Cancer (LEGOH)", PTDC/BTM-TEC/6706/2020; (2) European Union's Horizon 2020 research and innovation programme under grant agreement No 779257: "Solve-RD"; (3) COMPETE 2020—POCI, Lisboa Portugal Regional Operational Programme (Lisboa2020), Algarve Portugal Regional Operational Programme (CRESC Algarve2020), under the PORTUGAL 2020 Partnership Agreement, through the European Regional Development Fund (ERDF), and by FCT: "GenomePT", POCI-01-0145-FEDER-022184; (4) NORTE2020 and FEDER funds through the COMPETE 2020—POCI and Portugal 2020: "Diabetes & obesity at the crossroads between Oncological and Cardiovascular diseases—a system analysis NETtwork towards precision medicine (DOCnet)", NORTE-01-0145-FEDER-000003; (5) Portuguese Mass Spectrometry Network, integrated in the National Roadmap of Research Infrastructures of Strategic Relevance (ROTEIRO/0028/2013; LISBOA-01-0145-FEDER-022125); (6) FCT with salary support for PhD studentships to C.S.J. (Ref. SFRH/BD/140796/2018), C.P. (Ref. SFRH/BD/113031/2015), and M.F. (Ref. 2020.05763.BD).

Institutional Review Board Statement: The study was conducted in accordance with the Declaration of Helsinki and approved by the Ethics Committee of Centro Hospitalar Universitário de São João (CES072017 on 16 March 2017 and CE 305-19 on 8 May 2020).

Informed Consent Statement: Informed consent was obtained from all subjects involved in the study.

Data Availability Statement: Raw data supporting the findings of this study are available upon request from the corresponding author. Processed data are available in Supplementary Materials.

Acknowledgments: This work was supported by: (1) Doctoral Programme in Biomedicine—Faculty of Medicine, University of Porto, to C.S.J. and C.P.; (2) Doctoral Programme in Computational Sciences, Faculty of Sciences, University of Porto, to M.F. We would like to thank Rita Barbosa-Matos for performing E-cadherin immunohistochemistry.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Ivanov, D.B.; Philippova, M.P.; Tkachuk, V.A. Structure and Functions of Classical Cadherins. *Biochemistry* **2001**, *66*, 1174–1186. [CrossRef] [PubMed]
2. Bhatt, T.; Rizvi, A.; Batta, S.P.R.; Kataria, S.; Jamora, C. Signaling and Mechanical Roles of E-cadherin. *Cell Commun. Adhes.* **2013**, *20*, 189–199. [CrossRef]
3. Nollet, F.; Koolsa, P.; Van Roy, F. Phylogenetic analysis of the cadherin superfamily allows identification of six major subfamilies besides several solitary members. *J. Mol. Biol.* **2000**, *299*, 551–572. [CrossRef] [PubMed]
4. Cao, J.; O'day, D.R.; Pliner, H.A.; Kingsley, P.D.; Deng, M.; Daza, R.M.; Zager, M.A.; Aldinger, K.A.; Blecher-Gonen, R.; Zhang, F.; et al. A human cell atlas of fetal gene expression. *Science* **2020**, *370*, eaba7721. [CrossRef]
5. Daniel, C.W.; Strickland, P.; Friedmann, Y. Expression and Functional Role of E- and P-Cadherins in Mouse Mammary Ductal Morphogenesis and Growth. *Dev. Biol.* **1995**, *169*, 511–519. [CrossRef] [PubMed]
6. Park, J.; Park, E.; Han, S.-W.; Im, S.-A.; Kim, T.-Y.; Kim, W.-H.; Oh, D.-Y.; Bang, Y.-J. Down-regulation of P-cadherin with PF-03732010 inhibits cell migration and tumor growth in gastric cancer. *Investig. New Drugs* **2012**, *30*, 1404–1412. [CrossRef]
7. Kim, M.A.; Jung, E.J.; Lee, H.S.; Lee, H.E.; Yang, H.-K.; Oh, D.-Y.; Bang, Y.-J.; Kim, W.H. P-cadherin expression in gastric carcinoma: Its regulation mechanism and prognostic significance. *Hum. Pathol.* **2010**, *41*, 877–885. [CrossRef]
8. Eom, B.W.; Ryu, K.W.; Yoon, H.M.; Kook, M.-C. Predictive value of E-cadherin and EpCAM for detection of metastatic lymph node in early gastric cancer. *Chin. J. Cancer Res.* **2020**, *32*, 614–620. [CrossRef]
9. Liu, N.; Yu, Q.; Liu, T.J.; Gebreamlak, E.P.; Wang, S.L.; Zhang, R.J.; Zhang, J.; Niu, Y. P-cadherin expression and basal-like subtype in breast cancers. *Med. Oncol.* **2012**, *29*, 2606–2612. [CrossRef]
10. Margan, M.M.; Cimpean, A.M.; Ceausu, A.R.; Raica, M. Differential Expression of E-Cadherin and P-Cadherin in Breast Cancer Molecular Subtypes. *Anticancer Res.* **2020**, *40*, 5557–5566. [CrossRef]
11. Turashvili, G.; E McKinney, S.; Goktepe, O.; Leung, S.C.; Huntsman, D.G.; A Gelmon, K.; Los, G.; A Rejto, P.; Aparicio, S.A.J.R. P-cadherin expression as a prognostic biomarker in a 3992 case tissue microarray series of breast cancer. *Mod. Pathol.* **2011**, *24*, 64–81. [CrossRef]
12. Faria, G.; Cardoso, M.J.; Martins, D.; Bettencourt, H.; Amendoeira, I.; Schimitt, F. P-cadherin as prognostic factor for loco-regional relapse in breast cancer. *Acta Med. Port.* **2012**, *25*, 97–105.
13. Sridhar, S.; Rajesh, C.; Jishnu, P.V.; Jayaram, P.; Kabekkodu, S.P. Increased expression of P-cadherin is an indicator of poor prognosis in breast cancer: A systematic review and meta-analysis. *Breast Cancer Res. Treat.* **2020**, *179*, 301–313. [CrossRef] [PubMed]
14. Gravdal, K.; Halvorsen, O.J.; Haukaas, S.A.; Akslen, L.A. A Switch from E-Cadherin to N-Cadherin Expression Indicates Epithelial to Mesenchymal Transition and Is of Strong and Independent Importance for the Progress of Prostate Cancer. *Clin. Cancer Res.* **2007**, *13*, 7003–7011. [CrossRef]
15. Appolloni, I.; Barilari, M.; Caviglia, S.; Gambini, E.; Reisoli, E.; Malatesta, P. A cadherin switch underlies malignancy in high-grade gliomas. *Oncogene* **2015**, *34*, 1991–2002. [CrossRef]
16. Araki, K.; Shimura, T.; Suzuki, H.; Tsutsumi, S.; Wada, W.; Yajima, T.; Kobayahi, T.; Kubo, N.; Kuwano, H. E/N-cadherin switch mediates cancer progression via TGF-beta-induced epithelial-to-mesenchymal transition in extrahepatic cholangiocarcinoma. *Br. J. Cancer* **2011**, *105*, 1885–1893. [CrossRef]
17. Barber, M.; Murrell, A.; Ito, Y.; Maia, A.-T.; Hyland, S.; Oliveira, C.; Save, V.; Carneiro, F.; Paterson, A.; Grehan, N.; et al. Mechanisms and sequelae of E-cadherin silencing in hereditary diffuse gastric cancer. *J. Pathol.* **2008**, *216*, 295–306. [CrossRef]
18. Li, B.; Guo, X.; Li, N.; Chen, Q.; Shen, J.; Huang, X.; Huang, G.; Wang, F. WNT1, a target of miR-34a, promotes cervical squamous cell carcinoma proliferation and invasion by induction of an E-P cadherin switch via the WNT/β-catenin pathway. *Cell. Oncol.* **2020**, *43*, 489–503. [CrossRef] [PubMed]
19. D'Antonio, M.; Guerra, R.F.; Cereda, M.; Marchesi, S.; Montani, F.; Nicassio, F.; Di Fiore, P.P.; Ciccarelli, F.D. Recessive Cancer Genes Engage in Negative Genetic Interactions with Their Functional Paralogs. *Cell Rep.* **2013**, *5*, 1519–1526. [CrossRef] [PubMed]
20. Andrey, G.; Mundlos, S. The three-dimensional genome: Regulating gene expression during pluripotency and development. *Development* **2017**, *144*, 3646–3658. [CrossRef]
21. Will, A.J.; Cova, G.; Osterwalder, M.; Chan, W.L.; Wittler, L.; Brieske, N.; Heinrich, V.; de Villartay, J.P.; Vingron, M.; Klopocki, E.; et al. Composition and dosage of a multipartite enhancer cluster control developmental expression of Ihh (Indian hedgehog). *Nat Genet.* **2017**, *49*, 1539–1545. [CrossRef] [PubMed]
22. Lupiáñez, D.G.; Kraft, K.; Heinrich, V.; Krawitz, P.; Brancati, F.; Klopocki, E.; Horn, D.; Kayserili, H.; Opitz, J.M.; Laxova, R.; et al. Disruptions of Topological Chromatin Domains Cause Pathogenic Rewiring of Gene-Enhancer Interactions. *Cell* **2015**, *161*, 1012–1025. [CrossRef] [PubMed]
23. Spielmann, M.; Lupiáñez, D.G.; Mundlos, S. Structural variation in the 3D genome. *Nat. Rev. Genet.* **2018**, *19*, 453–467. [CrossRef] [PubMed]
24. Tennekes, M. tmap: Thematic Maps in R. *J. Stat. Softw.* **2018**, *84*, 1–39. [CrossRef]
25. Feingold, E.A.; Good, P.J.; Guyer, M.S.; Kamholz, S.; Liefer, L.; Wetterstrand, K.; Collins, F.S.; Gingeras, T.R.; Kampa, D.; Sekinger, E.A.; et al. The ENCODE (ENCyclopedia Of DNA Elements) Project. *Science* **2004**, *306*, 636–640. [CrossRef]

26. Zhang, X.O.; Gingeras, T.R.; Weng, Z. Genome-wide analysis of polymerase III–transcribed Alu elements suggests cell-type–specific enhancer function. *Genome Res.* **2019**, *29*, 1402–1414. [CrossRef]
27. Lee, D.; Zhang, J.; Liu, J.; Gerstein, M. Epigenome-based splicing prediction using a recurrent neural network. *PLoS Comput. Biol.* **2020**, *16*, e1008006. [CrossRef]
28. Ran, F.A.; Hsu, P.D.; Wright, J.; Agarwala, V.; Scott, D.A.; Zhang, F. Genome engineering using the CRISPR-Cas9 system. *Nat. Protoc.* **2013**, *8*, 2281–2308. [CrossRef]
29. Livak, K.J.; Schmittgen, T.D. Analysis of relative gene expression data using real-time quantitative PCR and the $2^{-\Delta\Delta CT}$ Method. *Methods* **2001**, *25*, 402–408. [CrossRef]
30. Hughes, C.S.; Moggridge, S.; Müller, T.; Sorensen, P.H.; Morin, G.B.; Krijgsveld, J. Single-pot, solid-phase-enhanced sample preparation for proteomics experiments. *Nat. Protoc.* **2019**, *14*, 68–85. [CrossRef]
31. Moore, J.E.; Purcaro, M.J.; Pratt, H.E.; Epstein, C.B.; Shoresh, N.; Adrian, J.; Kawli, T.; Davis, C.A.; Dobin, A. Expanded encyclopaedias of DNA elements in the human and mouse genomes. *Nature* **2020**, *583*, 699–710. [CrossRef]
32. Buenrostro, J.D.; Wu, B.; Chang, H.Y.; Greenleaf, W.J. ATAC-seq: A Method for Assaying Chromatin Accessibility Genome-Wide. *Curr. Protoc. Mol. Biol.* **2015**, *109*, 21.29.1–21.29.9. [CrossRef] [PubMed]
33. Van de Werken, H.J.; de Vree, P.J.; Splinter, E.; Holwerda, S.J.; Klous, P.; de Wit, E.; de Laat, W. 4C technology: Protocols and data analysis. *Methods Enzym.* **2012**, *513*, 89–112. [CrossRef]
34. Krijger, P.H.; Geeven, G.; Bianchi, V.; Hilvering, C.R.; de Laat, W. 4C-seq from beginning to end: A detailed protocol for sample preparation and data analysis. *Methods* **2020**, *170*, 17–32. [CrossRef]
35. Geeven, G.; Teunissen, H.; De Laat, W.; De Wit, E. peakC: A flexible, non-parametric peak calling package for 4C and Capture-C data. *Nucleic Acids Res.* **2018**, *46*, e91. [CrossRef] [PubMed]
36. Martins, E.P.; Gonçalves, C.S.; Pojo, M.; Carvalho, R.; Ribeiro, A.S.; Miranda-Gonçalves, V.; Taipa, R.; Pardal, F.; Pinto, A.A.; Custódia, C.; et al. *Cadherin-3* is a novel oncogenic biomarker with prognostic value in glioblastoma. *Mol. Oncol.* **2022**, *16*, 2611–2631. [CrossRef] [PubMed]
37. Van Marck, V.; Stove, C.; Jacobs, K.; Eynden, G.V.D.; Bracke, M. P-cadherin in adhesion and invasion: Opposite roles in colon and bladder carcinoma. *Int. J. Cancer* **2011**, *128*, 1031–1044. [CrossRef] [PubMed]
38. Ribeiro, A.S.; Sousa, B.; Carreto, L.; Mendes, N.; Nobre, A.R.; Ricardo, S.; Albergaria, A.; Cameselle-Teijeiro, J.F.; Gerhard, R.; Söderberg, O.; et al. P-cadherin functional role is dependent on E-cadherin cellular context: A proof of concept using the breast cancer model. *J. Pathol.* **2013**, *229*, 705–718. [CrossRef] [PubMed]
39. Christgen, M.; Bartels, S.; van Luttikhuizen, J.L.; Bublitz, J.; Rieger, L.U.; Christgen, H.; Stark, H.; Sander, B.; Lehmann, U.; Steinemann, D.; et al. E-cadherin to P-cadherin switching in lobular breast cancer with tubular elements. *Mod. Pathol.* **2020**, *33*, 2483–2498. [CrossRef]
40. Paliou, C.; Guckelberger, P.; Schöpflin, R.; Heinrich, V.; Esposito, A.; Chiariello, A.M.; Bianco, S.; Annunziatella, C.; Helmuth, J.; Haas, S.; et al. Preformed chromatin topology assists transcriptional robustness of *Shh* during limb development. *Proc. Natl. Acad. Sci. USA* **2019**, *116*, 12390–12399. [CrossRef]

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.