U.PORTO

DOUTORAMENTO

BIOTECNOLOGIA MOLECULAR E CELULAR APLICADA ÀS CIÊNCIAS DA SAÚDE

The primary genetic cause and beyond: the lifetime risk and phenotype modifers in HDGC

Ana Rita Barbosa de Matos

D

2024

U.PORTO

Ana Rita Barbosa de Matos. The primary genetic cause and beyond: the lifetime risk and phenotype modifers in HDGC

The primary genetic cause and beyond: the lifetime risk and phenotype modifers in HDGC

Ana Rita Barbosa de Matos

D.ICBAS 2024

SEDE ADMINISTRATIVA

INSTITUTO DE CIÊNCIAS BIOMÉDICAS ABEL SALAZAR

FACULDADE DE FARMÁCIA

ANA RITA BARBOSA DE MATOS

THE PRIMARY GENETIC CAUSE AND BEYOND: LIFETIME RISK AND PHENOTYPE MODIFERS IN HDGC

Tese de Candidatura ao grau de Doutor em Biotecnologia Molecular e Celular Aplicada às Ciências da Saúde;
Programa Doutoral da Universidade do Porto
(Instituto de Ciências Biomédicas de Abel Salazar e Faculdade de Farmácia)

Orientador – Doutor (Carla Oliveira)
Categoria – (Investigadora Principal, Professora Afiliada)
Afiliação – (Faculdade de Medicina da Universidade do Porto)

Co-orientador – Doutor (Gustavo Mostosvlasky)
Categoria – (Professor)
Afiliação – (Boston University Chobanian & Avedisian School of Medicine,Boston.

| Funding

PhD Fellowship

- Fundação para a Ciência e a Tecnologia (FCT): SFRH/BD/145132/2019

Research Project

- "LIFE-TIME RISK ESTIMATIONS AND GENETIC MODIFIERS OF HEREDITARY DIFFUSE GASTRIC CANCER". [LEGOH], (Grant Ref. PTDC/BTM-TEC/6706/2020). 2021-2024. Funded by Fundação para a Ciência e a Tecnologia. BUDGET: € 249.951,25: PTDC/BTM-TEC/6706/2020

- "DISCLOSING THE GENESIS OF HEREDITARY STOMACH CANCER THROUGH ORGAN-ON-A-CHIP TECHNOLOGY". [DETONATE], [Project Ref: 2022.02951.PTDC], 2023-2024. Portuguese Science & Technology Foundation. (PI: Daniel Ferreira, ERiC group, Ipatimup/i3S, Porto, Portugal)

Other funding sources

- **Fulbright Fellowship with the support from FCT**
- **No Stomach for Cancer financial support for accommodation** (Duration: 11 months in Boston, USA)
- **European Joint Programme on Rare Disease Research Mobility Fellowship** (Duration: 1 month in Lyon, France)
- **SPGH Award Professor Amândio Tavares 2021** (travel expenses and registration in the European Society of Human Genetics Meeting)

| Team of Supervisors

- ## Principal Supervisor

  PhD, Carla Oliveira, Principal Investigator & Affiliated Professor

  i3S – Instituto de Investigação e Inovação em Saúde (i3S), Porto, Portugal

  Ipatimup – Institute of Molecular Pathology and Immunology of the University of Porto, Porto, Portugal

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

- ## Co-Supervisor

  MD PhD, Gustavo Mostosvlasky

  Center for Regenerative Medicine, Boston University, Boston, USA

  Boston University Chobanian & Avedisian School of Medicine, Boston, USA

- ## PhD Advisory Committee

  PhD, Susana Seixas

  i3S – Instituto de Investigação e Inovação em Saúde (i3S), Porto, Portugal

  PhD, Bruno Pereira

  i3S – Instituto de Investigação e Inovação em Saúde (i3S), Porto, Portugal

| Thesis Committee

Presidente

Doutor Manuel João Rua Vilanova, Professor Catedrático do Instituto de Ciências Biomédicas Abel Salazar da Universidade do Porto.

Vogais

Doutor Patrick Benusiglio, Professor Associado da Sorbonne Université, França;

Doutora Joana Gonçalves de Gouveia Maia Xavier, Investigadora Auxiliar da Universidade do Algarve;

Doutor Diogo Pinto da Cruz Sampaio e Castro, Investigador Principal do i3S- Instituto de Investigação e Inovação em Saúde da Universidade do Porto;

Doutora Carla Isabel Gonçalves de Oliveira, Investigadora Principal do i3S- Instituto de Investigação e Inovação em Saúde da Universidade do Porto

Doutor José Carlos Lemos Machado, Professor Catedrático da Faculdade de Medicina da Universidade do Porto;

Doutora Paula Maria Viera Jorge, Professora Auxiliar do Instituto de Ciências Biomédicas Abel Salazar da Universidade do Porto.

| Publications

This thesis follows the terms of the n. º 2, alínea a) do artigo 31º do Decreto-Lei n. º 115/2013 de 7 de Agosto, as the work developed and hereby presented is published in peer-reviewed scientific journals or in preparation for publication, as following detailed:

Paper 1 – Published manuscript

Garcia-Pelaez J*, **Barbosa-Matos R***, São José C*, Sousa S, Gullo I, Hoogerbrugge N, Carneiro F, Oliveira C. **Gastric cancer genetic predisposition and clinical presentations: Established heritable causes and potential candidate genes**. Eur J Med Genet. 2022 Jan;65(1):104401. doi: 10.1016/j.ejmg.2021.104401. Epub 2021 Dec 3. PMID: 34871783.

*Equal contributors

Paper 2 – Published manuscript

Garcia-Pelaez J*, **Barbosa-Matos R***, Gullo I, Carneiro F, Oliveira C. **Histological and mutational profile of diffuse gastric cancer: current knowledge and future challenges**. Mol Oncol. 2021 Nov;15(11):2841-2867. doi: 10.1002/1878-0261.12948. Epub 2021 May 2. PMID: 33724653; PMCID: PMC8564639.

*Equal contributors

Paper 3- Published manuscript

Barbosa-Matos R*, Leal Silva R*, Garrido L*, Aguiar AC, Garcia-Pelaez J, André A, Seixas S, Sousa SP, Ferro L, Vilarinho L, Gullo I, Devezas V, Oliveira R, Fernandes S, Costa SC, Magalhães A, Baptista M, Carneiro F, Pinheiro H, Castedo S, Oliveira C. **The *CDH1* c.1901C>T Variant: A Founder Variant in the Portuguese Population with Severe Impact in mRNA Splicing. Cancers** (Basel). 2021 Sep 4;13(17):4464. doi: 10.3390/cancers13174464. PMID: 34503274; PMCID: PMC8430675.

*Equal contributors

Paper 4 – Published manuscript

Garcia-Pelaez J, **Barbosa-Matos R***, Lobo S*, Dias A*, Garrido L, Castedo S, Sousa S, Pinheiro H, Sousa L, Monteiro R, Maqueda JJ, Fernandes S, Carneiro F, Pinto N, Lemos C, Pinto C, Teixeira MR, Aretz S, Bajalica-Lagercrantz S, Balmaña J, Blatnik A, Benusiglio PR, Blanluet M, Bours V, Brems H, Brunet J, Calistri D, Capellá G,

Carrera S, Colas C, Dahan K, de Putter R, Desseignés C, Domínguez-Garrido E, Egas C, Evans DG, Feret D, Fewings E, Fitzgerald RC, Coulet F, Garcia-Barcina M, Genuardi M, Golmard L, Hackmann K, Hanson H, Holinski-Feder E, Hüneburg R, Krajc M, Lagerstedt-Robinson K, Lázaro C, Ligtenberg MJL, Martínez-Bouzas C, Merino S, Michils G, Novaković S, Patiño-García A, Ranzani GN, Schröck E, Silva I, Silveira C, Soto JL, Spier I, Steinke-Lange V, Tedaldi G, Tejada MI, Woodward ER, Tischkowitz M, Hoogerbrugge N, Oliveira C. **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. 2023 Jan;24(1):91-106. doi: 10.1016/S1470-2045(22)00643-X. Epub 2022 Nov 24. Erratum in: Lancet Oncol. 2023 Jan;24(1):e10. PMID: 36436516; PMCID: PMC9810541.

*Equal contributors

Paper 5 – Manuscript in preparation

Barbosa-Matos R, Fonseca J*, Garrido L*, Seixas S, Oliveira R, Freixo J, Quental R, Louro P, Fernandes S, Grangeia A, Baptista M, Carneiro F, Pinheiro H, Castedo S, Gullo I, Drouet Y, Oliveira C. **Lifetime cancer risk associated with the most recurrent *CDH1* germline variant in DGLBCS: A Retrospective study of a founder effect.**

*Equal contributors

Paper 6 – Manuscript in preparation

Barbosa-Matos R, Fonseca J*, Ferreira M*, Garrido L, Pedro A, Dias D, Pelaez, J, Martins N, Brito A, Moutinho A, Seixas S, Oliveira R, Freixo J, Quental R, Louro P, Fernandes S, Grangeia A, Baptista M, Gullo I, Carneiro, F Pinheiro H, Castedo S, Ferreira D, Oliveira C. **Combined whole-genome and transcriptome profiling of early onset cancer patients and asymptomatic carriers of a *CDH1* founder variant.**

*Equal contributor

Complementary publications in related fields outside of this thesis scope:

Alimohamed MZ, Mnika K, Adadey SM, **Barbosa-Matos R**, Avram E, Nevondwe P, Akurugu WA, Mupfururirwa W, de Miranda Cerqueira JX, Dore R, Săbău ID, Kalantari S, da Silva ARGF, Anzaku AA, Matimba A, Chauke PA, Johari M, Nembaware V, Mroczek M. **Advancing intercontinental collaboration in human genetics: success story of the African and European Young Investigator Forum**. Eur J Hum Genet. 2024 Jan;32(1):3-5. doi: 10.1038/s41431-023-01487-6. Epub 2023 Oct 26. PMID: 37880422; PMCID: PMC10772082.

Elena Avram, Can Ding, Juliana Xavier de Miranda Cerqueira, Mridul Johari, Ana Raquel Gouveia Freitas da Silva, Ileana-Delia Săbău, Nuru Noor, Silvia Kalantari, Rhys Dore, **Rita Barbosa-Matos**, Magdalena Mroczek, Francesca Tonini & **European Society of Human Genetics - Young Committee. The European Society of Human Genetics—Young committee- activities and achievements between 2019–2022**. *Eur J Hum Genet* **31**, 383–386 (2023). https://doi.org/10.1038/s41431-023-01300-4

Tedaldi G, Molinari C, São José C, **Barbosa-Matos R**, André A, Danesi R, Arcangeli V, Ravegnani M, Saragoni L, Morgagni P, Rebuzzi F, Canale M, Pignatta S, Ferracci E, Martinelli G, Ranzani GN, Oliveira C, Calistri D, Ulivi P. **Genetic and Epigenetic Alterations of *CDH1* Regulatory Regions in Hereditary and Sporadic Gastric Cancer. Pharmaceuticals** (Basel). 2021 May 12;14(5):457. doi: 10.3390/ph14050457. PMID: 34066170; PMCID: PMC8151134.

| Acknowledgements

Serve este espaço para agradecer a todos os que fizeram parte desta etapa e que recordarei sempre com muito carinho.

Começo por agradecer à pessoa sem a qual esta tese nunca poderia existir, que teve a ideia deste projeto e me aguçou a curiosidade para o levar avante: Carla Oliveira, conhecida por transformar em ouro qualquer projeto de investigação em que toca. Das várias conversas que tivemos, onde de múltiplas maneiras aprendi, há uma que faço questão de recordar até hoje. Num momento após a minha defesa de Mestrado disseste-me que apesar de não saber exatamente o que queria, sabia o que não queria. Uso este exemplo porque sinto que representa duas constantes nestes anos do meu doutoramento que passamos juntas: 1) Teres visto em mim algumas qualidades que eu não sabia ter, até confrontada com os desafios por ti lançados e 2) Teres me sempre apoiado e incentivado para arriscar, explorar, riscar mais umas coisas da lista "do que, entretanto, fiquei a saber que não queria", mas sem ter medo de continuar à procura daquilo que quero. Vou sempre recordar com alegria a tua gargalhada contagiante, assim como a paixão pela ciência que me transmitiste como se de um "Imprinting" se tratasse. Espero que o nosso entusiasmo de discutir resultados nunca se gaste!

Foi também para mim muito importante nunca esquecer a humanização da investigação, razão pela qual quis desenvolver este projeto. A todas as famílias que conheci no Hospital de S. João, cujas células foram imortalizadas ainda que as suas vidas não, obrigada por acreditarem neste projeto e espero estar à altura do vosso altruísmo.

From the moment, Dr. Gustavo Mostoslavsky accepted to co-supervise me, I knew my stay in the US would be sweet (especially because you are the most extraordinary "chocoholic" that I met). Thank you for always making me feel as part of the team. I will never forget your guidance and your contagious passion for research: "*You need to have a special love for what you do because when you do experiments, you need to do them many times…once in a while, you get a great result that you can enjoy, and it makes it all worth it.*"

A huge thank you for the No Stomach For Cancer for their support. They allowed me to have the best learning experience so far in my career! It was a

pleasure to met Jon Florin, his family and Karen. I'm so grateful for your help and inspiring dedication to help others.

A very important thank you note to Youenn and Jeremie for hosting me in the Léon Bérard Centre. Specially to Youenn, it was really an honour to collaborate with you: "*It's hard to do good science, but what a pleasure it is when it works!*" Thank you for everything.

Aos membros da minha comissão de acompanhamento, em primeiro lugar agradeço terem aceitado este desafio! À Susana Seixas agradeço a simpatia, a disponibilidade, as perguntas aguçadas e vontade de saber sempre como estava a correr o trabalho e como podia ajudar! Foi sempre muito bom sentir o teu apoio, obrigada! Ao Bruno Pereira agradeço todo o tempo que roubei, por teres sempre uma postura irrepreensível. És para mim um modelo a seguir como investigador e fiquei muito feliz por podermos discutir ciência juntos!

Agradeço também a todos os membros do grupo Expression Regulation in Cancer por toda a disponibilidade e boa disposição (mesmo para comer frango frio com coca-cola quente). Uma nota especial para os colegas que me ajudaram a realizar grande parte do trabalho presente nesta tese. Foram uma série de grandes desafios que conquistamos em equipa, com alguns tubos de sangue "estourados" na centrifuga, erros de código e horas infinitas nas culturas à mistura. Agradeço muito o tempo e a paciência do João, da Marta, do Alexandre, do José, da Carminho, da Moutinho, da Ana Pedro, do Nelson, da Irene, da Lúzia e do Daniel.

Importa salientar o aluno de licenciatura que apareceu no laboratório porque um dia me ouviu nas Jornadas da Bioquímica e ficou curioso por saber mais. Obrigada por todos os momentos Nelsinho, venham as próximas conquistas juntos, tenho uma enorme admiração e carinho por ti! Também não podia esquecer *mis hermanos*: José, Joaquim e Sérgio (por costela direita ou esquerda), foram a verdadeira santíssima trindade que milagrosamente alegrou tantos dos meus dias!

Obrigada ao Miguel, Paulo e Karina que fazem do Lusco Fusco uma segunda casa. E aos amigos que saltaram do i3s para os finos, tremoços, bilhar e setas…. Obrigada pelas gargalhadas e descontração, em especial para a Taci e o Pedro.

To my good friends in South Boston: Yvonne, Kavya, Meng, Charith, and Adrian, I really miss you guys. Thanks for all the wonderful moments. All the car trips, lunches, special cooking dinners and starry nights are in my heart.

To my Coolidge Corner Family: Fraukazinha (& Anderson), Elizabeth and Jeff. Words are not enough to express how special you are. Thank you for taking care of me like family and making me feel at home. Specially you Frauke! Thanks for all your love, flowers, and our sacred Starbucks time.

Aos tugas no US: João, Luís, Sara, António e companhia obrigada pela diversão!

Aos meus queridos amigos de um "Futuro mais que ao léu" devo um agradecimento especial, fizeram-se sempre perto mesmo quando eu estive do outro lado do mundo. Ao Kiko, Miguel, Inês, Kika, Dani, Manel, Morais, Maria, França, Babuxa, Seixas, Sandocha (& Diogo!!). Crescemos juntos, e nestes quatro anos não foi diferente… que bem me fizeram as nossas férias, a vossa visita a Lyon, os miminhos que chegaram a Boston e todas as horas em que me ouviram reclamar da FCT e da precariedade científica. Obrigada, sou uma sortuda por vos ter na minha vida!

Agradeço também ao "quereis-queríamos", vivemos a loucura dos anos 20, o ínicio da depressão dos quase 30, mas que venham os próximos! À Tete, Joana, Ana Isabel, Patrícia, Rita, Sofs, Isabel, Xica, Mimas, Teresinha, Digs e Cuzinho o meu carinho por vocês é eterno. Por mais aventuras, desventuras e conquistas ao vosso lado! Obrigada por estarem sempre lá quando precisei.

Ao Alexandre que acompanhou toda esta jornada ao meu lado, agradeço todo o apoio e amor. Admiro a tua tranquilidade, inteligência, curiosidade e essência. Sou mais feliz por ter partilhado contigo tantos momentos bonitos. Graças a ti, tive também o prazer de conhecer a tua família incrível, que sinto também como minha. Obrigada, Aida, Pedrinho, Tité, Tia Nita, Tio João e Vó Laide por me receberem de malas e bagagens, sempre de braços e coração abertos.

Por último, ao meu porto de abrigo, os meus pais, tias, tio, primas e avós… obrigada por tudo o que fizeram por mim e que fizeram de mim o que sou hoje. O vosso amor incondicional foi essencial. A todos os que já partiram e ficaram as saudades, gosto de acreditar que estariam orgulhosos. Aos meus pais agradeço por terem sido sempre capazes de respeitar a minha insaciável vontade de ser livre. Agora percebo que nunca o somos totalmente, pois o meu coração fica sempre apertado quando vivo longe dos que me são mais importantes.

Foram quatro anos e pico em três folhas…. Sempre falei de mais….

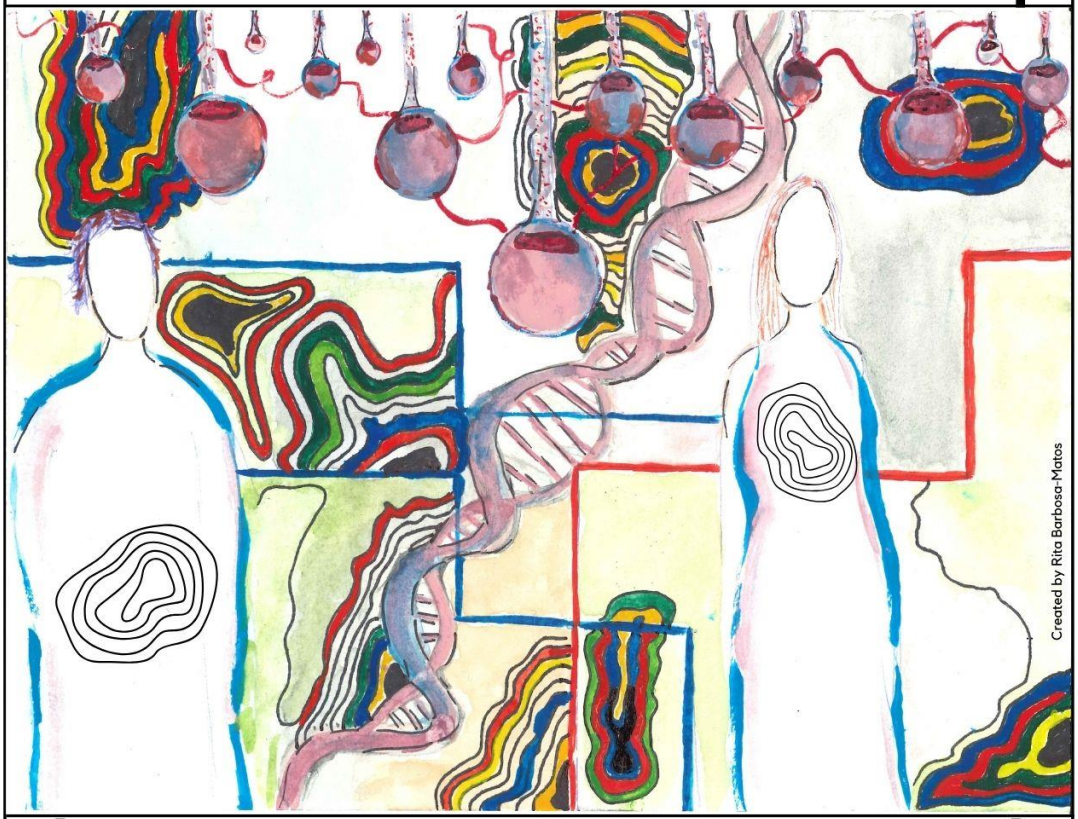
OBRIGADA A TODOS

X

| Motivational Context

The primary genetic cause and beyond:

Lifetime risk and phenotype modifiers in HDGC



"The most beautiful experience we can have is the mysterious. It is the fundamental emotion that stands at the cradle of true art and true science."
Albert Einstein

"Pela arte e pela ciência, nós vamos criar uma nova humanidade"
Graça Morais

| Table of Contents

|Thesis Outline XV
| Abstract XVI
| Sumário XIX
| Acronyms and Abbreviations XXII
Chapter I| Introduction 1
1| Gastric Cancer 2
1.1| Sporadic, Familial, and Hereditary Gastric Cancer 4
1.2| Histological, anatomic, and molecular classification of Gastric Cancer 7
1.3| Gastric Cancer Epidemiology 9
1.4. Diffuse Gastric Cancer Subtype: Pathogenicity, Clinical Manifestations, Incidence, and Beyond 13
2| Breast Cancer 15
2.1| Sporadic, Familial and Hereditary Breast Cancer 17
2.2| Breast Cancer Clinical, Histological and Molecular Classifications 19
2.3| Breast Cancer Epidemiology 21
2.4| Lobular Breast Cancer Subtype: Pathogenicity, Clinical Manifestations, Incidence, and Beyond 24
3| DGLBCS (Diffuse Gastric Lobular Breast Cancer Syndrome) previously known as Hereditary Diffuse Gastric Cancer (HDGC): Molecular Genetics 26
3.1| Causal genes associated with DGLBCS 26
3.2| Candidate genes 35
4| Clinical management of DGLBCS families 36
4.1| Surveillance and risk reducing approaches for asymptomatic carriers. 37
4.2| Treatment of clinical manifestations and supportive care 40
4.3| Challenges in management of families 42

5| Study models in DGLBCS....43
6| Challenging the Mendelian inheritance with genetic modifiers of disease....44
6.1| Modified penetrance by coding variants (SNVs and CNVs)....45
7| References....46
Chapter II| Rationale and Aims....67
Chapter III| Results....70
1. Search for DGLBCS candidate genes....71
1.1. Paper 1, Chapter III....72
1.2. Paper 2, Chapter III....91
2. *CDH1* germline alterations and phenotypes associated with DGLBCS....119
2.1. Paper 3, Chapter III....120
2.2. Paper 4, Chapter III....139
3. Lifetime risk in *CDH1* carriers and diagnostic signatures: insights from the Portuguese founder Population....156
3.1. Paper 5, Chapter III....157
3.2. Original Paper 6, Chapter III....192
4. DGLBCS experimental resources....219
4.1. iPSCs Biobank Establishment....219
Chapter IV|....231
Discussion & Conclusion....231

|Thesis Outline

This thesis can be divided into four chapters and summarizes the work developed in four years that originated two review articles, two published manuscripts and two manuscripts currently in preparation for publication.

Chapter I introduces the research topic: Diffuse Gastric Lobular Breast Cancer Syndrome (DGLBCS). Within its subchapters, it explains the current understanding of the clinical, molecular, and epidemiological aspects of the disease phenotypes associated with this hereditary disease. Moreover, it presents the ongoing challenges faced by those working on this syndrome.

Chapter II outlines the general and specific aims to address the DGLBCS challenges identified in the scope of this PhD thesis.

Chapter III presents the results obtained during this project, including a systematic search for DGLBCS candidate genes, and the clinical importance of variant classification, along with genotype–phenotype correlations. In more detail, taking advantage of a particular context in Portugal, we use a founder population to study lifetime risk and identify modifiers of disease penetrance and expressivity.

Finally, **Chapter IV** integrates all the generated data, providing an overview of the accomplishments, limitations, and future directions of the presented research work.

| Abstract

Diffuse Gastric and Lobular Breast Cancer Syndrome (DGLBCS, previously known as HDGC) is an autosomal dominant cancer susceptibility syndrome caused *CDH1* or CTNNA1 germline alterations. *CDH1*-DGLBCS predisposes individuals to early-onset aggressive Diffuse Gastric Cancer (DGC) and Lobular Breast Cancer (LBC). DGLBCS-like families share clinical presentation with DGLBCS, but lack molecular diagnosis, which prevents identifying at-risk individuals for disease prevention.

The current understanding of *CDH1*-DGLBCS families' disease lifetime risk is limited due to incomplete penetrance, differential severity and expressivity among carriers, and unknown mechanisms of disease progression. Disease surveillance in asymptomatic carriers at-risk is quite effective for LBC, by using Magnetic Resonance Imaging (MRI). However, gastro-endoscopy with multiple biopsies fails to detect DGC in >50% of carriers. Therefore, prophylactic total gastrectomy and mastectomy from *CDH1* mutation carriers remain the most efficient measures to prevent disease, however a non-negligible portion of carriers remain asymptomatic throughout their lives. It is urgent to better understand cancer lifetime risks to improve carriers' selection for risk-reducing surgeries. Current clinical guidelines reflect a one-fits-all approach for genetic testing and patient management.

Simultaneously, understanding which *CDH1* germline variants are causative, which clinical presentations are missing in *CDH1*-carrier families, and finding the molecular cause of DGLBCS-like families, remain unsolved issues in the field. In both *CDH1*-DGLBCS and DGLBCS-like families, it is essential to understand the disease modifiers that promote or prevent cancer development.

This thesis aims to provide more details on the genetic landscape of heritable GC, exploring both germline and somatic contexts, trying to solve currently molecularly undiagnosed families, and identifying candidate genes as disease modifiers. Furthermore, we seek to improve DGLBCS families' management, by establishing genotype-phenotype correlations and updating current clinical criteria, by estimating less biased lifetime cancer risks for *CDH1* carriers, and identifying distinct genomic expression profiles that may explain incomplete disease penetrance.

We addressed the genetic landscape of heritable GC in Papers 1 and 2 in Chapter III. Paper 1 focuses on frequently mutated genes in cases with a strong GC family history, while the second delves into the somatic landscape of DGC and its subtypes. In both, mutations in genes beyond *CDH1* were identified, establishing a list of candidates relevant for germline testing interpretation and identifying potential genetic hallmarks for disease severity at the somatic level.

In Paper 3 of Chapter III, we reclassified the *CDH1* c.1901 C>T variant from missense to truncating. This work illustrates the clinical impact of variant molecular classification in the management of DGLBCS families. Additionally, we described the first *CDH1* founder effect in Europe, based on several families from the Northern region of Portugal. Going further in variant classification, Paper 4 presented the first multicentre genotype-phenotype study for individuals carrying *CDH1* germline variants. Our findings underscore an association between truncating variants (Pathogenic/Likely-pathogenic) and typical DGLBCS cancers. This work also demonstrates that some families, that did not fulfil clinical DGLBCS criteria, were notably enriched in Lobular Breast Cancer (LBC) or Breast Cancer (BC) cases, creating an opportunity to expand current clinical criteria to accommodate at-risk families.

The identification of a Portuguese founder population from the same geographical location was pivotal in our research strategy. By having access to a cohort with low genetic and environmental heterogeneity, we were in a less biased context that allowed us to conduct a prospective study on DGLBCS families followed at a single institution, accessing both cancer risk and genetic modifiers.

Therefore, in Paper 5, we calculated a less biased cancer lifetime risk for DGC and LBC, using the largest cohort of families sharing the same causal variant. Importantly, we could overcome several limitations of previous studies including lack of histological subtype, country specific cancer incidence, and an exclusive *CDH1* variant type. This study also provided valuable insights into the outcomes of preventive measures, shedding light on their effectiveness, potential complications, and overall impact on the disease course.

In Paper 6, we collected biological material from affected and asymptomatic carriers. Ultimately by resorting to whole genome and total RNA sequencing of different sources (blood, peripheral blood mononuclear cells, and induced-pluripotent stem cells), we identified germline alterations and differentially expressed genes that can distinguish affected from asymptomatic carriers. The

results highlight a cancer predisposing genomic profile specific to individuals early affected by DGC (before 35 years old) and include several candidate gene modifiers that can be relevant biomarkers of disease onset.

The comprehensive work conducted in this thesis yielded pertinent genotype-phenotype correlations, lifetime risk data and disease expression profiles, laying the groundwork for future discussions and considerations in the clinical management of DGLBCS.

| Sumário

A Síndrome de Cancro Gástrico Difuso e Cancro da Mama Lobular (DGLBCS, anteriormente conhecido como HDGC) é uma síndrome de suscetibilidade ao cancro autossómica dominante, causada por alterações germinativas no gene *CDH1* ou *CTNNA1*. As famílias *CDH1*-DGLBCS têm risco aumentado para cancro gástrico difuso (DGC) agressivo de início precoce e cancro invasivo lobular da mama (LBC). Famílias semelhantes a DGLBCS apresentam uma apresentação clínica semelhante, mas carecem de um diagnóstico molecular, o que impede a identificação de indivíduos em risco para a prevenção da doença. A compreensão atual do risco de vida da doença em famílias *CDH1*-DGLBCS é limitada, devido à penetrância incompleta, e diferenças de expressividade de fénotipo entre portadores, com mecanismos desconhecidos de progressão da doença. A vigilância da doença em portadores assintomáticos em risco é eficaz para LBC, utilizando a ressonância magnética (MRI), mas a endoscopia digestiva alta com múltiplas biópsias falha em detetar DGC em >50% dos portadores. Portanto, a gastrectomia total e a mastectomia profilática em portadores de mutações no *CDH1* continuam a ser a medida mais eficaz para prevenir a doença, no entanto uma porção não-negligenciável dos portadores permanece assintomático ao longo da sua vida. É urgente compreender os riscos ao longo da vida para melhorar a estratificação de riscos para cirurgias redutoras de risco. As recomendações clínicas refletem uma abordagem generalista para todos os portadores quer nos critérios para teste genético como gestão de pacientes. Ao mesmo tempo, entender quais variantes germinativas no *CDH1* que são verdadeiramente causativas, quais apresentações clínicas estão ausentes em famílias de portadores de *CDH1* e encontrar a causa molecular de famílias semelhantes a DGLBCS são questões não resolvidas nesta área. Em ambas as famílias *CDH1*-DGLBCS e semelhantes a DGLBCS, é essencial compreender os modificadores da doença que promovem/previnem o cancro.

Esta tese visa fornecer mais detalhes sobre o panorama genético do cancro gástrico hereditário, explorando ambos os contextos germinais e somáticos, tentando resolver famílias atualmente não diagnosticadas molecularmente e identificar genes candidatos como modificadores. Além disso, tentaremos melhorar o acompanhamento clínico das famílias DGLBCS, estabelecendo correlações genótipo-fenótipo, atualizando e definindo novos critérios clínicos de doença, estimando o risco ao longo da vida para uma variante específica de *CDH1*

e identificando diferentes perfis de expressão genéticos que possam explicar a penetrância incompleta da doença.

Abordamos o panorama genético do cancro gástrico hereditário nos Papers 1 e 2 do Capítulo III. O primeiro foca-se em genes frequentemente mutados em casos com uma forte história familiar de cancro gástrico, enquanto o segundo explora o perfil somático de DGC e seus subtipos. Em ambas, foram identificadas mutações em genes além do *CDH1*, estabelecendo uma lista de candidatos a DGLBCS para a interpretação de resultados genéticos germinativos, bem como, identificar potenciais genes discriminadores de severidade da doença a nível somático. No Capítulo III: Paper 3, reclassificamos a variante *CDH1* c.1901 C>T de missense para truncante. Este trabalho ilustra o impacto clínico da classificação de variantes na gestão de famílias DGLBCS. Além disso, descrevemos o primeiro efeito fundador *CDH1* descrito na Europa, com base em várias famílias da região norte de Portugal. Avançamos ainda mais na classificação de variantes e, no Paper 4, realizamos o primeiro estudo multicêntrico genótipo-fenótipo para indivíduos portadores de variantes germinativas em *CDH1*. Nas nossas descobertas, destaca-se a associação entre variantes truncantes (patogênicas/provavelmente patogênicas) a cancros típicos de DGLBCS. Este trabalho também demonstra que algumas famílias que não preenchiam os critérios clínicos de DGLBCS eram notavelmente enriquecidas em casos de cancro da mama lobular (LBC) ou cancro da mama (BC) e portadores de variantes patogénicas no *CDH1,* estabelecendo uma oportunidade de expansão dos critérios atuais para acomodar estas famílias em risco.

A identificação de uma população fundadora portuguesa da mesma localização geográfica foi crucial em nossa estratégia. Tendo acesso a um grupo de estudo com baixa heterogeneidade genética e ambiental, tivemos um contexto mais imparcial que nos permite criar um estudo prospetivo em famílias com DGLBCS seguidos numa única instituição, avaliando o risco para cancro e a presença de modificadores genéticos. Portanto, no Paper 5, pudemos calcular o risco ao longo da vida para DGC e LBC, usando a maior coorte de famílias com a mesma variante causal. Importa salientar que conseguimos ultrapassar várias limitações presentes em estudos anteriores incluindo a ausência de subtipos histológicos, incidências para cancro específicas de cada região e considerar apenas uma variante do *CDH1* em específico. Este estudo forneceu ainda informações valiosas sobre o acompanhamento clínico destas famílias, os resultados das medidas preventivas, lançando luz sobre sua eficácia, complicações

potenciais e impacto geral no curso da doença. No Paper 6, recolhemos material biológico relevante de portadores afetados e portadores assintomáticos. Finalmente, recorrendo à sequenciação de genoma inteiro e RNA total de diferentes fontes (sangue, iPSCs), identificamos alterações germinativas e genes diferencialmente expressos que podem distinguir portadores afetados de assintomáticos. Os nossos resultados salientam a existência de um perfil genómico de predisposição para cancro, específico em indivíduos afetados por DGC em idades jovens (inferior a 35 anos), que incluem vários genes modificadores candidatos para biomarcadores relevantes na manifestação de doença.

O trabalho abrangente realizado até agora resultou em correlações genótipo-fenótipo pertinentes e dados de risco ao longo da vida, lançando as bases para futuras discussões e considerações gestão clínico de doentes com DGLBCS.

| Acronyms and Abbreviations

ACMG/AMP The American College of Medical Genetics and the Association for Molecular Pathology

ANOVA Analysis of Variance

BC Breast cancer

BMI Body Mass Index

BP Biological Process

BRCA1/2 Breast Cancer Associated gene 1/2

BRRM Bilateral Risk Reducing Mastectomy

CC Cellular Component

CDH1 Cadherin 1

CHUSJ Centro Hospitalar e Universitário de São João

CI Confidence Interval

CRC Colorectal Cancer

CTNNA1 Catenin alpha-1

DE Differential Expression

DGC Diffuse Gastric Cancer

ECIS European Cancer Information System

FCT Fundação para a Ciência e a Tecnologia

FDR False Discovery Rate

FFPE Formalin-Fixed Paraffin-Embedded

GC Gastric Cancer

GCO Global Cancer Observatory

GO Gene Ontology

GRL Genotype Restricted Likelihood

HDGC Hereditary Diffuse Gastric Cancer

ID Identification

IGC Intestinal Gastric Cancer

IGCLC International Gastric Cancer Linkage Consortium

IHC Imunohistochemical

IPATIMUP Instituto de Patologia e Imunologia Molecular da Universidade do Porto
Instituto de Patologia e Imunologia Molecular da Universidade do Porto

LBC Lobular Breast Cancer

LIN Lobular Intraepithelial Neoplasia

lncRNA Long non-coding RNA

MF Molecular Function

MRI Magnetic Resonance Imaging

NGS Next Generation Sequencing

NP Nonparametric

OC Obligate Carrier

PCA Principal Component Analysis

PSCA Prostate Stem Cell Antigen

RNA Ribonucleic Acid

RRG Risk Reducing Gastrectomy

RRM Risk Reducing Mastectomy

SEER Surveillance, Epidemiology and End Results Program

SIR Standard Incidence Ratio

SRC Signet Ring Cells

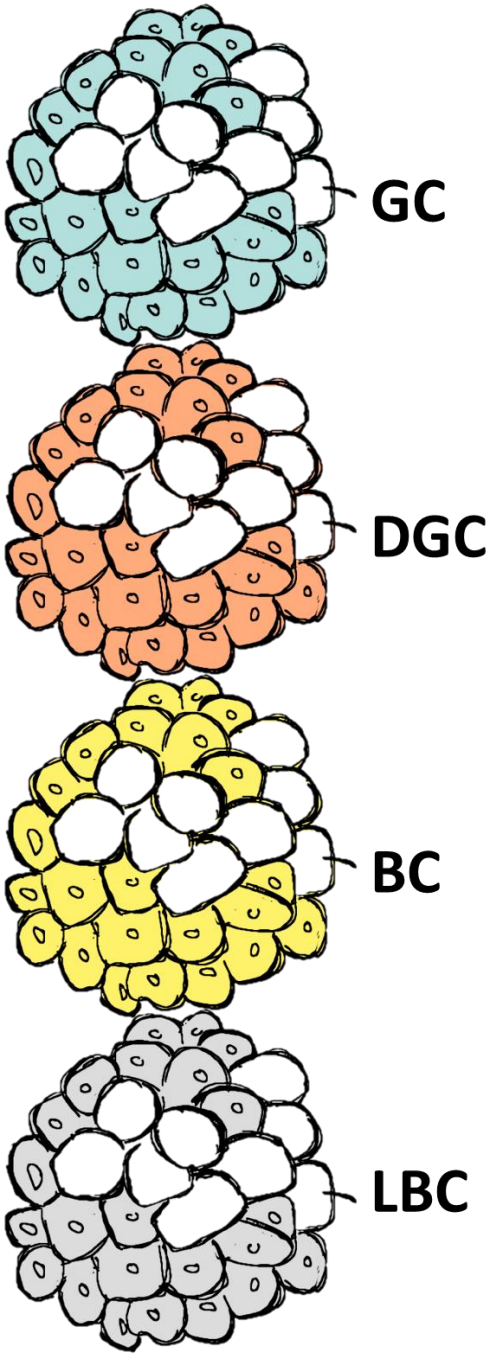
TAC Transcriptome Analysis console

TNM Tumour, Node, Metastasis Classification

UnkH Unknown Histological classification

DGLBCS Diffuse Gastric Cancer Lobular Breast Syndrome

Chapter I| Introduction



1| Gastric Cancer

Stomach cancer can affect any part of the stomach, but the most common types, representing around 90-95% of cases, begins in the glandular tissue of the stomach lining and are classified as gastric adenocarcinomas. Beyond this common form, other less prevalent types of stomach cancer exist, including gastrointestinal stromal tumours, neuroendocrine tumours, lymphomas, and exceedingly rare variations like small cell carcinomas, squamous cell carcinomas, or leiomyosarcomas. Gastric adenocarcinoma, commonly referred to as Gastric Cancer (GC), is a complex disease where both genetic and environmental factors impact its development and prognosis (Ajani et al., 2017; Gullo et al., 2020; Gullo, Carneiro, et al., 2018). Survival outcomes are closely tied to the stage at the time of clinical diagnosis.

Currently, GC staging and therapeutic approaches are determined by the TNM (Tumour, Lymph nodes, and Metastasis) system which considers the extent of tumour invasion/ tumour size (pT), the number of metastatic lymph nodes (pN), and the presence of distant metastasis (pM) (Amin et al., 2017). Premalignant lesions are considered stage zero, and this stage is also known as carcinoma in situ (pTis) when cancer cells are only in the top layer of the mucosa. In cases of early GC (pT1a or pT1b), where the cancerous lesion is confined to the mucosa and/or submucosa without lymph nodes involvement, patients typically experience a more favourable prognosis compared to those with advanced GC. Advanced GC (≥pT2) is characterized by tumour invasion into the muscularis propria or beyond, capable of metastasizing to distant organs (**Figure 1A**)(Gullo et al., 2020; Hu et al., 2012). Several factors, acting independently or collaboratively, can contribute to GC onset, emphasizing the need for better management strategies for high-risk individuals (**Figure 1B**) (López et al., 2023; Yang et al., 2023) Predominantly impacting individuals over 60 years old, with less than 2% of cases diagnosed before the age of 34, GC still poses unique challenges in understanding its etiology and manifestations (Pocurull et al., 2021).

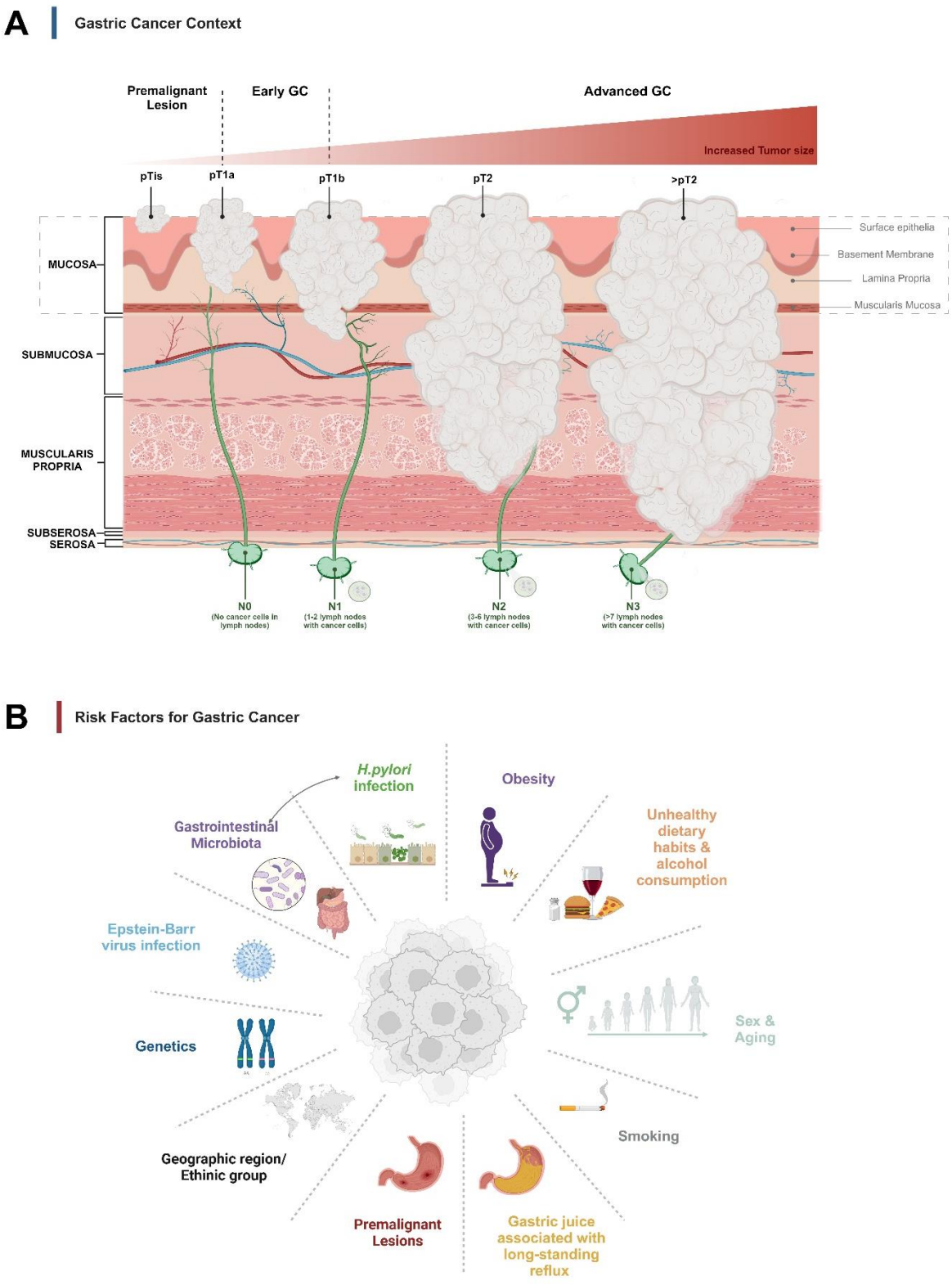


Figure 1| A. Stages of GC focusing on tumour size and lymph nodes involvement. Based on the AJCC cancer staging manual (8th ed.), American Joint Committee on Cancer, (2017), Springer. **B. Main factors associated with GC.** Adapted from Yang WJ *et al.*, World J Gastroenterol. 2023

1.1| Sporadic, Familial, and Hereditary Gastric Cancer

Sporadic GC constitutes 90% of all cases and remains the predominant manifestation associated with environmental factors and low-risk genetic elements (Corso et al., 2011; Song et al., 2018). Approximately 10% of GC cases exhibit familial aggregation, indicating an interplay between shared genetic susceptibilities, environmental influences, and lifestyle habits. Within this subset, 1–3% are classified as hereditary, implying that a pattern of inheritance was identified, and a germline causal variant may be identified in affected individuals. Typically, there is an autosomal dominant pattern of inheritance, and disease phenotypes are manifested at young ages. Recently, we revisited the current knowledge on heritable GC tumour risk syndromes reviewing their germline testing strategy, and patient management approaches and shedding light on the current diagnostic challenges, particularly reinforcing the utility of panel testing/exome sequencing for syndromes with overlapping cancer spectra, and the need for collection and publication of extended family pedigrees (**Original Review 1**).

Generically, there are three main heritable tumour risk syndromes associated with gastric cancer (**Figure 1**): (1) Diffuse Gastric and Lobular Breast Cancer Syndrome (DGLBCS) formerly known as Hereditary Diffuse Gastric Cancer; (2) Gastric Adenocarcinoma and Proximal Polyposis of the Stomach (GAPPS), and; (3) Familial Intestinal Gastric Cancer (FIGC) (Caldas et al., 1999; Corso et al., 2011; Gullo et al., 2020; Linhares et al., 2023; Oliveira et al., 2015b; Oliveira, Seruca, et al., 2009).

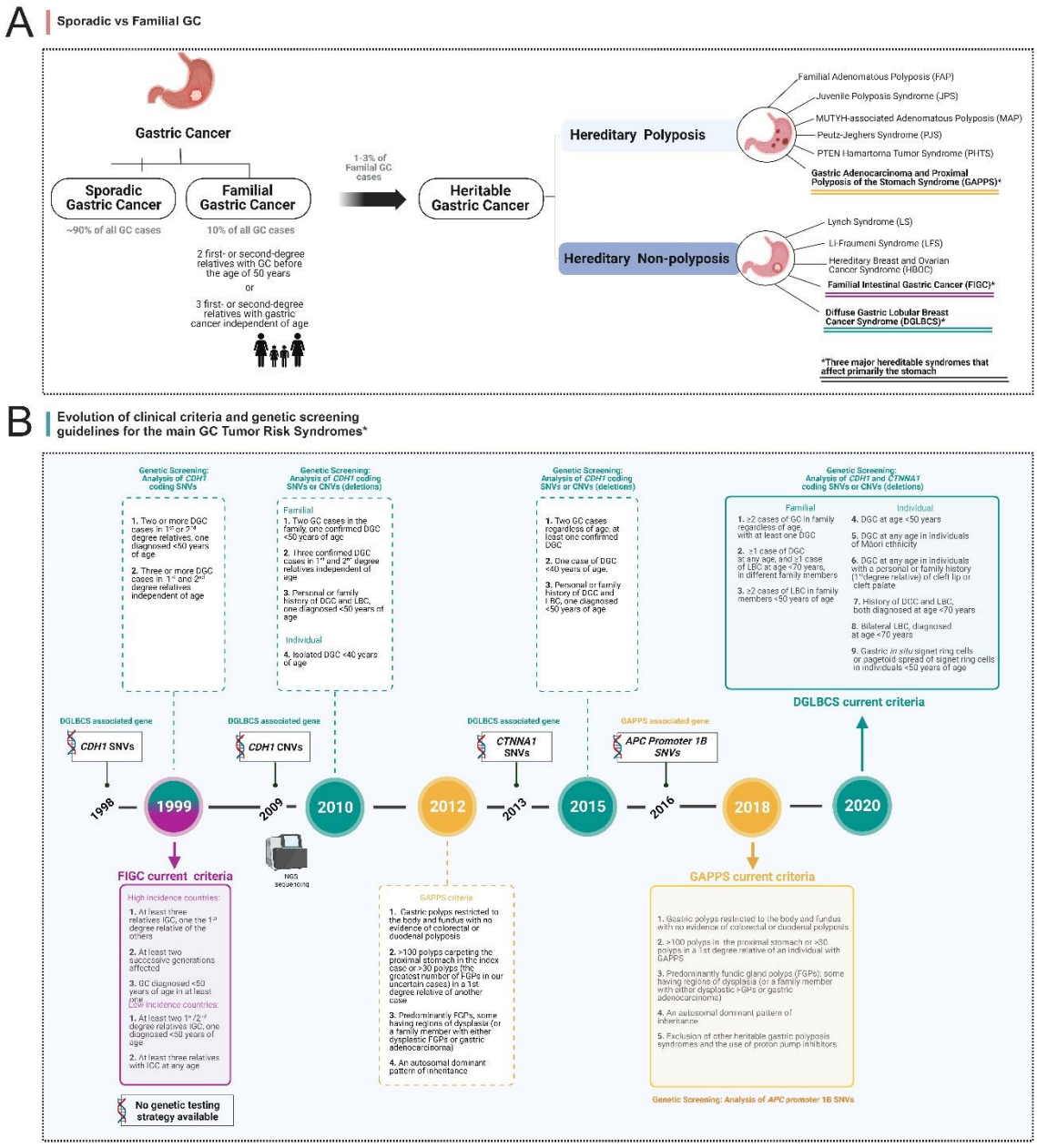


Figure 2| A. Overview of heritable tumour risk syndromes with GC as part of the disease spectrum. B. Current knowledge on clinical criteria and genetic testing screening guideline for Diffuse Gastric Lobular Breast Cancer Syndrome (DGLBCS), Gastric Adenocarcinoma and Proximal Polyposis of the Stomach (GAPPS) and Familial Intestinal Gastric Cancer (FIGC). Adapted with permission from: (Garcia-Pelaez et al., 2022)

GC and other gastric lesions can also constitute a significant aspect of the tumour spectrum associated with various inherited predispositions 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). A noticeable familial clustering of GC cases

has been observed for JPS, PJS and LS and less frequently in FAP, LFS, MAP, PHTS, or HBOC (**Table 1**) (Gullo et al., 2020; Kim et al., 2022; Lerner & Llor, 2020). The overlapping disease spectrum may hinder the correct diagnosis of a family TRS and therefore the identification of germline causative events is crucial.

Table 1| TRS presenting GC as part of their disease spectrum (without primary predisposition to GC).

	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	BRAC1 BRAC2	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 MSH2 MSH6 PMS2 EPCAM	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, stomach, 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).

Despite being rare in the general population, heritable forms of gastric cancer are associated with high morbidity and mortality rates. Therefore, the identification of at-risk individuals can be crucial to prevent disease onset and is achievable by improving genetic testing, surveillance and/or risk-reduction measures offered to patients and their families. Adapted with permission from (Garcia-Pelaez et al., 2022).

1.2| Histological, anatomic, and molecular classification of Gastric Cancer

GC displays variable morphological and molecular features between patients which is also reflected within each tumour variability (Gullo, Carneiro, et al., 2018). Despite several classification systems have been proposed, no unifying classification is used (Costache 2023).

One of the most popular systems is the Lauren classification which distinguishes two major types, intestinal gastric cancer (IGC) and DGC (Laurén, 1965). The former is constituted by tubular or papillary structures, while the latter is characterized by poorly cohesive and infiltrative tumour cells that may or may not have a signet-ring cell (SRC) morphology. These subtypes also correspond to distinct metastatic patterns: while IGC tends to metastasize haematogenously to the liver, DGC usually disseminates through peritoneal surfaces. The intestinal subtype is associated with intestinal metaplasia of the gastric mucosa and the occurrence of H. pylori infection (Carneiro, 1997). Moreover, its cancer process is well-established starting with pre-malignant lesions (atrophic gastritis, intestinal metaplasia, dysplasia) that develop into a malignant state in a well-known process defined as the Corea Cascade (Correa & Houghton, 2007). DGC is considered to have a worse prognosis, with a chain of events originating in the normal gastric mucosa and leading to advanced disease that remains to be fully determined with the clarified mechanism of disease initiation and progression (J. L. Monster et al., 2022). It is common in younger individuals with an equal gender distribution of affected individuals, while IGC is predominantly found in older males (Assumpção et al., 2020; Faria et al., 2013; Henson et al., 2004; Lauren & Nevalainen, 1993; Yao et al., 2020).

Tumours presenting both intestinal and diffuse components are termed mixed carcinomas. GCs that do not fit in one of these subtypes are placed in the indeterminate category. Mixed GC shows a poorer prognosis compared with IGC or DGC types (Zheng et al., 2008). Approximately half of the GC cases are classified as IGC, with 30% falling under the DGC type, and the remaining 15–20% exhibiting mixed or indeterminate characteristics. Although originating in 1965, the Laurén classification is still widely accepted and used, as it distinguishes subtypes with distinct epidemiologic settings, clinicopathologic profiles, and biological behaviours.

Another frequently employed classification system is the World Health Organization (WHO), 5th edition (2019), which recognizes five major types of GC (tubular, papillary, mucinous, poorly cohesive and mixed), as well as other rarer subtypes (gastric carcinoma with lymphoid stroma, hepatoid adenocarcinoma and related entities, micropapillary adenocarcinoma, gastric adenocarcinoma of fundic gland type, etc.)(Nagtegaal et al., 2020). Moreover, GC classification can be guided by the anatomical location of the tumour. Specifically, the cardia type is characterized by tumours located in in the proximal stomach near the gastro-oesophageal junction, and non-cardia GC, where tumours are situated in the distal stomach. (Hess et al., 2023).

The application of next-generation sequencing (NGS) technologies has significantly enhanced our understanding of the mutational burden and molecular signatures in gastric cancer (GC). This advancement has led to the formulation of GC molecular classifications, which rely on genomic, transcriptomic, and/or epigenomic profiles. There was an attempt to correlate the molecular subtypes with GC main histotypes, which did not provide a straightforward overlap.

The Cancer Genome Atlas (TCGA) consortium published in 2014 a deep and integrative molecular profiling with data from array-based somatic copy number analysis, microsatellite instability (MSI) testing, low-pass whole-genome sequencing, whole-exome sequencing, array-based DNA methylation profiling, messenger RNA (mRNA) sequencing, microRNA (miRNA) sequencing and reverse-phase protein array (RPPA) analysis (Bass et al., 2014). In their classification four GC categories were proposed: 1) The Epstein-Barr virus (EBV) positive tumours (~9%), 2) Microsatellite instability-high tumours (~22%), 3) Chromosomal instable tumours (CIN) (~50%) and 4) Genomically stable tumours (~20%). Despite providing novel information for patient stratification and distinct therapeutic targets within each patients group, no association with clinical outcome or prognosis was found within this cohort (Bass et al., 2014).

In 2015, the Asian Cancer Research Group (ACRG) generated comprehensive genomic and transcriptomic data sets from 300 GC cases, mainly from East Asia, producing the first GC classification based on integrated molecular data analysis and clinical outcomes (Cristescu et al., 2015). In this study, other four GC subtypes

were defined into MSI high and MSS types; the latter was further portioned into epithelial–mesenchymal transition (EMT), TP53+, and TP53− groups.

The research efforts that followed the publication of the TCGA and ACRG molecular classifications led not only to the development of cost-effective strategies to stratify GC patients according to these classifications but also to validate the molecular components from different molecular subgroups. This further allowed correlating molecular subgroups or alterations with clinical outcomes and response to therapy.

1.3| Gastric Cancer Epidemiology

According to the most recent GLOBOCAN estimates, GC remains the fifth most prevalent type of cancer globally, ranking as the fourth leading cause of cancer-related mortality. It is estimated that 1,089,000 cases were reached in 2020, causing 769,000 deaths (Sung et al., 2021a). The substantial lethality associated with GC can be attributed to its frequent late-stage diagnosis and suboptimal response to treatment. These challenges underscore the critical need for improved understanding of risk factors, enhanced early detection methods, and the identification of novel molecular targets for therapy to address this complex condition. Traditionally, the regions with the highest incidence of gastric cancer (GC) are Asia, the Pacific region of South America, and Central/Eastern Europe while North America and Africa present lower age-standardized incidence rates (ASIR) (**Figure 3**) (Linhares et al., 2023; Thrift et al., 2023)

The **incidence rate (IR)** reflects the number of new cases diagnosed within a fixed period, divided by the number of persons at risk during that same period of time.

$$IR = \frac{\#new\ cases\ during\ observation\ period}{person - time * \ at\ risk\ during\ observation\ period}$$

*Person-time is a concept used in epidemiology to measure the amount of time that individuals in a study population are at risk for a particular event or outcome. It is calculated by summing the time each individual is at risk while fulfilling the eligibility criteria.

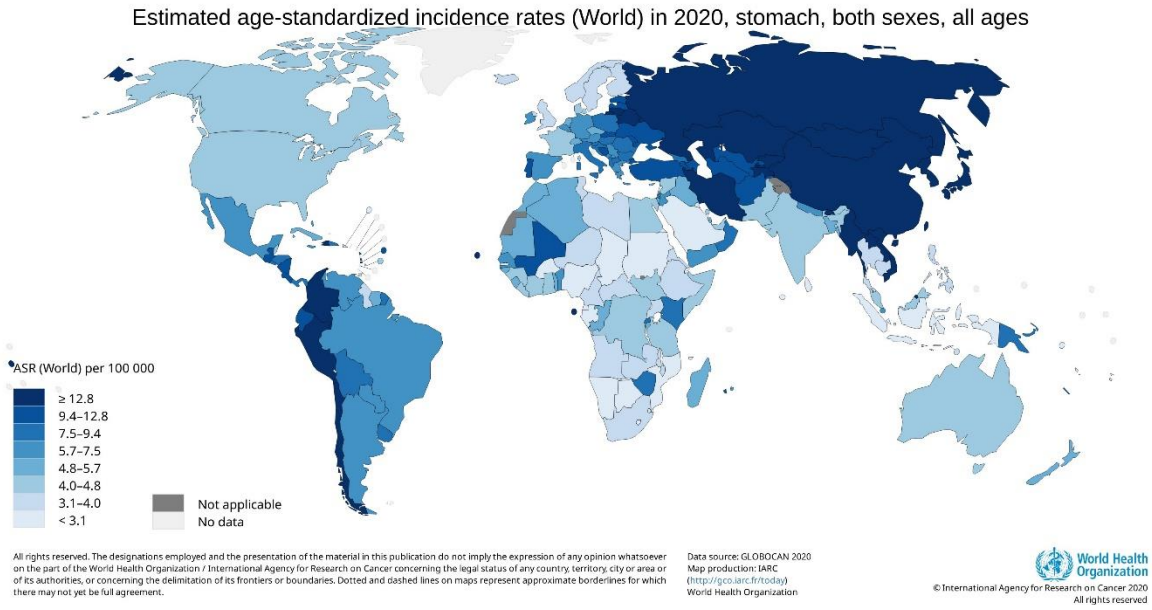


Figure 3| Worldwide GC incidence in 2020. The shading indicates estimated age-standardized incidence rates per 100,000 persons based on data from GLOBOCAN 2020 (Sung et al., 2021a), (accessed 6 January 2024; ©International Agency for Research on Cancer, 2020)

The ASIR of GC, encompassing both genders and all age groups, positions Portugal among the top 10 European populations with the highest estimated value. This corresponds to an ASIR of 11 affected per 100,000 persons. Notably, when focusing on the age group between 0-49 years, commonly referred to as Early-onset GC, Portugal ascends to the third position (**Figure 4A**). According to GLOBOCAN, GC emerges as one of the top five most prevalent cancers, excluding non-melanoma skin cancer, with a total of 2950 new cases diagnosed in 2020 across all ages and genders. Of these cases, 59% (1742 out of 2950) affect men (**Figure 4B**).

The **age-standardized incidence rate (ASIR)** is obtained by dividing the incidence rate of a specific population in a particular age group and multiplying the result by 100,000.

This allows us to compare distinct populations with different age profiles.

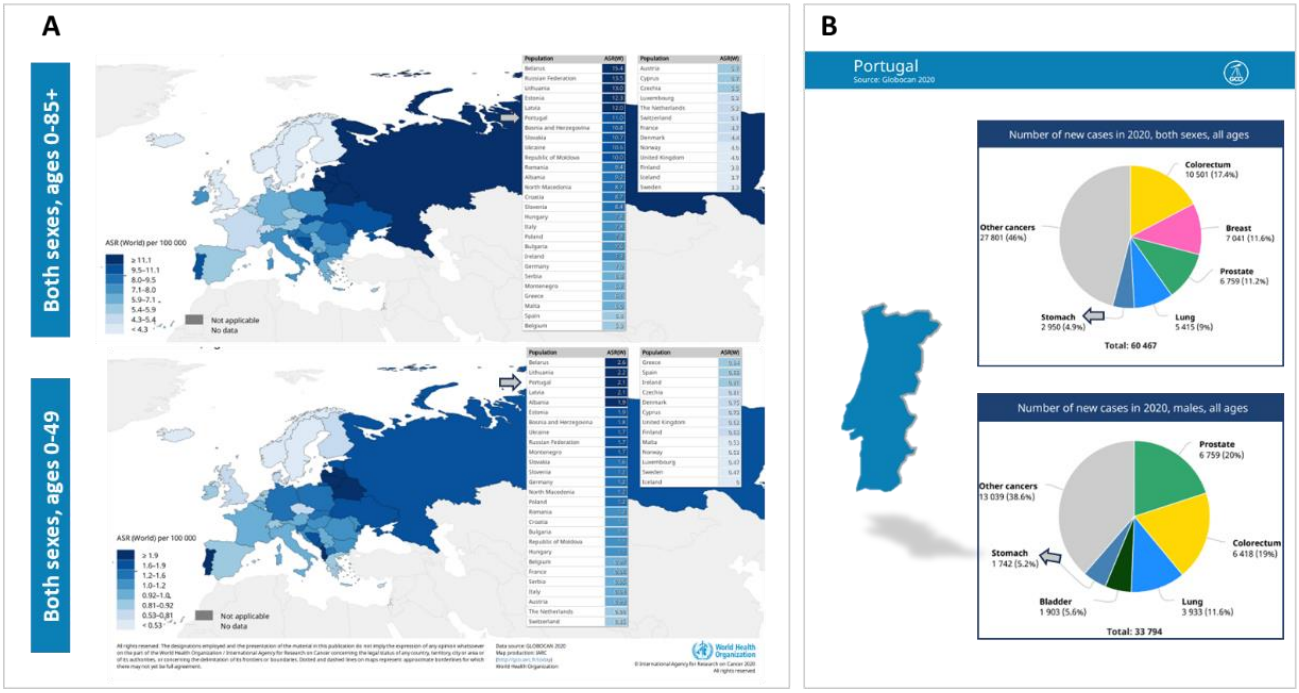


Figure 4| A. 2020 GC incidence in European populations. The shading indicates estimated age-standardized incidence rates per 100,000 persons based on data from GLOBOCAN 2020; **B. Distribution of newly diagnosed cancer cases in Portugal in 2020.** Both accessed on 6 January 2024; ©International Agency for Research on Cancer, 2020 (Sung et al., 2021a),

It is important to emphasize that the geographical variation of GC underscores the need for further evaluation of the usefulness and cost-effectiveness of applying different disease management strategies between low and high-risk populations, especially to improve disease management in regions with high incidence rates and risk factors (Linhares et al., 2023). For example, the Northern region of Portugal has a higher incidence of GC (**Figure 5**), but the healthcare systems are similar throughout all districts, leading to equivalent prevalence and mortality rates.

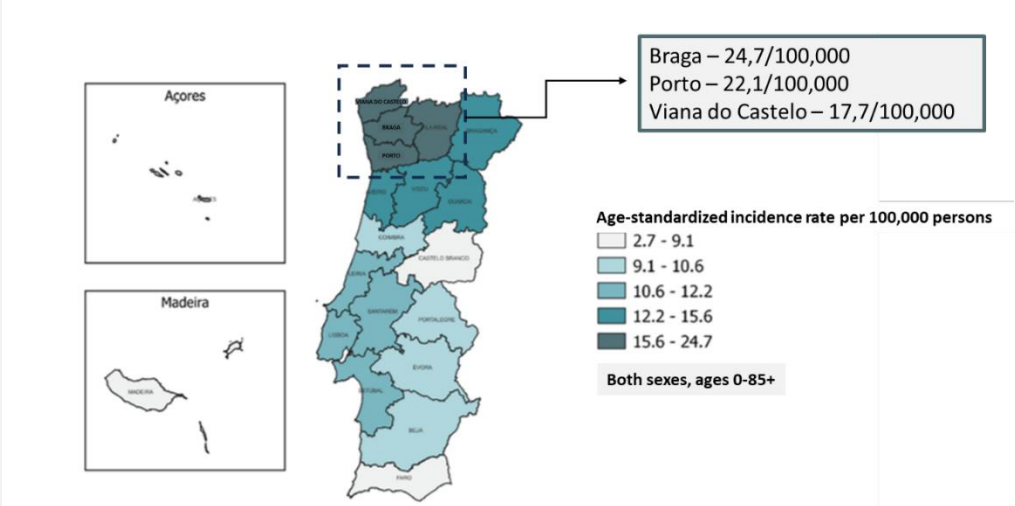


Figure 5| Geographical distribution of GC incidence across Portugal in 2020. Adapted from: RON. Registo Oncológico Nacional de Todos os Tumoures na População Residente em Portugal, em 2020. Instituto Português de Oncologia do Porto Francisco Gentil – EPE, ed. Porto, 2023

Interestingly, although the incidence of GC is declining over time, largely due to better quality of life (better hygiene, improved food preservation, changes in diet, widespread antibiotic use, and other socio-economic metrics), it has been described that this tendency is not applicable for all countries(Arnold, et al., 2020). This may reflect a limitation present in most epidemiological studies which have been hampered by the lack of a uniform practise in the nomenclature of diagnosis and registration of GCs. Most cases are reported with unspecified topography or GCs with distinct histological entities are considered together. Frequently in these studies, subsite-specific analysis is not possible and the tendency for overall GC may not reflect the tendencies for its particular subtypes (Castro et al., 2017; Libânio et al., 2022; Arnold, et al., 2020). Moreover, there is an observed trend in recent decades indicating a shift from IGC to DGC in non-cardia (López et al., 2023; Lv et al., 2021; van Velzen et al., 2023). Therefore, to consider diffuse and intestinal-type GC as one single entity raises data incidence inconsistencies.

Since this thesis focus is to explore Diffuse Gastric Lobular Breast Cancer Syndrome, in the next section, we will take a closer look at the details of DGC.

1.4. Diffuse Gastric Cancer Subtype: Pathogenicity, Clinical Manifestations, Incidence, and Beyond

The classical picture of DGC displays a tumour composed of poorly cohesive signet-ring cells (SRCs). The presence of signet cells arises from the convergence and compression of multiple mucin vacuoles, which ultimately push the nucleus against the plasma membrane. Various SRC types exist, characterized by cytoplasmic mucin nature, yet no specific type correlates with prognosis(Gullo et al., 2021; Sugihara et al., 1987). *In-situ* SRCC (Tis) present depolarised and disorganised SRCs replacing the normal epithelial cells. In this context, the presence of a linear proliferation of SRCs between the gastric foveola and/or glands and the basal membrane is defined as a pagetoid spread of SRCs. The invasion of the lamina propria giving rise to pT1a foci can occur without the morphological detection of these preinvasive lesions since their detection rate is much lower (Gullo et al., 2021). Both the T1a stage and in situ foci show significant overexpression of cytokeratin 7 (Decourtye-espiard & Guilford, 2023).

Advanced lesions may present cells that do not present classic SRC morphology but instead acquire a pleomorphic and diffusely infiltrative neoplastic transformation. sometimes with a lymphohistiocytic or sarcomatoid appearance. This subtype is associated with a worse prognosis when compared to classic SRC carcinomas, and frequently presents abnormal or absent E-cadherin immunohistochemistry. The recognition of different SRCs types and DGC subtypes variability between different tumours but also within different areas of the same tumour reflects the heterogeneity of this disease. It remains as an unsolved question whether advanced DGC stems from the malignant transformation of an early lesion or if it represents a distinct event. (Gullo et al., 2021; Li et al., 2020).

Symptoms of DGC are nonspecific in the early stages and consequently tend to be dismissed both by affected individuals and physicians. By the time specific symptoms appear, affected individuals present with advanced disease (Iyer et al., 2020). Symptoms at late-stage DGC may include abdominal pain, nausea, vomiting, dysphagia, postprandial fullness, loss of appetite, and weight loss. A palpable mass may be present late during DGC. Advanced DGC often involves metastatic spread into the gastrointestinal tract, peritoneum, omentum, mesocolon, and female reproductive tract, spreading to other organs is rare but

can happen to bone, lung and liver. Conventional therapies are usually ineffective, there are no targeted drugs for DGC, and it is associated with less chemosensitivity. Novel therapeutic agents are needed, with encouraging results from two phase II trials for anti-CLDN18.2 (zolbetuximab) and FGFR2-IIIb (bemarituzumab) antibodies holding promise as effective strategies for individuals with overexpression of these proteins in cancer cells (Decourtye-espiard & Guilford, 2023; J. Monster et al., 2022; Ooki, 2022).

DGC is linked more strongly to genetic factors than environmental aspects, as indicated by an earlier onset age. Interestingly, Despite DGC ethology does not seem to follow the Correa hypothesis, the role of H. pylori infection, EBV infection, tobacco along other environmental and lifestyle factors as potential contributors to disease development remains a subject of ongoing debate (Assumpção et al., 2020; Iyer et al., 2020; J. L. Monster et al., 2022).

Manifesting a marked variation in incidence among diverse ethnic groups, DGC exhibits the highest incidence in African populations reaching 60.4%, followed by Asia at 46.4%, and the United States at 40.5%. Conversely, Latin America reports lower incidences at 36.5%, while in Caucasians DGC cases account for nearly 30% of all gastric carcinoma cases (López et al., 2023).

Multiple studies involving diverse populations have identified genetic alterations associated with DGC tumours, with *CDH1* and *RHOA* gene mutations being frequently reported and highlighted for their relevance. These mutations contribute to alterations in cell adhesion and the actomyosin network, underscoring their pivotal roles in the disease progression (Kakiuchi et al., 2014; J. L. Monster et al., 2022; Wang et al., 2014).

Having explored DGC, we will now transition to a comprehensive overview of Breast Cancer (BC), laying the foundation for a more detailed examination of Lobular Breast Cancer.

2| Breast Cancer

Worldwide, female Breast Cancer (BC) is the most common type of cancer leading to 2.3 million new cases and causing the death of 685,000 women in 2020. BC in males is rare, being less than 1% of all BCs (Fox et al., 2022). Currently, most of the deaths (80%) happened in women aged 50 and older, and 70% of them were diagnosed within the same age range, depicting the main impact in this specific age group (Arnold et al., 2022; Łukasiewicz et al., 2021). Importantly, the majority of the deaths caused by BC are the result of long distant organ metastasis rather than the primary breast tumour itself. The preferential target sites are the bone and lungs, liver, and brain (Ivanova et al., 2023; Scully et al., 2012).

Early-stage BC is defined by the tumour being localized within the breast and not have spread to nearby lymph nodes (**Figure 6**A). The detection of such cases has been increasing due to heightened surveillance awareness and improved diagnostic methodologies (Cui et al., 2020; Jarvandi et al., 2016). Moreover, BC can be divided into non-invasive and invasive forms representing distinct disease stages. Non-invasive BC confines cancer cells to ducts and lobules, while invasive cancer infiltrates surrounding adipose and connective tissues. Both lesions show distinct immune infiltrate compositions (Tarighati et al., 2023).

BC is associated with many risk factors, including environmental and genetic predisposition (**Figure 6**B)(Obeagu & Obeagu, 2023). This contributes to its significant heterogeneity, featuring multiple disease subtypes. Disease treatment depends on the tumour molecular features since no therapeutic approach is effective for all disease subtypes. Survival rates will largely depend on the cancer stage and molecular subtype. BC treatment requires a multidisciplinary team merging the expertise from different specialties involving surgery, radiotherapy, chemotherapy, hormone therapy and targeted therapies combined differently according to each patient treatment plan (Łukasiewicz et al., 2021).

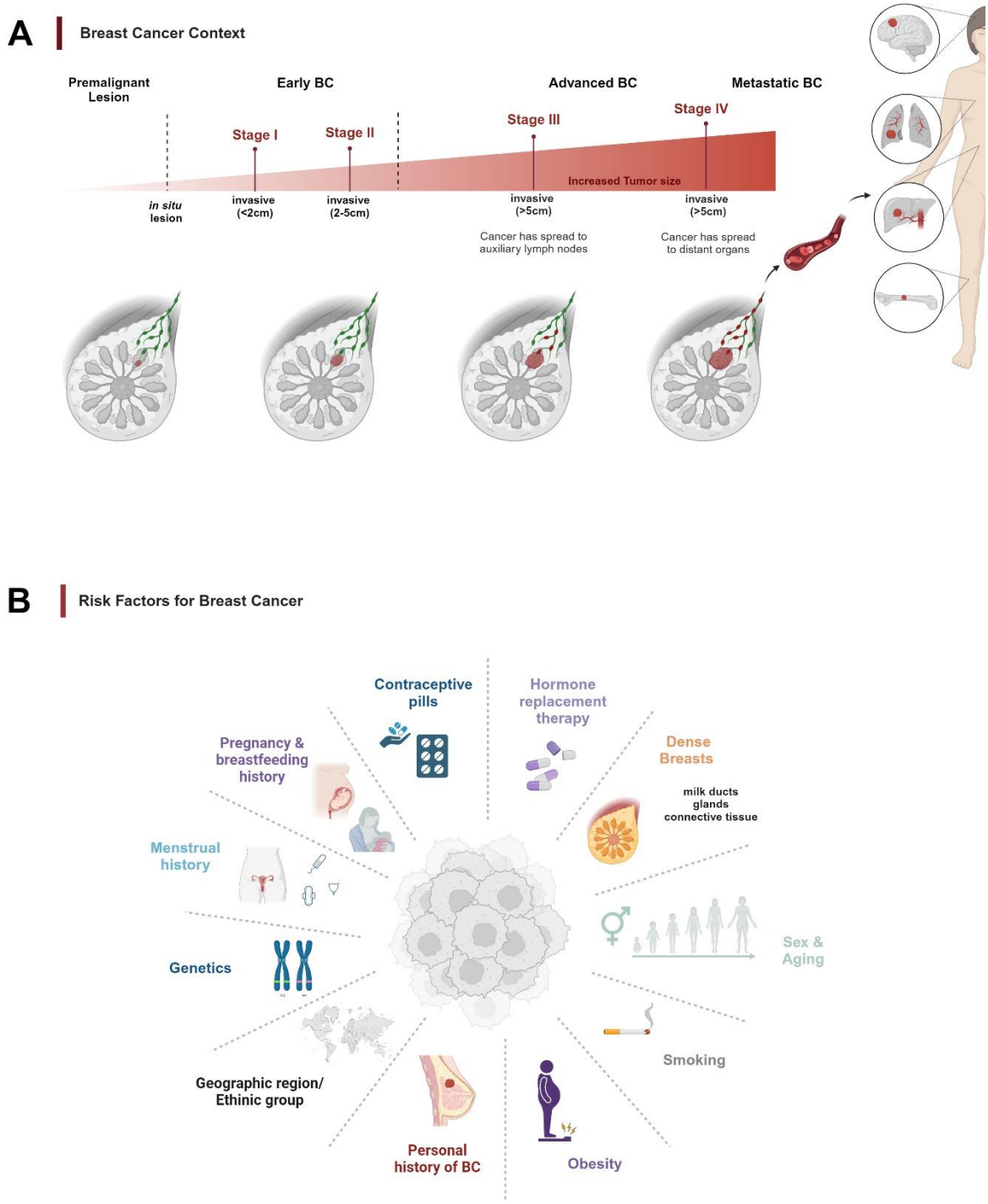


Figure 6| A. Stages of BC focusing on tumour size and lymph nodes involvement. Based on the AJCC cancer staging manual (8th ed.), American Joint Committee on Cancer, (2017), Springer. **B. Main factors associated with BC.** Adapted from Obeagu *et al.*, 2023.

2.1| Sporadic, Familial and Hereditary Breast Cancer

Understanding the differences among sporadic, familial, and hereditary breast cancer forms allows an accurate diagnosis and effective treatment. Being most BCs considered sporadic, 5%–20% of cases depict familial clustering of disease depending on specific genetic and geographical contexts (Farahat et al., 2017; Lakhani, 1999). Only 5–10% of BC cases are hereditary but in the last few decades, many BC susceptibility genes have been identified, especially due to the development of novel sequencing technologies (Fairoosa & Witharana, 2020).

Due to the disease high prevalence and impact on public health, significant efforts have been made to underpin the genetic susceptibility to BC. Genetic variants linked to BC risk can be high-penetrance (rare in the population but associated with substantial risk), moderate penetrance variants (modest increases in risk) and low-penetrance polymorphisms (common variants in the population but associated with slight risk increments). A minor portion of the hereditary cases exhibit germline alterations in genes classified as high penetrance, with the most well-known being *BRCA1* and *BRCA2* (Apostolou & Fostira, 2013; Lalloo & Evans, 2012) Moreover, several hereditary syndromes are associated with an increased risk of BC development, including DGCLBC associated with *CDH1* germline alterations (**Table 3**). They are associated with BC but also with the development of other types of cancer, shared in an autosomal dominant pattern. This shows that diverse combinations of multiple variants are responsible for hereditary BC susceptibility constituting a polygenic risk, instead of being accounted for a single variant in a specific gene. Approximately, 18% of hereditary BC cases are explained by polygenic risk while 25–30% are due to moderate and high penetrance genes (Sokolova et al., 2023). Interestingly, most genes are involved in DNA damage and repair (Apostolou & Fostira, 2013; Fairoosa & Witharana, 2020). These findings underscore the complexity of BC associated risk and highlight the need for personalized management strategies appropriate for the type of germline alterations identified in carriers and their families. For example, risk-reducing surgery is favoured in females carrying high penetrant genes while surveillance by Magnetic Resonance Imaging (MRI) is offered for carriers of low-penetrant variants (Daly et al., 2021; Hong & Xu, 2022).

Table 3. Major hereditary syndromes predisposing to BC

	Syndrome	Inheritance	Cancer Spectrum	Major Associated Genes	References (PMID)
Hereditary Syndromes predisposing to breast cancer	Hereditary breast and ovarian cancer syndrome (HBOC)	AD	Breast, ovarian, prostate, pancreatic and gastric cancer.	*BRAC1* *BRAC2*	32259785 16369438 12373604
	Li-Fraumeni syndrome (LFS)	AD	Breast cancer, CNS cancer, soft tissue sarcoma, osteosarcoma, adrenal gland carcinoma, leukemia, gastric, lung, and adrenocortical cancers, lymphomas.	*TP53*	32457520 32259785 3409256 26014290 21552135 36468211
	DGLBCS	AD	Breast and diffuse gastric cancer	*CDH1*	36436516
	Neurofibromatosis Type 1 (NF1)	AD	Optic pathway glioma, breast cancer, High-grade glioma, malignant peripheral nerve sheath tumours (MPNSTs)	*NF1*	26926675 30006586 28230061
	Peutz-Jeghers syndrome (PJS)	AD	Colorectal, stomach, breast, small intestine, pancreatic, ovarian, uterus, cervical, and lung cancer	*STK11*	24143323 20051941 20581245 *NCCN 2021
	PTEN hamartoma tumor syndrome (PHTS) / Cowden syndrome (CS)	AD	Breast, endometrial and follicular thyroid cancer; gastrointestinal hamartomas or ganglioneuromas, colon cancer, lipomas, renal cell carcinoma and	*PTEN*	32533092 32259785 23335809 28741261 36468211
	Louis-Bar syndrome (LBS)	AR	Lymphomas and lymphoid leukemias, breast cancer	*ATM*	26662178 9442910 27112364
	Nijmegen Breakage Syndrome - like disorder	AR	Large intestine, skin, breast, stomach, breast, endometrium, urinary tract, upper aer tract, thyroid, ovary cancer	*RAD51D* *RAD52*	30099541 36107942
	Nijmegen Breakage Syndrome (NBS)	AR	Lymphoma, squamous cell carcinoma, breast cancer	*NBN*	17957789 18788073 17957789

Footnote: AD, Autosomal Dominant; AR, Autosomal Recessive; CNS, Central Nervous System; NCCN

2.2| Breast Cancer Clinical, Histological and Molecular Classifications

BC tumours present diverse morphological, pathological, and biological characteristics. To facilitate patient diagnosis and guide treatment decisions, cancer classification systems categorize tumours into subtypes. In addition to the traditional clinicopathological features included in the TNM staging, the staging of BC involves evaluating the expression of specific hormone receptors through immunohistochemistry (IHC). Clinically, BCs are commonly classified into four immunohistochemical phenotypes based on the expression of these hormone receptors: 1) Estrogen receptor-positive, 2) Progesterone receptor-positive, 3) Human Epidermal Growth Factor Receptor-2 (HER-2) positive, and 4) triple-negative, characterized by the absence of expression of any of the mentioned receptors which are also considered prognostic markers (Orrantia-Borunda E. et al., 2022).

Moreover, thanks to microarrays and gene expression profiling development, an intrinsic subtyping system was developed and originated a molecular classification system. In total, four intrinsic subtypes are usually referred to, namely, Luminal A, Luminal B, HER2-overexpressing, basal-like breast cancers, and normal-like tumours. Several studies have explored these subtypes assigning different gene sets to each signature. Nevertheless, due to technical simplicity and lower costs, in daily clinical the markers for IHC-based molecular classification rely on evaluating the expression of the above-mentioned hormone receptors together with Ki-67 for tumour proliferation intensity (Sarhangi et al., 2022).

Regarding the histological classification system defined by the 2019 World Health Organization 2019, at least 20 different histological types exist based on the pathologic growth pattern of the tumour. Nevertheless, tumours are distributed into two major groups. One is represented by non-invasive carcinomas which have a better prognosis and can be considered premalignant including atypical lobular hyperplasia, lobular or ductal carcinoma *in-situ*. Nevertheless, the presence of precursor lesions is always associated with an increased risk for invasive cancer. The second group is invasive carcinoma, which means these tumours are capable

of invading surrounding and distant tissues. Invasive BC is classified in two groups subtypes: the non-special type, also known as Invasive Ductal Carcinoma, and the special subtypes including Invasive Lobular Carcinoma, Tubular Carcinoma, Mucinous Carcinoma, and several others (Eliyatkin et al., 2015; Makki, 2015; McCart Reed et al., 2021; Nascimento & Otoni, 2020; Tsang & Tse, 2020; Weigelt et al., 2010)(**Figure 7**).

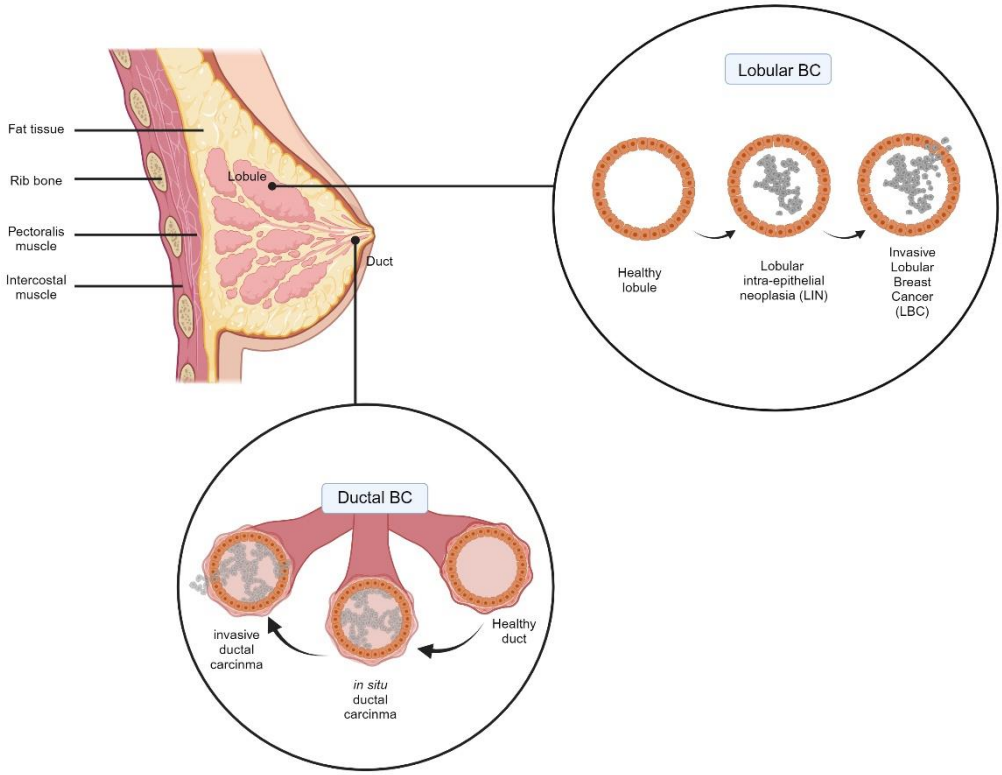


Figure 7| A. Invasive and non-invasive ductal and lobular breast cancer

Invasive carcinoma of no special type represents 60-85% of all BCs, being the most common histological subtype in hereditary breast cancer and frequently observed in the context of Li-Fraumeni syndrome and neurofibromatosis type 1. The second most frequent is Lobular Breast Cancer, typical of DGLBCS, accounting for 10% of BCs cases being frequently associated with *CDH1* germline alterations but not with *BRCA1* or *TP53* (Benusiglio et al., 2013; Ditchi et al., 2019). The remaining 18 subtypes are considered rare (0.5%-5%), however, they exhibit less variability in their transcriptomic patterns compared to invasive carcinoma of no special type (Hong & Xu, 2022; Sokolova et al., 2023; Weigelt et al., 2010).

2.3| Breast Cancer Epidemiology

Globally, BC ranks as the second most diagnosed cancer. In 2020, 2.3 million women were diagnosed, accounting for 11.7% of all diagnosed cancers. Early detection and improved therapy are decreasing breast cancer mortality. Nevertheless, BC still is the fifth most deadly cancer site (Sung et al., 2021b; WHO & Organization, n.d.).

There are significant variations in BC incidence among countries and world regions. According to GLOBOCAN 2020 the top 5 regions in the world with the highest ASIR (all above 70 cases per 100,000 females), were Australia/New Zealand, Northern America, and Europe (Northern, Southern, and Western). On the other hand, the lowest ASIR were observed in Asia (South-Central and South-Eastern), in Africa (South-Eastern, Eastern, and Middle Africa), and Central America (Globocan, 2020; Łukasiewicz et al., 2021)(**Figure 8**).

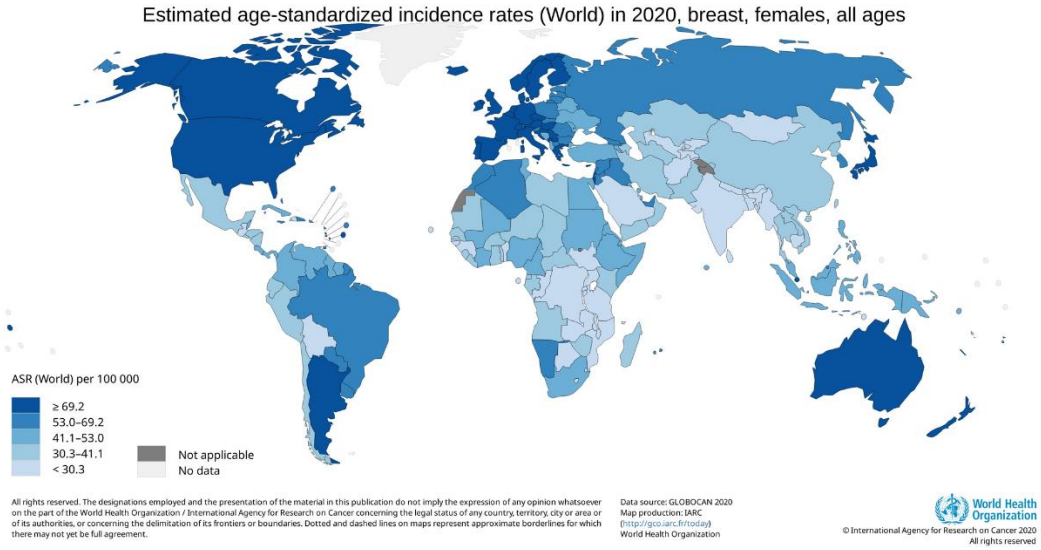


Figure 8| Worldwide BC incidence in 2020. The shading indicates estimated age-standardized incidence rates per 100,000 persons based on data from GLOBOCAN 2020 (Sung et al., 2021a), (accessed 6 January 2024; ©International Agency for Research on Cancer, 2020)

Taking a closer look at Europe, the country with the highest ASIR was Belgium followed by the Netherlands and Luxembourg. Portugal ranks 23rd when considering all ages with an ASIR of 70.8 per 100,000. Interestingly, when considering the age interval of 0–49 years, Southern European countries take over. Portugal rises to the 9th highest ASIR, still with the top 3 countries being Belgium, France and the Netherlands (**Figure 9**) all with ASIR below 45/100,000.

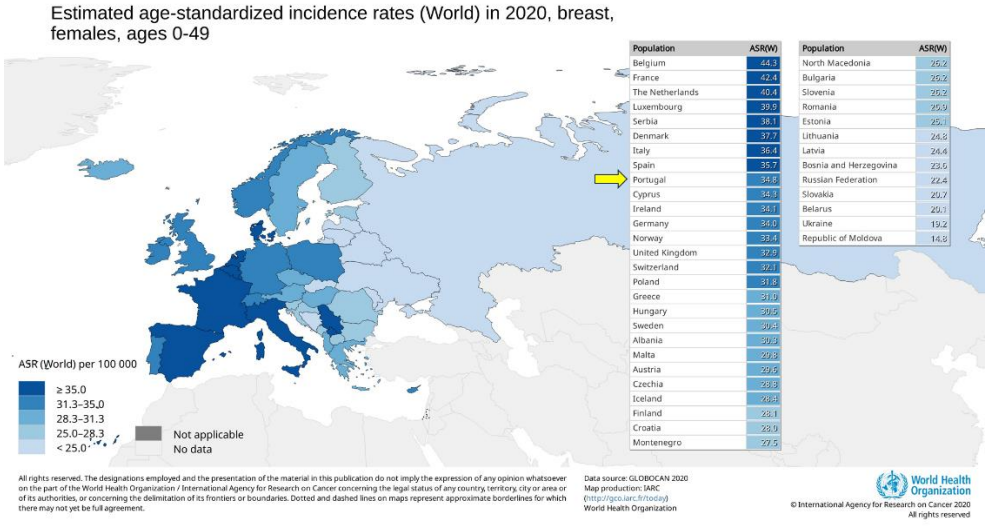


Figure 9| 2020 BC incidence in European populations. The shading indicates estimated age-standardized incidence rates per 100,000 persons based on data from GLOBOCAN 2020; Accessed on 6 January 2024; ©International Agency for Research on Cancer, 2020 (Sung et al., 2021a),

In Portugal, BC was the most incident type of cancer affecting females with 7425 new cases diagnosed. The districts that are more affected by BC are Coimbra, followed by Madeira and Vila-Real (**Figure 10**). All of them have an ASIR for all ages above 100/100,000.

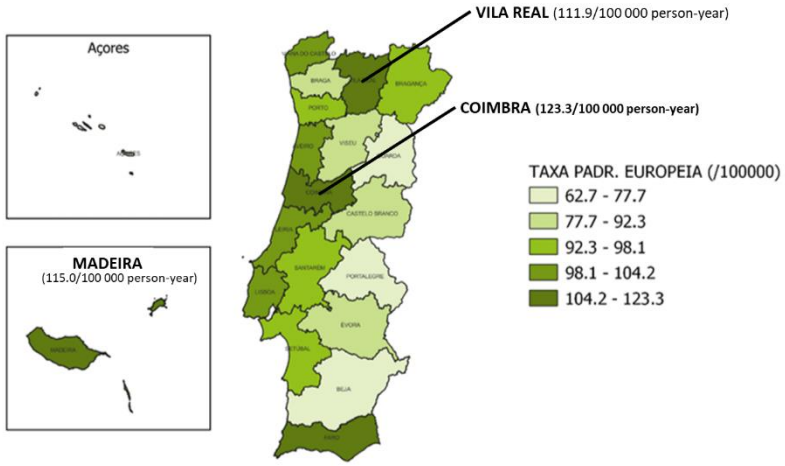


Figure 10| Geographical distribution for female BC in Portugal 2020. Adapted from: RON. Registo Oncológico Nacional de Todos os Tumoures na População Residente em Portugal, em 2020. Instituto Português de Oncologia do Porto Francisco Gentil - EPE, ed. Porto, 2023

Despite the evidence of BC heterogeneity from clinical, pathological, and molecular data, most epidemiological studies disregard BC subtypes and consider it as a single entity with several risk factors (Abhishek & Vidhyarthy, 2023; Forjaz de Lacerda et al., 2018) Females diagnosed and treated with BC in Europe are older than in Asia and Latin America, and more likely to develop long-distance metastasis. Additionally, Lobular Breast Cancer (LBC) is more frequently diagnosed in Europe than in Asia or Latin America. In low-income and transitioning countries, detection rates are higher for locally advanced BC and metastatic lesions (Kadys et al., 2023)(Forjaz de Lacerda et al., 2018).

2.4| Lobular Breast Cancer Subtype: Pathogenicity, Clinical Manifestations, Incidence, and Beyond

Invasive LBC has specific clinical and biological features, typically tumours are estrogen and progesterone receptor-positive, HER2 negative, low proliferation (Ki67 low), and thus belong to the luminal A molecular subtype. One of the hallmarks of classic LBC is the loss of E-cadherin expression (frequently as the result of biallelic inactivation of the *CDH1* gene), originating a characteristic presence of small tumour cells that are loosely dispersed in the stroma or arranged in single infiltrating patterns (Christgen et al., 2016; McCart Reed et al., 2021). Interestingly, in addition to the typical sites of metastasis seen in breast cancer cases, LBC exhibits an unusual tendency to colonize the gastrointestinal tract, peritoneum, and ovaries (Mathew et al., 2017). Other special morphological subtypes of LBC exist (Pleomorphic, Solid, Alveolar, and Tubulo-lobular), which are considered variants of the classic LBC but frequently observed together. This diversity of *in-situ* and invasive lesions can result from distinct molecular alterations driving disease progression (McCart Reed et al., 2021). In classic LBC, the most common mutations across 5 different studies are represented in **Table 4**.

Table 4. Somatic alterations reported in LBC.

Somatic Alteration	Gene	LBC (n = 127)	LBC (n = 144)	LBC (n = 413)	LBC (n = 32)	LBC (n = 127)	**Average**
Mutations	*CDH1*	63,0%	42,7%	65,4%	53,0%	82,0%	**61,2%**
	PIK3CA	48,0%	33,8%	43,3%	44,0%	57,0%	**45,2%**
	CCND1	17,0%	15,0%	38,0%	33,3%	17,0%	**24,1%**
	FGFR1	NR	NR	25,3%	6,2%	7,0%	**12,8%**
	TBX3	9,0%	8,0%	13,3%	21,8%	10,0%	**12,4%**
	ARID1A	17,0%	7,0%	6,3%	12,5%	8,0%	**10,2%**
	FOXA1	7,0%	NR	9,0%	15,6%	8,0%	**9,9%**
	TP53	8,0%	3,6%	7,3%	18,7%	9,0%	**9,3%**
	ERBB2	7,0%	4,0%	0,0%	33,3%	2,0%	**9,3%**
	MAP3K1	6,0%	5,1%	5,0%	18,7%	10,0%	**9,0%**
	KMT2C	7,0%	10,0%	8,0%	NR	NR	**8,3%**
	GATA3	5,0%	5,1%	7,3%	15,6%	3,0%	**7,2%**
	RUNX1	10,0%	NR	3,4%	3,0%	9,0%	**6,4%**
	ERBB2	4,0%	4,3%	5,1%	12,5%	2,0%	**5,6%**
	ESR1	NR	1,4%	NR	12,5%	2,0%	**5,3%**
	PTEN	7,0%	1,4%	3,9%	3,1%	9,0%	**4,9%**
	AKT1	2,4%	5,1%	4,1%	6,2%	NR	**4,5%**
Amp	*ERBB3*	NR	2,9%	3,6%	3,0%	NR	**3,2%**
	NF1	NR	4,3%	1,0%	3,1%	2,0%	**2,6%**
	CBFB	2,0%	NR	NR	NR	NR	**2,0%**
Reported in:		Ciriello et al. [2015]	Michaut et al. [2016]	Desmedt et al. [2016]	Richard et al. [2020]	Pareja et al. [2020]	

Adapted from:(McCart Reed et al., 2021)

The genes more frequently mutated are *CDH1, PIK3CA,* and *CCND1*, while no gene amplification is particularly found in invasive LBC (Ciriello et al., 2015; Desmedt et al., 2016; Michaut et al., 2016; Pareja et al., 2020; Richard et al., 2020). Invasive LBC is particularly associated with *CDH1* germline alterations in DGLBCS and Hereditary Lobular Breast Cancer (HLBC) syndrome. Usually, in HLBC cases LBC onset is earlier than in typical DGLBCS families.

Clinical symptoms for LBC may include a thickened or swollen area within the breast. This may be accompanied by alterations in the nipple, such as inversion or a change in shape. Another characteristic is the development of dimpling on the skin over the affected region. Breast magnetic resonance imaging is used as the most accurate imaging system to diagnose LBC, followed by breast biopsy for confirmation. Current treatment relies on initial surgery with endocrine therapy in the neoadjuvant setting. After surgery, patients may need adjuvant radiotherapy, chemotherapy, and/or endocrine therapy(Luveta et al., 2020).

Of the invasive BC histological special subtypes, invasive Lobular Breast Carcinoma (LBC) is the most frequent, accounting for nearly 10% of all BC cases. It is more associated with multifocal, multicentric, and bilateral lesions (Dossus & Benusiglio, 2015).Previous studies have documented an increase in LBC incidence, attributed to advances in diagnostic methods. This was followed by a subsequent decrease, associated with reduced use of hormone replacement therapy to alleviate menopausal symptoms in women (Biglia et al., 2007; Dossus & Benusiglio, 2015; Eheman et al., 2009). Enhancing our comprehension of the distinctive molecular characteristics of invasive LBC will allow us to establish an improved biomarker profile and lead to more effective treatments.

Hereditary Lobular Breast Cancer syndrome was defined to classify patients with LBC carrying *CDH1* germline alteration that do not present GC family history, in particular DGC.

From this point onward, DGLBC syndrome and its intricacies take center stage as the main interest of this thesis.

3| DGLBCS (Diffuse Gastric Lobular Breast Cancer Syndrome) previously known as Hereditary Diffuse Gastric Cancer (HDGC): Molecular Genetics

Diffuse Gastric Lobular Breast Cancer Syndrome is an autosomal dominant TRS, whose families have an increased risk of developing DGC and LBC. The definition and name of the disease evolved over the years, concomitantly with the clinical criteria driving genetic testing. Classically, DGLBCS is known for its association with germline variants in the *CDH1* gene, and more recently *CTNNA1* variants (Blair et al., 2020; Guilford et al., 1998; Majewski et al., 2013)

3.1| Causal genes associated with DGLBCS

The diagnosis of DGLBCS can be established in an individual with genetic testing clinical diagnostic criteria or by an unexpected identification of a germline heterozygous pathogenic variant in *CDH1* or *CTNNA1* by cancer gene panels. Consensus genetic testing criteria have been published for DGLBCS suspected individuals (Blair et al., 2020). Current genetic testing criteria consider familiar and individual conditions and even contemplate the presence of DGC in patients with personal or family history of cleft lip and/or cleft palate. Individuals without a known pathogenic variant in *CDH1* or *CTNNA1* but meeting genetic testing criteria should be recognized as having Diffuse Gastric and Lobular Breast Cancer Syndrome of Unknown Genetic Cause (this is also referred to as DGLBCS-like or HDGC-like) (**Figure 2B**). Despite *CDH1* being the major causative gene in DGLBCS

context, the majority of the families do not have an established genetic cause remaining molecularly unresolved (Blair et al., 2020).

Molecular genetic testing approaches can include a combination of gene-targeted testing (serial single gene testing, multigene panel) and comprehensive genomic testing (exome sequencing, genome sequencing). In a gene-targeted approach, the clinician decides which gene(s) to test for the presence of a single nucleotide variant (SNV), copy number variants (CNV), and structural variants (SV) associated with the disease phenotype. Sequence analysis and deletion/duplication analysis for *CDH1* are performed first. If no pathogenic variant is detected, an analysis of *CTNNA1* should be performed. Importantly, in individuals without a pathogenic variant identified at the DNA level, RNA analysis of *CDH1* and *CTNNA1* may help to identify pathogenic variants that cause splicing defects (Blair et al., 2020; Cosma et al., 2022)

Additionally, a multigene panel that includes *CDH1*, *CTNNA1*, and other genes of interest may be considered by clinicians to identify the disease's genetic cause or requested as direct-to-consumer testing. However, this condition may be associated with the identification of variants of uncertain significance, and/or pathogenic variants in genes that are not associated with the underlying phenotype (Blair et al., 2020; Clark et al., 2020).

3.1.1| *CDH1* and *CTNNA1*

The ***CDH1*** gene encodes the E-cadherin protein, a type I transmembrane glycoprotein, that constitutes adherent junctions in a Ca^{2+}dependent manner (Harris & Tepass, 2010). This protein comprises three distinct functional domains: a cytoplasmic domain, a transmembrane domain, and an extracellular domain. The cytoplasmic domain engages with p120 and β-catenin proteins to facilitate stabilization, while the extracellular domain establishes homophilic interactions with E-cadherin proteins on adjacent cells (Decourtye-espiard & Guilford, 2023; Wijnhoven et al., 2000). Loss of function of this gene/gene product in the epithelial tissue impairs cell adhesion and promotes the migratory and invasive potential in cancer cells (Aladin et al., 2021; Decourtye-espiard & Guilford, 2023). This tumour behaviour is considered a hallmark of DGC and LBC in DGLBCS families(Bruner & Derksen, 2018; Graziano et al., 2003).

Depending on the geographical region, *CDH1* SNVs or CNVs occur in 10–40% of DGLBCS families (**Table 5**) (Guilford et al., 1998; Oliveira, Senz, et al., 2009). Most DGLBCS-associated variants are truncating SNVs identified in the nonsense-mediated decay (NMD) competent region. These include frameshift, nonsense, and numerous splice-site variants. Truncating SNVs have a clear clinical significance and are thus generally considered actionable in DGLBCS families. Nevertheless, large deletions, encompassing both 5′- and 3′-regulatory or coding regions of the *CDH1* gene were also identified (Table 1) (Oliveira, Senz, et al., 2009). So far, no hotspot was identified in *CDH1* (Oliveira et al., 2015a). In two large studies, a germline *CDH1* pathogenic variant was identified in 0.1%–0.3% of women with LBC (Adib et al., 2021; Veen et al., 2022). Approximately 7% of individuals with DGC have been found to have a germline *CDH1* pathogenic variant (Adib et al., 2021).

Table 5. Molecular Genetic Testing Used in DGLBCS

Gene	Proportion of DGLBCS Attributed to Pathogenic Variants in Gene	Proportion of Probands with a Pathogenic Variant detectable by:	
		Sequence analysis [1]	Gene-targeted deletion/duplication analysis [2]
CDH1	10–40% [3]	82–95% [3]	5–18% [3]
CTNNA1	<2% [4]	100% [4]	Not available
Diffuse Gastric and Lobular Breast Cancer Syndrome of Unknown Genetic Cause [5]	>50% [3]	Not Available	

1. Sequence analysis detects variants that are benign, likely benign, of uncertain significance, likely pathogenic, or pathogenic. Variants may include small intragenic deletions/insertions and missense, nonsense, and splice site variants; typically, exon or whole-gene deletions/duplications are not detected.
2. Gene-targeted deletion/duplication analysis detects intragenic deletions or duplications. Methods used may include a range of techniques such as quantitative PCR, long-range PCR, multiplex ligation-dependent probe amplification (MLPA), and a gene-targeted microarray designed to detect single-exon deletions or duplications.
3. (Lobo et al., 2021)(Hansford et al., 2015) (Kaurah et al., 2007)(Oliveira et al., 2015a) (Oliveira, Senz, et al., 2009)
4. (Lobo et al., 2021; Majewski et al., 2013)
5. Known as DGLBC-like or HDGC-like in (Blair et al., 2020)

Given *CDH1* is considered a tumour suppressor, its inactivation is expected to follow a two-hit model. It is expected that the wild-type allele is targeted,

through a second hit mechanism, and *CDH1* somatic expression is completely lost, leading to cancer development arise often at an early age (Oliveira et al., 2015a).

Frequent mechanisms causing somatic inactivation or down-regulation of *CDH1* expression in DGLBCS are promoter hypermethylation and somatic mutation (e.g., deletions) causing loss of heterozygosity (LOH) of *CDH1* (Decourtye-espiard & Guilford, 2023; Machado et al., 2001). Primary tumours frequently have *CDH1* promoter methylation, while somatic deletion of *CDH1* is more recurrent in lymph node metastases (Oliveira, Senz, et al., 2009).

The ***CTNNA1*** gene encodes for the αE-catenin protein, which functions as the primary protein link between cadherins, and the actin cytoskeleton (Drees et al., 2005; Takeichi, 2018). It interacts with several proteins in at least four signaling cancer-associated pathways (Wnt/B-catenin, NF-kB, Hippo-YAP, and Hedgehog)(Takeichi, 2018). CTNNA1 is reported to be essential in inhibiting cell proliferation, promoting apoptosis, and repressing invasion, and metastasis in malignancies (Takeichi, 2018). Recently it was suggested that pathogenic, but not likely-pathogenic variants in *CTNNA1* seem to be associated with DGC development. These pathogenic variants explain the genetic causal event in <2% of families/individuals fulfilling DGLBCS criteria **(Table 5)**. In *CTNNA1*-associated DGLBCS families, there is not a clear association with LBC or cleft lip, with or without cleft palate (Lobo et al., 2021). Other phenotypes have been associated with germline variants in *CDH1* (Blepharocheilodontic syndrome 1; OMIM 119580) and *CTNNA1* (Patterned macular dystrophy 2; OMIM 608970).

3.1.2| Variant Classification System – *CDH1*

Germline variants classification is challenging, and given the diversity of variant types in DGLBCS, it remains a controversial territory in clinical practice. To standardize *CDH1* variants' classification and to attribute a value to its actionability, specific variant curation guidelines were established for the analysis of germline *CDH1* sequence variants, according to the American College of Medical Genetics and Genomics and the Association for Molecular Pathology (ACMG/AMP). 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 are

considered clinically actionable in DGLBCS disease context (Blair et al., 2020; K. Lee et al., 2018; Richards et al., 2015). Currently, one of the biggest challenges is predicting VUS actionability, particularly of missense variants, due to the scarce clinical data available for these types of variants. There is a high frequency of missense variants found in DGLBCS families, it is of uttermost importance to understand which of them may induce cryptic splicing with severe consequences such as premature truncation (Corso et al., 2021). This would be very important for the management of these families since 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).

Classification using ACMG criteria is not available for *CTNNA1* variants, thus, these have been mainly classified using the general ACMG criteria (Richards et al., 2015). The lack of specific guidelines complicates the prediction of actionability and their interpretation when occurring in patients and families (Lobo et al., 2021).

Despite *in vivo* and *in vitro* assays that can demonstrate causality and/or loss of function mechanisms, pathogenicity classification is very challenging since these experiments have limited assessment and are interpreted in a single, oversimplified context. This is particularly relevant for variants with very low penetrance, in which most carriers will not develop disease (Ciesielski et al., 2024).

3.1.3| Founder effects in DGLBCS

As previously mentioned, DGC and LBC incidence highly vary according to the geographical region and/or ethnic differences. One possible explanation for the observation of high-risk populations for specific inherited diseases is the effect of a founder mutation. Founder effects result from a reduction in the genetic variation in a population, which might be caused by a bottleneck effect (related to environmental events) or due to the isolation of a few members of a certain population giving rise to a new colony/community (Neuhausen, 2000). These events allow for a rare pathogenic variant to become more frequent, and its identification has a huge impact on genetic screening and disease surveillance in these areas (Hakkaart et al., 2019).

The first description of a founder mutation was reported in multiple families from Newfoundland in Canada (Benesch et al., 2021). Additionally, five distinct *CDH1* germline alterations have been ID identified in the Māori population (**Table 6)**. Given the high frequency of such variants in the Māori in comparison with the low-risk GC considered for New Zealand, it is reasonable to imply a founder effect (Scalzi et al., 2019). It can be confirmed if the disease is co-segregating with shared regions of the genome where the variant of interest is located.

It is expected that better estimates of cancer risk are calculated in populations with founder mutations since access to a higher number of patients and families carrying a rare variant in the population is facilitated. Another advantage of their study is the diminished genetic heterogeneity and shared environmental factors between individuals. Overall, those are a great model of study for disease penetrance, genetic, and environmental modifiers, and may be crucial for developing a more personalized individual risk assessment (Neuhausen, 2000).

Table 6. DGLBCS Founder variants in *CDH1*

Gene	Reference Sequences	DNA Nucleotide Change	Predicted Protein Change	Observation (Reference)
CDH1	NM_004360.5 NP_004351.1	c.190C>T	p.Gln64Ter	Founder variants in the Māori population (Hakkaart et al., 2019)
		c.1792C>T	p.Arg598Ter	
		c.2287G>T	p.Glu763Ter	
		c.2386dupC	p.Arg796ProfsTer11	
		c.2398delC	p.Arg800AlafsTer16	Founder variant in the Newfoundland population (Kaurah et al., 2007)
		c.1901C>T #	p.Ala634ProfsTer7#	Founder variant in the Portuguese population (Barbosa-Matos et al., 2021)

This variant has been previously known as p.Ala634Val.

3.1.4| Genotype-phenotype associations

Germline ***CDH1*** pathogenic variants have been described in individuals with other primary cancers besides DGC and LBC (e.g., other types of GC, other types of BC, colorectal cancer (CRC), and others. In the 2020 guidelines, one of the questions that remain to be solved was the genotype-phenotype correlations in DGLBCS families (Blair et al., 2020).

In a US study of 212,944 individuals, 141 individuals (0.06%) had a germline *CDH1* pathogenic variant (Adib et al., 2021). Breast cancer (54.6% of individuals) and gastric cancer (39.7% of individuals) were the most common cancer types among individuals with a germline *CDH1* pathogenic variant; colorectal cancer **(CRC)** was reported in 9.9% of individuals (3.8% had signet-ring cell colorectal cancer), and other types of cancer were reported in 3.5% of individuals. In a second US study, there was no history of gastric or breast cancer in 12% of families with a germline *CDH1* pathogenic variant. Moreover, a significantly high frequency of these cancer-free individuals were carriers of missense variants. All variants identified in probands with GC were truncating, while for BC 7% of missense were identified (Xicola et al., 2019). No clear association with CRC was possible, despite being the third most frequent, after gastric and breast cancer in these families.

Interestingly, *CDH1* variants located in the linker regions between the extracellular domains were associated with cleft lip/palate. A specific protein region responsible for calcium ion chelation displayed high intolerance to missense substitutions (between amino acids 253-260)(Obermair et al., 2019; Selvanathan et al., 2020).

For ***CTNNA1***, due to its more recent discovery, data is still being collected. Nevertheless, a systematic review including 10541 individuals from 41 families with a germline *CTNNA1* truncating variant, found that only 24% of families met the criteria for DGLBCS, and affected individuals presented most often with DGC at around age 40 years. Affected individuals from the remaining 76% of families with a germline *CTNNA1* truncating variant presented other cancer types, mostly unspecified breast cancer, or remained cancer-free (Lobo et al., 2021).

3.1.5| Lifetime risk estimates for *CDH1* and *CTNNA1* carriers

Lifetime risk estimations reflect the cumulative risk for an individual to develop a certain disease over one's lifespan.

Nevertheless, there are some complexities in risk assessment for hereditary diseases, such as DGLBCS, regarding its incomplete penetrance (not all carriers develop disease), and variable expressivity (Oliveira et al., 2013).

Accurate estimates in the context of autosomal dominant conditions are highly valuable for genetic counseling and patient management (Kosztolányi, 2008). One major limitation of this type of study is the lack of large pedigrees and ascertainment bias. The selection criteria for genetic testing depends on having an affected individual, so families with multiple affected individuals tend to be more studied (Spargo et al., 2022).

Usually, most studies observe how the variant of interest is segregated with the disease phenotype within families or study disease incidence in a cohort of unrelated people who harbour a genetic variant of interest (Gong et al., 2010). Additional significant constraints include the lack of knowledge about the baseline risk for the disease in a specific population/ geographic location, or the allelic frequency of the variants under consideration in a particular population (Whiffin et al., 2017).

Distinct statistical methods can be used for estimating penetrance, a frequently employed method for cumulative risk calculations, due to its simplicity, is the Kaplan–Meier estimator. However, this approach is limited since it is impossible to correct for data ascertainment bias, and therefore lifetime risks are frequently overestimated (Jonker et al., 2019).

For *CDH1* several penetrance studies have been conducted. The differences observed regarding lifetime risks could arise from the study population, the mutation types included and the overall analysis strategy. However, since different penetrance studies lead to different risk estimates it is challenging for risk counseling to provide a reasonable value for patients and their families (Turner & Jackson, 2020).

The first families with *CDH1* germline alterations were reported in New Zealand in 1998 (Guilford et al., 1998). From the beginning DGLBCS families were described, based on the kindred of Māori descendants, to be associated with an

autosomal dominant inheritance of a susceptibility gene that displayed incomplete penetrance.

The first penetrance/cumulative risk study was conducted in 2001, with 11 families carrying 11 distinct mutation types (three splice-site, five non-sense and three frameshift) in a total of 476 individuals. The selection criteria were highly enriched in DGC cases since an eligible family had to present three cases of DGC and at least one *CDH1* carrier. At least 80 GC cases (average age of onset of 40 years) being 44 confirmed DGC cases and 2 unspecified carcinomas. For BC, there were two confirmed LBC in a total of seven. For the age of 80 years, the cumulative risk of GC was 67% (95% Confidence Interval [CI] 39-99) and 83% (95% CI 58-100) for women, while BC risk in women was 39% (95% CI 12-84). The authors recognize an overestimation of risk due to ascertainment bias, since in other studies like BRCA1 and BRCA2 estimates derived from family-based studies were higher than population-based studies (Pharoah et al., 2001). Another study from 2007, considering clinically detected gastric and breast cancer cases, calculated the penetrance of a specific variant (2398delC) present in four families, two of them from a *CDH1* founder population in Newfoundland, North America. Their calculated risk at 75 years of age for GC was 40% (95% CI 12-91) for males and 63% (95% CI 19-99) for females while BC risk was 52% (95% CI, 29-94) (Kaurah et al., 2007). In 2015, penetrance studies from 75 families reported a GC risk of 70% (95% CI 59-80) for men and 56% (95% CI 44-69) for women by the age of 80 years, while BC risk was 42% (95% CI 23-68). In this study, all *CDH1* carriers had DGLBCS criteria (Hansford et al., 2015).

More recently, penetrance studies for *CDH1* carriers were carried out with population-based studies, using *CDH1* carriers independently of DGLBCS criteria. This analysis included 25 families fulfilling DGLBCS criteria out of 41 used for risk estimations. The risk of GC was 42% (95% CI 30-56) for men and 33% (95% CI 21-43) for women by age 80 years. For BC, the cumulative risk was 55% (95% CI 39-68). Since multiple *CDH1* variants were considered, the impact of each variant is the same. However, different *CDH1* germline mutations can be more penetrant than others (São José et al., 2023).

Initial estimates from individuals from 13 families with a germline *CTNNA1* truncating variant were performed, showing a cumulative risk of DGC of 49% and 57%, by age 80 years depending on the statistical method used (Coudert et al., 2022)

3.2| Candidate genes

Several strategies have been applied to identify the molecular cause behind hereditary syndromes, such as next-generation sequencing (NGS) technologies and functional screening. By definition, a candidate is defined as a gene whose known biological function is deemed to have a direct or indirect regulation of the manifestation of a particular disease phenotype (Zhu & Zhao, 2007). For example, gene expression signatures can be used to identify signalling pathways relevant to disease development and progression (Dominguez-Valentin et al., 2013). Many cancer-predisposing genes were identified by genome-wide linkage analysis with sequencing to identify familiar segregation of pathogenic variants in candidate *loci*. However, this strategy is limited to rare diseases with few family members suitable for linkage analysis. With the development of NGS, candidates for gene sequencing were selected based on having similar functions or being involved in shared molecular pathways, as a targeted approach, but many failed to be associated with hereditary cancer syndromes (Alenezi et al., 2020).

Importantly, also due to the implementation of NGS technologies, such as whole-exome sequencing and whole-genome sequencing, many variants of unknown clinical significance are identified. Therefore, is crucial to perform a functional validation of the identified variant or gene for proof of pathogenicity (Rodenburg, 2018). Currently, non-coding regions of the genome and regulatory landscapes in the loci of disease causal genes are being explored to disclose the missing heritability in tumour risk syndromes such as DGLBCS (São José et al., 2023).

Another interesting approach, more common in cancers frequently occurring, such as BC, are genome-wide association studies. These can be performed to identify novel disease-associated risk with moderate and low penetrance. Unexpectedly or not, many of those hits are well-known cancer genes, frequently mutated at the somatic level. For example, *FGFR2* and *MAP3K1* were reported to increase BC risk and were frequently mutated in several types of cancer including BC (Bonifaci et al., 2010).

Given that 1-40% of the families suspected of DGLBCS present *CDH1* alterations and ~2% present *CTNNA1* germline mutations, there is still a huge portion of families considered genetically undiagnosed or DGLBCS-like (Table 5). These families have to present at least one confirmed DGC plus another GC case

or LBC (before 70 years) case in a first or second-degree relative (Blair et al., 2020). It is urgent to understand the genetic cause behind these cases, so that appropriate genetic screening studies, surveillance and risk reduction measures can be applied. An effective genetic screening is crucial to identify families with hereditary predisposition and reduce the risk of advanced disease due to late diagnosis(van Dieren, Jolanda, Bisseling et al., 2023).

4| Clinical management of DGLBCS families

Clinical practice guidelines have been 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).

Management of *CDH1*-DGLBCS aims to prevent DGC and LBC development. Despite not being too frequent, cleft lip with or without cleft palate has also been included as part of DGLBCS disease spectrum (**Table 7**). Information regarding the type of clefting and severity is limited, and its presence highly varies among affected individuals from the same family. Audiology and speech are evaluated in affected individuals(Green et al., 2022).

Table 7. DGLBCS Frequency of disease phenotypes

Disease phenotype	**Frequency of disease phenotype by gene**	
	CDH1	***CTNNA1***
DGC	42%–70% of males [1] 33%–56% of females [1]	49%–57% [2] both males and females
LBC	42% of females [1]	NA
Cleft lip w/or w/o cleft palate	<3% [3]	NA

NA = not reported to date
1.(Roberts et al., 2019)
2. (Coudert et al., 2022)
3. Cleft lip/palate has been described in at least one individual in 19% of families with *CDH1*-related DGLBCS (Green et al., 2022)

4.1| Surveillance and risk reducing approaches for asymptomatic carriers.

The uncertainty regarding cancer risk should always be kept in mind. Therefore, available surveillance measures should be offered based on family history and biopsy/surveillance results, while discussing risk-reduction measures.

In *CDH1* VUS carriers: gastric and/or breast surveillance are offered on a case-by-case basis. Risk-reduction gastrectomy or mastectomy is not recommended, unless in the presence of a heavy family history only. However, the relatives of the carrier do not have indications for genetic tests and are not clinically followed. In DGLBCS-like families, cancer screening age can be delayed or intervals between screenings can be extended if no positive findings are observed. Again, risk-reduction surgeries are less likely to be recommended unless the family history is strong (Blair et al., 2020).

4.1.1| For DGC

Endoscopy allows a direct inspection of the stomach, and several biopsies are collected of suspicious areas; however, advanced DGC lesions mostly grow into the submucosa, becoming hard to identify through a biopsy. Signet ring cancer (SRC) cells are often identified at the initial endoscopy in individuals with a germline *CDH1* pathogenic variant, being reported frequently in cases without a first-degree relative diagnosed with DGC. (Gullo et al., 2020; Jacobs et al., 2019). The decision between long-term surveillance for DGC or risk reduction surgery is challenging since until now it is not clear which or when intramucosal lesions will progress into a more aggressive and deadly state. Recent studies support the safety of relying on endoscopic surveillance if performed at expert centers to detect early DGC (Asif et al., 2023; Friedman et al., 2021; C. Y. C. Lee et al., 2023). However, more studies with greater statistical power are needed before this approach is implemented in clinical practice.

Currently, there are ongoing trials to investigate the utility of probe-based confocal laser endomicroscopy (pCLE) for DGC diagnosis, particularly in identifying early SRCC lesions (Jacobs et al., 2019). According to the 2020 guidelines, the following surveillance measures are recommended for pathogenic and likely-pathogenic *CDH1* and *CTNNA1* variant carriers:

Table 8. *CDH1*- or *CTNNA1*-Related DGLBCS recommended surveillance measures.

Target organ	Evaluation	Observations
Stomach	Referral to a high-risk gastric-screening program	• After a positive genetic test
	Detailed 30-minute upper endoscopy w/multiple random biopsies & biopsies of subtle lesions [1]	• Every 6-12 months beginning at age 18 yrs (or 5-10 yrs prior to the earliest gastric cancer diagnosis in the family (minimum age of 18 yrs). • If the risk for DGC is unclear, the endoscopist has the discretion to extend surveillance intervals after two consecutive negative endoscopies. This decision is informed by individual findings in previous endoscopic examinations and consideration of the patient's family history.

1. The endoscopy surveillance protocol for International Gastric Cancer Linkage Consortium (IGCLC; Cambridge method) recommends a thorough high-definition white light examination of at least 30 minutes in a DGLBCS center of expertise. The session should begin with targeted biopsies from all suspicious lesions, followed by five random biopsies from six specific anatomical regions (prepyloric, antrum, transitional zone, body, fundus, and cardia) (Buckley et al., 2023).

Risk-reducing total gastrectomy for DGC is an option from early adulthood in those with normal endoscopy/gastric biopsies in individuals with: DGLBCS-related *CDH1* pathogenic variant regardless of family history; DGLBCS-related *CTNNA1* pathogenic variant and a family history of DGC and/or LBC. Risk-reducing total gastrectomy is not typically recommended in individuals with *CTNNA1*-related DGLBCS and no family history of DGC. It is possible that molecularly unsolved individuals who meet genetic testing criteria of DGLBCS and undergo risk-reduction measures before a causing genetic event is discovered, turn out negative once the cause is identified (São José et al., 2023). This is the reason why risk-reduction surgeries are less likely to be recommended and should be done on a case-by-case basis. Total gastrectomy with a multidisciplinary team including gastric surgery, gastroenterology, pathology, nutrition, oncology, and palliative care for preoperative and postoperative care. The multidisciplinary team members can counsel individuals on the risks and benefits of total gastrectomy. There should be a careful approach with reduced emphasis on prophylactic total gastrectomy, considering endoscopic surveillance in cases where the risk for DGC is unclear. This

includes individuals such as carriers of a *CDH1* PV/LPV who do not meet genetic testing criteria, carriers of a CTNNA1 PV/LPV, those with a family history or personal history of DGC and a *CDH1* VUS, and individuals with Diffuse Gastric and Lobular Breast Cancer Syndrome of Unknown Genetic Cause and their first-degree relatives. Additionally, total gastrectomy in young children is not recommended until full physical maturity is reached and should be avoided in adolescents, due to its high risk of complications. Also, it is not typically recommended for individuals older than age 70 years (Blair et al., 2020).

The morbidity from risk-reducing total gastrectomy is 100% (Muir et al., 2016; Worster et al., 2014). All individuals have immediate and long-term complications including rapid intestinal transit, dumping syndrome, diarrhea, eating habit alterations, and weight loss [van der Kaaij et al 2018, Gallanis & Davis 2023]. The risk for malabsorption increases, including an increased incidence of osteoporosis, osteomalacia, and malnutrition. Individuals who have undergone risk-reducing total gastrectomy in expert centers express minimal to no regret, emphasizing that the perceived benefits outweigh the associated risks [Kaurah et al 2019, Hallowell et al 2017, Muir et al 2016].

Following identification of DGC on gastric biopsy, treatment with total gastrectomy is an option. (example in: São José et al 2023). The affected individual and his/her physicians should consider to be tested and treated for H. pylori infection (if identified) to eradicate it prior to total gastrectomy. Current guidelines suggest that individuals with a DGLBCS-related *CDH1* pathogenic variant should undergo a baseline upper endoscopy before risk-reducing total gastrectomy.

Following total gastrectomy, surveillance with laboratory examinations and clinical, nutritional, and psychological monitoring is recommended. Laboratory assessments should include a complete blood count, electrolyte levels, blood urea nitrogen, creatinine, and evaluations of liver function. An annual endoscopy to survey the anastomosis may be considered. Vitamin and mineral supplementation are required due to the increased risk of nutrient deficiency (e.g., vitamin B12, iron, calcium, and vitamin D) [Stillman et al 2022]

4.1.2| For LBC

Target organ	Evaluation	Observations
Breast (in those w/ *-CDH1*-related DGLBCS or *-CTNNA1*-related DGLBCS & personal or family history of breast cancer)	Referral to a high-risk breast-screening program	
	Self-breast examinations	Monthly in females beginning at age 20 yrs
	• Clinical breast examinations • Education on clinical features of LBC (e.g., thickening, indrawn nipple, or change in breast skin).	Annually in females beginning at age 30 yrs
	Bilateral breast MRI w/contrast [3,4]	Annually in females from age 30-40 years
	Bilateral breast MRI w/contrast combined w/mammography & breast ultrasound	Annually in females beginning at age 40 yrs

Mammography has a low sensitivity for detecting lobular breast cancer. MRI with or without mammography and ultrasound is effective in detecting early LBC lesions during surveillance (Corso et al 2023).

For *CTNNA1* carriers (according to DGLBCS criteria), lifetime risk for BC has not yet been studied, both surveillance and risk reduction mastectomies should be considered on a case-by-case basis, with emphasis on personal and family history of breast cancer.

4.2| Treatment of clinical manifestations and supportive care

Clinical trials, targeted therapies and predictive markers of therapy response directed to DGC or LBC are scarce. For DGC, tumour resection is still the main curative treatment for early and advanced stages of the disease and therapeutic agents are mainly used as palliative control (Corso et al 2023).

4.2.1| For DGC

The average age of onset of DGC in the largest cohorts of individuals with *CDH1*-related DGLBCS carrier families ranges from age 42 to 47 years (Xicola et al 2019, Blair et al. 2020). DGC onset may occur as young as age 18 years (Gullo et al 2018, Xicola et al 2019, São José et al 2023).

A single study assessed the survival of individuals with *CDH1*–related DGC and reported these individuals had shorter survival than those who met genetic testing clinical criteria but did not have a germline *CDH1* pathogenic variant (van der Post et al 2015). Disease-free survival of individuals with *CDH1*- and *CTNNA1*-related advanced DGC is believed to be the same as individuals with advanced sporadic DGC. The five-year survival is less than 30% when DGC is identified at a late stage (Stiekema et al 2013, Pereira et al 2022).

Ongoing studies are mainly investigating alternate medication doses and combining therapeutic agents that are approved for sporadic gastric cancer and other solid tumours (e.g., platinum compounds, fluropyrimidines, and topoisomerase I inhibitors). Immunotherapy for anti-HER2, anti-VEGDR2, and anti-PD1 has been approved for treatment of gastric adenocarcinoma however, treatment data for DGC is weak. The MONO study (NCT01197885), a Phase II clinical trial, examined the outcome of zolbetuximab (monoclonal antibody against CLDN18.2: IMAB362) monotherapy in a series of individuals with recurrent or refractory, locally advanced or metastatic, CLDN18.2-positive gastric, gastro-esophageal junction, or esophageal adenocarcinoma. Of the 54 individuals enrolled in this trial, 22 had advanced DGC with CLDN18.2 expression in ≥50% of tumour cells. Zolbetuximab was well-tolerated and exhibited antitumour activity with a clinical benefit rate of 23% (Türeci et al 2019); whether individuals benefiting from this targeted therapy had DGC was not reported.

4.2.2 |For LBC

The average age of onset for BC in females with a germline *CDH1* pathogenic variant ranges from 45 to 54 years (Pharoah et al 2001, Xicola et al 2019)

The reported five-year overall survival rate for infiltrating LBC is approximately 93%, however, the impact of a germline *CDH1* pathogenic variant on

the survival rate of individuals with LBC remains uncertain (Han et al 2022). While some studies have suggested that a somatic *CDH1* variant does not significantly affect the prognosis of individuals with invasive LBC, the presence of both *CDH1* and *ERBB2* variants in LBC tumour tissue has been associated with a poorer prognosis (Ping et al 2016). There are active clinical trials for individuals with invasive LBC or enriched in individuals with lobular histology. These studies tested small molecules including tyrosine kinase ROS1 inhibition, CDK4/6 inhibition, endocrine therapy strategies, immunotherapy with checkpoint inhibition, and inhibition of HER2 in individuals with HER2 alterations (Mukhtar et al 2021). Novel imaging tools are also being studied to decrease rates of positive margins and improve surgical outcomes (Jones et al 2020).

4.3| Challenges in management of families

Once a molecular diagnosis of DGLBCS has been established in the proband, it is appropriate to offer at-risk relatives predictive testing for the familial *CDH1* or *CTNNA1* pathogenic variant. Most individuals diagnosed with *CDH1*-related DGLBCS inherited the *CDH1* pathogenic variant from a parent. Due to incomplete penetrance or to the early death of the parent before the onset of symptoms, the family history of an individual with DGLBCS may appear to be negative because of failure to recognize the disorder in family members. Intrafamilial clinical variability, differences in age of onset and manifestations may differ among sibs with the same *CDH1* or *CTNNA1* pathogenic variant (Blair et al., 2020; Gullo, Devezas, et al., 2018).

The uncertainty regarding disease risk, the lack of efficient therapeutic options, the missing mechanism of disease progression and the great percentage of molecularly undiagnosed families are the top burning questions in the field (Blair et al., 2020).

5| Study models in DGLBCS

In DGC research, diverse *in vitro* and *in vivo* models have been employed while trying to unravel the complexities of this disease. Several in vivo models such as in *Drosophila* systems, have been instrumental in elucidating the effect of certain genetic variations and its associated risk for DGC, shedding light on disease susceptibility (Wang et al., 2010). Additionally, *in vivo* studies have highlighted the aggressive nature of DGC, emphasizing the urgency for effective therapeutic interventions (Nadauld et al., 2014; Mitchell et al., 2017).

Many research models have been developed to study cancer development in DGLBCS families, however, most studies have relied on immortalized cell lines, genetically engineered animal models, or mouse tissue-derived organoids. Recently, extensive reviews on synthetic lethal models and mouse models for DGC development have been performed (Lim et al., 2023; J. L. Monster et al., 2022). Most cell line models derive from patients with advanced or metastatic disease and have been in culture for very long periods, sometimes without information on the germline alterations or specific tumor characteristics (Nanki et al., 2018). The genetic mouse models, despite being useful for understanding novel details on GC pathogenesis, fail to recapitulate the human genetic background. The discovery of reprogramming human somatic cells to induced-pluripotent stem cells by Kazutoshi Takahashi and Shinya Yamanaka allowed us to overlap some of the previous limitations, revolutionizing disease modeling and taking us one step closer to translating research findings to clinics(Nanki et al., 2018).

Human iPSCs provide an unlimited supply of tissue-specific cell types and are extremely useful in the context of complex genetic diseases for disease development studies, drug testing, or toxicity (Anderson & Francis, 2018). Moreover, human age does not affect somatic cells reprogramming which allows to generate iPSCs from donors of different ages carrying a particular germline alteration (Pitrez et al., 2024). To date, only one patient-iPSCs was generated from a *CDH1* carrier affected by both gastric and colon carcinomas. No reports exist for asymptomatic individuals or family biobanks (Li et al., 2021).

Patient-derived gastric cancer organoids proved to have similar transcriptomic profiles to their matching in vivo tumours, but most DGC cases are sporadic or a *CDH1* germline alteration is not described. Organoid models have

been utilized to model tumour molecular profiling and evolution (Nadauld et al., 2014). These models have also been employed to analyze the genomic profiles of DGC, providing insights into the aggressive progression of this cancer subtype (Lee et al., 2014). Several models are now dedicated to the study of early development of DGC in *CDH1* mutation carriers, trying novel approaches to improve patient screening and management (Carneiro et al., 2004). Mouse models of *CDH1* and *TP53* deficiency have been used to study the progression of early gastric signet-ring cell carcinoma into advanced DGC, shedding light on the molecular mechanisms underlying the initiation of DGC (Humar et al., 2009).

In vivo xenograft models have been utilized to investigate the effects of estrogen on tumour growth in DGC, revealing its potential role in aggravating tumour progression (Lee et al., 2021). In the context of drug resistance, both *in vitro* and *in vivo* models have been used to explore mechanisms underlying chemoresistance in DGC, shedding light on potential targets for sensitizing chemotherapy (Zhang et al., 2020; Peng et al., 2019).

In summary, the utilization of in vitro and in vivo models, including organoids, mouse models, and cell lines, has significantly advanced our understanding of Diffuse Gastric Cancer. These models have provided valuable insights into the genetic, molecular, and environmental factors influencing DGC susceptibility, progression, and treatment response, highlighting the importance of a multifaceted approach in DGC research.

6| Challenging the Mendelian inheritance with genetic modifiers of disease

Both incomplete penetrance and variable expressivity are thought to be caused by a range of factors, including common variants, variants in regulatory regions, epigenetics, environmental factors, and lifestyle. Many thousands of genetic variants have been identified as the cause of monogenic disorders, mostly determined through small clinical studies, and thus, the penetrance and expressivity of these variants may be overestimated when compared to their effect on the general population (Kousi & Katsanis, 2015; Rahit & Tarailo-Graovac, 2020).

6.1| Modified penetrance by coding variants (SNVs and CNVs)

Deleterious genomic alterations can be caused by single nucleotide variants (SNVs), copy number variants (CNVs) and even structural variants (SVs). These variants can be present in unaffected individuals and only lead to disease development when co-occurring with other known causative variants present in affected carriers. At an even higher complexity level, this effect may require the combination of moderate and/or low effect variants to be recognized as a disease manifestation. Additionally, the same modifier locus can be associated with multiple phenotypes (Glessner et al., 2017; Jensen & Girirajan, 2017).

On the other side, genetic modifiers can also protect against the presence of a disease predisposing variant by direct or indirect interaction to reduce its penetrance or effect (Andrews et al., 2019).

Whole-genome sequencing analysis give us access to almost all the genomic alterations present in and individual, including coding and non-coding regions. It has become an attractive approach specially in the context of rare diseases. Moreover, a link can be established with the transcriptome to reach a more precise understanding of the variant effect. In this type of approaches, variant prioritization is essential. It will lead to the selection of candidate variants based on a knowledge-driven approach considering a list of relevant genes or based on a set of filters considering predefined parameters (variant allele frequency, variant classification, alignment quality and others)(Rahit & Tarailo-Graovac, 2020)

For LBC, six common variants in the UK population were strongly associated with invasive LBC and in situ lesions but not with the ductal type. The strongest associations were observed for rs11977670 (7q34) considered a lobular specific single nucleotide polymorphism. Unfortunately, the same type of data is not available for DGC (Sawyer et al., 2014).

7| References

Abhishek, K., & Vidhyarthy, A. K. (2023). *A CLINICO EPIDEMOLOGICAL STUDY OF BREAST CANCER WITH SPECIFIC REFRENCE TO ER / PR*. 1313–1317. https://doi.org/10.47009/jamp.2023.5.2.277

Adib, E., Zarif, T. El, Nassar, A. H., Akl, E. W., Alaiwi, S. A., Mouhieddine, T. H., & Esplin, E. D. (2021). *CDH1 germline variants are enriched in patients with colorectal cancer , gastric cancer , and breast cancer*. *December*. https://doi.org/10.1038/s41416-021-01673-7

Ajani, J. A., Lee, J., Sano, T., Janjigian, Y. Y., Fan, D., & Song, S. (2017). Gastric adenocarcinoma. *Nature Reviews Disease Primers*, *3*. https://doi.org/10.1038/nrdp.2017.36

Aladin, D., Robinson, C., Dufour, S., Viasnoff, V., & Borghi, N. (2021). *Extracellular domains of E-cadherin determine key mechanical phenotypes of an epithelium through cell- and non-cell- autonomous outside-in signaling*. 1–22. https://doi.org/10.1371/journal.pone.0260593

Alenezi, W. M., Fierheller, C. T., Recio, N., & Tonin, P. N. (2020). Literature review of BARD1 as a cancer predisposing gene with a focus on breast and ovarian cancers. *Genes*, *11*(8), 1–24. https://doi.org/10.3390/genes11080856

Amin, M. B., Greene, F. L., Edge, S. B., Compton, C. C., Gershenwald, J. E., Brookland, R. K., Meyer, L., Gress, D. M., Byrd, D. R., & Winchester, D. P. (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: A Cancer Journal for Clinicians*, *67*(2), 93–99. https://doi.org/10.3322/caac.21388

Anderson, R. H., & Francis, K. R. (2018). Modeling rare diseases with induced pluripotent stem cell technology. *Molecular and Cellular Probes*, *40*(September 2017), 52–59. https://doi.org/10.1016/j.mcp.2018.01.001

Andrews, S. J., Fulton-Howard, B., & Goate, A. (2019). Protective Variants in Alzheimer's Disease. *Current Genetic Medicine Reports*, *7*(1), 1–12. https://doi.org/10.1007/s40142-019-0156-2

Apostolou, P., & Fostira, F. (2013). Hereditary breast cancer: The Era of new

susceptibility genes. *BioMed Research International*, *2013*. https://doi.org/10.1155/2013/747318

Arnold, M., Morgan, E., Rumgay, H., Mafra, A., Singh, D., Laversanne, M., Vignat, J., Gralow, J. R., Cardoso, F., Siesling, S., & Soerjomataram, I. (2022). Current and future burden of breast cancer : Global statistics for 2020 and 2040. *The Breast*, *66*(August), 15–23. https://doi.org/10.1016/j.breast.2022.08.010

Asif, B., Sarvestani, A. L., Gamble, L. A., Samaranayake, S. G., Famiglietti, A. L., Fasaye, G. A., Quezado, M., Miettinen, M., Korman, L., Koh, C., Heller, T., & Davis, J. L. (2023). Cancer surveillance as an alternative to prophylactic total gastrectomy in hereditary diffuse gastric cancer: a prospective cohort study. *The Lancet Oncology*, *24*(4), 383–391. https://doi.org/10.1016/S1470-2045(23)00057-8

Assumpção, P. P., Barra, W. F., Ishak, G., Coelho, L. G. V., Coimbra, F. J. F., Freitas, H. C., Dias-Neto, E., Camargo, M. C., & Szklo, M. (2020). The diffuse-type gastric cancer epidemiology enigma. *BMC Gastroenterology*, *20*(1), 1–7. https://doi.org/10.1186/s12876-020-01354-4

Barbosa-Matos, R., Silva, R. L., 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., … Oliveira, C. (2021). The cdh1 c.1901c>t variant: A founder variant in the portuguese population with severe impact in mrna splicing. *Cancers*, *13*(17), 1–18. https://doi.org/10.3390/cancers13174464

Bass, A. J., Thorsson, V., Shmulevich, I., Reynolds, S. M., Miller, M., Bernard, B., Hinoue, T., Laird, P. W., Curtis, C., Shen, H., Weisenberger, D. J., Schultz, N., Shen, R., Weinhold, N., Kelsen, D. P., Bowlby, R., Chu, A., Kasaian, K., Mungall, A. J., … Liu, J. (2014). Comprehensive molecular characterization of gastric adenocarcinoma. *Nature*, *513*(7517), 202–209. https://doi.org/10.1038/nature13480

Benesch, M. G. K., Bursey, S. R., O'connell, A. C., Ryan, M. G., Howard, C. L., Stockley, C. C., & Mathieson, A. (2021). Cdh1 gene mutation hereditary diffuse gastric cancer outcomes: Analysis of a large cohort, systematic review of endoscopic surveillance and secondary cancer risk postulation. *Cancers*, *13*(11), 1–18. https://doi.org/10.3390/cancers13112622

Benusiglio, P. R., Malka, D., Rouleau, E., Pauw, A. De, Buecher, B., Noguès, C., Fourme, E., Colas, C., Coulet, F., Warcoin, M., Grandjouan, S., Sezeur, A., Laurent-Puig, P., Molière, D., Tlemsani, C., Maria, M. Di, Byrde, V., Delaloge, S., Blayau, M., & Caron, O. (2013). CDH1 germline mutations and the hereditary diffuse gastric and lobular breast cancer syndrome: A multicentre study. *Journal of Medical Genetics*, *50*(7), 486–489. https://doi.org/10.1136/jmedgenet-2012-101472

Biglia, N., Mariani, L., Sgro, L., Mininanni, P., Moggio, G., & Sismondi, P. (2007). Increased incidence of lobular breast cancer in women treated with hormone replacement therapy: Implications for diagnosis, surgical and medical treatment. *Endocrine-Related Cancer*, *14*(3), 549–567. https://doi.org/10.1677/ERC-06-0060

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. di, … Guilford, P. (2020). Hereditary diffuse gastric cancer: updated clinical practice guidelines. *The Lancet Oncology*, *21*(8), e386–e397. https://doi.org/10.1016/S1470-2045(20)30219-9

Bonifaci, N., Górski, B., Masojć, B., Wokołorczyk, D., Jakubowska, A., Debniak, T., Berenguer, A., Musach, J. S., Brunet, J., Dopazo, J., Narod, S. A., Lubiński, J., Lázaro, C., Cybulski, C., & Pujana, M. A. (2010). Exploring the link between germline and somatic genetic alterations in breast carcinogenesis. *PLoS ONE*, *5*(11). https://doi.org/10.1371/journal.pone.0014078

Bruner, H. C., & Derksen, P. W. B. (2018). *Adhesion and the Development and Progression of Cancer*. *1*(Table 1), 1–20.

Buckley, K. H., Tondon, R., Guilford, P., & Katona, B. W. (2023). Non-targeted biopsies in hereditary diffuse gastric cancer: Necessary, but not enough. *Translational Cancer Research*, *12*(7), 1883–1886. https://doi.org/10.21037/tcr-23-541

Caldas, C., Carneiro, F., Lynch, H. T., Yokota, J., Wiesner, G. L., Powell, S. M., Lewis, F. R., Huntsman, D. G., Pharoah, P. D. P., Jankowski, J. A., MacLeod, P., Vogelsang, H., Keller, G., Park, K. G. M., Richards, F. M., Maher, E. R., Gayther, S. A., Oliveira, C., Grehan, N., … Jackson, C. E. (1999). Familial

gastric cancer: Overview and guidelines for management. *Journal of Medical Genetics*, *36*(12), 873–880.

Carneiro, F. (1997). Classification of gastric carcinoma. *Current Diagnostic Pathology*, *4*, 51–59. https://doi.org/10.1097/00008469-199908000-00033

Castro, C., Peleteiro, B., Bento, M. J., & Lunet, N. (2017). Trends in gastric and esophageal cancer incidence in northern Portugal (1994–2009) by subsite and histology, and predictions for 2015. *Tumori*, *103*(2), 155–163. https://doi.org/10.5301/tj.5000542

Christgen, M., Steinemann, D., Kühnle, E., Länger, F., Gluz, O., Harbeck, N., & Kreipe, H. (2016). Lobular breast cancer: Clinical, molecular and morphological characteristics. *Pathology Research and Practice*, *212*(7), 583–597. https://doi.org/10.1016/j.prp.2016.05.002

Ciesielski, T., Sirugo, G., Iyengar, S., & Williams, S. (2024). Characterizing the pathogenicity of genetic variants : the consequences of context. *Genome Medicine*, 1–11. https://doi.org/10.1038/s41525-023-00386-5

Ciriello, G., Gatza, M. L., Beck, A. H., Wilkerson, M. D., Rhie, S. K., Pastore, A., Zhang, H., McLellan, M., Yau, C., Kandoth, C., Bowlby, R., Shen, H., Hayat, S., Fieldhouse, R., Lester, S. C., Tse, G. M. K., Factor, R. E., Collins, L. C., Allison, K. H., … Perou, C. M. (2015). Comprehensive Molecular Portraits of Invasive Lobular Breast Cancer. *Cell*, *163*(2), 506–519. https://doi.org/10.1016/j.cell.2015.09.033

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. *Genetics in Medicine*, *22*(5), 840–846. https://doi.org/10.1038/s41436-020-0753-1

Correa, P., & Houghton, J. M. (2007). Carcinogenesis of Helicobacter pylori. *Gastroenterology*, *133*(2), 659–672. https://doi.org/10.1053/j.gastro.2007.06.026

Corso, G., Corso, F., Bellerba, F., Seixas, S., Cioffi, A., Vecchia, C. La, Magnoni, F., Bonanni, B., Veronesi, P., Gandini, S., & Figueiredo, J. (2021). *Geographical*

Distribution of E-cadherin Germline Mutations in the Context of Diffuse Gastric Cancer : A Systematic Review.

Corso, G., Marrelli, D., & Roviello, F. (2011). Familial gastric cancer: Update for practice management. *Familial Cancer*, *10*(2), 391–396. https://doi.org/10.1007/s10689-010-9410-1

Cosma, L. S., Schlosser, S., Tews, H. C., Müller, M., & Kandulski, A. (2022). Hereditary Diffuse Gastric Cancer: Molecular Genetics, Biological Mechanisms and Current Therapeutic Approaches. *International Journal of Molecular Sciences*, *23*(14). https://doi.org/10.3390/ijms23147821

Coudert, M., Drouet, Y., Delhomelle, H., Svrcek, M., Benusiglio, P. R., Coulet, F., Clark, D. F., Katona, B. W., Van Hest, L. P., Van Der Kolk, L. E., Cats, A., Van Dieren, J. M., Nehoray, B., Slavin, T., Spier, I., Hüneburg, R., Lobo, S., Oliveira, C., Boussemart, L., … Colas, C. (2022). First estimates of diffuse gastric cancer risks for carriers of CTNNA1 germline pathogenic variants. *Journal of Medical Genetics*, *59*(12), 1189–1195. https://doi.org/10.1136/jmg-2022-108740

Cristescu, R., Lee, J., Nebozhyn, M., Kim, K. M., Ting, J. C., Wong, S. S., Liu, J., Yue, Y. G., Wang, J., Yu, K., Ye, X. S., Do, I. G., Liu, S., Gong, L., Fu, J., Jin, J. G., Choi, M. G., Sohn, T. S., Lee, J. H., … Aggarwal, A. (2015). Molecular analysis of gastric cancer identifies subtypes associated with distinct clinical outcomes. *Nature Medicine*, *21*(5), 449–456. https://doi.org/10.1038/nm.3850

Cui, Q., Dai, L., Li, J., & Xue, J. (2020). *Accuracy of CEUS-guided sentinel lymph node biopsy in early-stage breast cancer : a study review and meta-analysis*. 1–11.

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., … Dwyer, M. A. (2021). Genetic/familial high-risk assessment: Breast, ovarian, and pancreatic, version 2.2021. *JNCCN Journal of the National Comprehensive Cancer Network*, *19*(1), 77–102. https://doi.org/10.6004/JNCCN.2021.0001

Decourtye-espiard, L., & Guilford, P. (2023). Hereditary Diffuse Gastric Cancer.

Gastroenterology, *164*(5), 719–735. https://doi.org/10.1053/j.gastro.2023.01.038

Desmedt, C., Zoppoli, G., Gundem, G., Pruneri, G., Larsimont, D., Fornili, M., Fumagalli, D., Brown, D., Rothé, F., Vincent, D., Kheddoumi, N., Rouas, G., Majjaj, S., Brohée, S., Van Loo, P., Maisonneuve, P., Salgado, R., Van Brussel, T., Lambrechts, D., … Sotiriou, C. (2016). Genomic Characterization of Primary Invasive Lobular Breast Cancer. *Journal of Clinical Oncology*, *34*(16), 1872–1880. https://doi.org/10.1200/JCO.2015.64.0334

Ditchi, Y., Broudin, C., El Dakdouki, Y., Muller, M., Lavaud, P., Caron, O., Lejri, D., Baynes, C., Mathieu, M. C., Salleron, J., & Benusiglio, P. R. (2019). Low risk of invasive lobular carcinoma of the breast in carriers of BRCA1 (hereditary breast and ovarian cancer) and TP53 (Li–Fraumeni syndrome) germline mutations. *Breast Journal*, *25*(1), 16–19. https://doi.org/10.1111/tbj.13154

Dominguez-Valentin, M., Therkildsen, C., Veerla, S., Jönsson, M., Bernstein, I., Borg, Å., & Nilbert, M. (2013). Distinct Gene Expression Signatures in Lynch Syndrome and Familial Colorectal Cancer Type X. *PLoS ONE*, *8*(8), 1–7. https://doi.org/10.1371/journal.pone.0071755

Dossus, L., & Benusiglio, P. R. (2015). Lobular breast cancer: Incidence and genetic and non-genetic risk factors. *Breast Cancer Research*, *17*(1), 1–8. https://doi.org/10.1186/s13058-015-0546-7

Drees, F., Pokutta, S., Yamada, S., Nelson, W. J., & Weis, W. I. (2005). *a -Catenin Is a Molecular Switch that Binds E-Cadherin- b -Catenin and Regulates Actin-Filament Assembly*. 903–915. https://doi.org/10.1016/j.cell.2005.09.021

Eheman, C. R., Shaw, K. M., Ryerson, A. B., Miller, J. W., Ajani, U. A., & White, M. C. (2009). The changing incidence of in situ and invasive ductal and lobular breast carcinomas: United States, 1999–2004. *Cancer Epidemiology Biomarkers and Prevention*, *18*(6), 1763–1769. https://doi.org/10.1158/1055-9965.EPI-08-1082

Eliyatkin, N., Yalcin, E., Zengel, B., Aktaş, S., & Vardar, E. (2015). Molecular Classification of Breast Carcinoma: From Traditional, Old-Fashioned Way to A New Age, and A New Way. *Journal of Breast Health*, *11*(2), 59–66. https://doi.org/10.5152/tjbh.2015.1669

Fairoosa, P., & Witharana, C. (2020). Gene Mutations in Hereditary Breast Cancer-

A Review. *European Journal of Medical and Health Sciences*, *2*(3). https://doi.org/10.24018/ejmed.2020.2.3.286

Farahat, T., Shaheen, H., Khalil, N., Bary, N., & Yassmin, H. (2017). Effect of Family History on Clinical and Pathological Characteristics of Breast Cancer: a Hospital-Based Study in Egypt. *The Egyptian Family Medicine Journal*, *1*(2), 1–14. https://doi.org/10.21608/efmj.2017.67591

Faria, G. R., Pinto-de-Sousa, J., Preto, J. R., Sousa, H. S., Barbosa, J. A., & Costa-Maia, J. (2013). Three decades of clinical-pathological trends in gastric cancer: Prospective data from a Portuguese hospital. *International Journal of Surgery*, *11*(6), 472–476. https://doi.org/10.1016/j.ijsu.2013.04.004

Forjaz de Lacerda, G., Kelly, S. P., Bastos, J., Castro, C., Mariotto, A. B., Anderson, W. F., Sciences, P., Registry, R. C., Centre, A. O., Registry, R. C., Registry, R. C., & Registry, R. C. (2018). Breast Cancer in Portugal: Temporal trends and age-specific incidence by geographic regions. *Cancer Epidemiology*, *54*, 12–18. https://doi.org/10.1016/j.canep.2018.03.003.Breast

Fox, S., Speirs, V., & Shaaban, A. M. (2022). Male breast cancer : an update. *Virchows Archiv*, 85–93. https://doi.org/10.1007/s00428-021-03190-7

Friedman, M., Adar, T., Patel, D., Lauwers, G. Y., Yoon, S. S., Mullen, J. T., & Chung, D. C. (2021). Surveillance Endoscopy in the Management of Hereditary Diffuse Gastric Cancer Syndrome. *Clinical Gastroenterology and Hepatology*, *19*(1), 189–191. https://doi.org/10.1016/j.cgh.2019.10.033

Garcia-Pelaez, J., Barbosa-Matos, R., São José, C., Sousa, S., Gullo, I., Hoogerbrugge, N., Carneiro, F., & Oliveira, C. (2022). Gastric cancer genetic predisposition and clinical presentations: Established heritable causes and potential candidate genes. *European Journal of Medical Genetics*, *65*(1). https://doi.org/10.1016/j.ejmg.2021.104401

Glessner, J. T., Li, J., Wang, D., March, M., Lima, L., Desai, A., Hadley, D., Kao, C., Gur, R. E., Cohen, N., Sleiman, P. M. A., Li, Q., Hakonarson, H., Wu, Z., Kim, C., Thomas, K., Favis, R., Fu, D. J., Chung, H., … Gopal, S. (2017). Copy number variation meta-analysis reveals a novel duplication at 9p24 associated with multiple neurodevelopmental disorders. *Genome Medicine*, *9*(1), 1–11. https://doi.org/10.1186/s13073-017-0494-1

Globocan. (2020). *Breast GLOBOCAN*. Breast Cancer GLOBOCAN 2020. chrome-

extension://efaidnbmnnnibpcajpcglclefindmkaj/https://gco.iarc.fr/today/data/factsheets/cancers/20-Breast-fact-sheet.pdf

Gong, G., Hannon, N., & Whittemore, A. S. (2010). Estimating gene penetrance from family data. *Genetic Epidemiology*, *34*(4), 373–381. https://doi.org/10.1002/gepi.20493

Graziano, F., Humar, B., & Guilford, P. (2003). The role of the E-cadherin gene (CDH1) in diffuse gastric cancer susceptibility : from the laboratory to clinical practice. *Annals of Oncology*, *14*(12), 1705–1713. https://doi.org/10.1093/annonc/mdg486

Green, B. L., Fasaye, G. A., Samaranayake, S. G., Duemler, A., Gamble, L. A., & Davis, J. L. (2022). Frequent cleft lip and palate in families with pathogenic germline CDH1 variants. *Frontiers in Genetics*, *13*(September), 1–5. https://doi.org/10.3389/fgene.2022.1012025

Guilford, P., Hopkins, J., & Harraway, J. (1998). E-cadherin germline mutations in familial gastric cancer. *Nature*, *69*(3), 203–208.

Gullo, I., Carneiro, F., Oliveira, C., & Almeida, G. M. (2018). *Heterogeneity in Gastric Cancer : From Pure Morphology to Molecular Classifications*. 50–63. https://doi.org/10.1159/000473881

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. do C., Amil Dias, J., Figueiredo, C., … Carneiro, F. (2018). Phenotypic heterogeneity of hereditary diffuse gastric cancer: report of a family with early-onset disease. *Gastrointestinal Endoscopy*, *87*(6), 1566–1575. https://doi.org/10.1016/j.gie.2018.02.008

Gullo, I., Grillo, F., Mastracci, L., Vanoli, A., Carneiro, F., Saragoni, L., Limarzi, F., Ferro, J., Parente, P., & Fassan, M. (2020). Precancerous lesions of the stomach, gastric cancer and hereditary gastric cancer syndromes. *Pathologica*, *112*(3), 166–185. https://doi.org/10.32074/1591-951X-166

Gullo, I., van der Post, R. S., & Carneiro, F. (2021). *Recent advances in the pathology of heritable gastric cancer syndromes*. 125–147. https://doi.org/10.1111/his.14228

Hakkaart, C., Ellison-Loschmann, L., Day, R., Sporle, A., Koea, J., Harawira, P., Cheng, S., Gray, M., Whaanga, T., Pearce, N., & Guilford, P. (2019). Germline CDH1 mutations are a significant contributor to the high frequency of early-onset diffuse gastric cancer cases in New Zealand Māori. *Familial Cancer*, *18*(1), 83–90. https://doi.org/10.1007/s10689-018-0080-8

Hansford, S., Kaurah, P., Li-chang, H., Woo, M., Senz, J., Zogopoulos, G., Santos, T. A., Claro, I., Carvalho, J., Nielsen, C., Padilla, S., Lum, A., Talhouk, A., Baker-lange, K., Fernandez, B., Keller, G., Lynch, H., Shah, S. P., Guilford, P., … Oliveira, C. (2015). *Hereditary Diffuse Gastric Cancer Syndrome*. *1*(1), 23–32. https://doi.org/10.1001/jamaoncol.2014.168

Harris, T. J. C., & Tepass, U. (2010). *Adherens junctions : from molecules to morphogenesis*. *11*(July). https://doi.org/10.1038/nrm2927

Henson, D. E., Dittus, C., Younes, M., Nguyen, H., & Albores-Saavedra, J. (2004). Differential trends in the intestinal diffuse types of gastric carcinoma in the United States, 1973–2000: Increase in the signet ring cell type. *Archives of Pathology and Laboratory Medicine*, *128*(7), 765–770. https://doi.org/10.5858/2004-128-765-dtitia

Hess, T., Maj, C., Gehlen, J., Borisov, O., Haas, S. L., Gockel, I., Vieth, M., Piessen, G., Alakus, H., Vashist, Y., Pereira, C., Knapp, M., Schüller, V., Quaas, A., Grabsch, H. I., Trautmann, J., Malecka-Wojciesko, E., Mokrowiecka, A., Speller, J., … Schumacher, J. (2023). Dissecting the genetic heterogeneity of gastric cancer. *EBioMedicine*, *92*, 1–13. https://doi.org/10.1016/j.ebiom.2023.104616

Hong, R., & Xu, B. (2022). Breast cancer: an up-to-date review and future perspectives. *Cancer Communications*, *42*(10), 913–936. https://doi.org/10.1002/cac2.12358

Hu, B., Hajj, N. El, Sittler, S., Lammert, N., Barnes, R., & Meloni-Ehrig, A. (2012). Gastric cancer: Classification, histology and application of molecular pathology. *Journal of Gastrointestinal Oncology*, *3*(3), 251–261. https://doi.org/10.3978/j.issn.2078-6891.2012.021

Ivanova, M., Porta, F. M., Giugliano, F., Frascarelli, C., Sajjadi, E., Venetis, K., Cursano, G., Mazzarol, G., Guerini-Rocco, E., Curigliano, G., Criscitiello, C., & Fusco, N. (2023). Breast Cancer with Brain Metastasis: Molecular Insights and

Clinical Management. *Genes*, *14*(6). https://doi.org/10.3390/genes14061160

Iyer, P., Moslim, M., Farma, J. M., & Denlinger, C. S. (2020). *Diffuse gastric cancer : histologic , molecular , and genetic basis of disease*. *2*, 20–22. https://doi.org/10.21037/tgh.2020.01.02

Jacobs, M. F., Dust, H., Koeppe, E., Wong, S., Mulholland, M., Choi, E. Y., Appelman, H., & Stoffel, E. M. (2019). Outcomes of Endoscopic Surveillance in Individuals With Genetic Predisposition to Hereditary Diffuse Gastric Cancer. *Gastroenterology*, *157*(1), 87–96. https://doi.org/10.1053/j.gastro.2019.03.047

Jarvandi, S., Pérez, M., Schootman, M., & Jeffe, D. B. (2016). Pre-Existing Diabetes in Early Stage Breast Cancer Patients is Associated with Lack of Improvement in Quality of Life 2 Years After Diagnosis. *International Journal of Behavioral Medicine*, *23*(6), 722–729. https://doi.org/10.1007/s12529-016-9577-4

Jensen, M., & Girirajan, S. (2017). Mapping a shared genetic basis for neurodevelopmental disorders. *Genome Medicine*, *9*(1), 9–11. https://doi.org/10.1186/s13073-017-0503-4

Jonker, M. A., Rijken, J. A., Hes, F. J., Putter, H., & Hensen, E. F. (2019). Estimating the penetrance of pathogenic gene variants in families with missing pedigree information. *Statistical Methods in Medical Research*, *28*(10–11), 2924–2936. https://doi.org/10.1177/0962280218791338

Kadys, A., Gremke, N., Schnetter, L., Kostev, K., & Kalder, M. (2023). Intercontinental comparison of women with breast cancer treated by oncologists in Europe, Asia, and Latin America: a retrospective study of 99,571 patients. *Journal of Cancer Research and Clinical Oncology*, *149*(10), 7319–7326. https://doi.org/10.1007/s00432-023-04681-7

Kakiuchi, M., Nishizawa, T., Ueda, H., Gotoh, K., Tanaka, A., Hayashi, A., Yamamoto, S., Tatsuno, K., Katoh, H., Watanabe, Y., Ichimura, T., Ushiku, T., Funahashi, S., Tateishi, K., Wada, I., Shimizu, N., Nomura, S., Koike, K., Seto, Y., … Ishikawa, S. (2014). letters Recurrent gain-of-function mutations of RHOA in diffuse-type gastric carcinoma. *Nature Publishing Group*, *46*(6). https://doi.org/10.1038/ng.2984

Kaurah, P., Gilpin, C., Nikkel, S., Connolly-wilson, M., Stroop, J., & Yim, D. (2007).

Founder and Recurrent CDH1 Mutations. *297*(21).

Kim, W., Kidambi, T., Lin, J., & Idos, G. (2022). Genetic Syndromes Associated with Gastric Cancer. *Gastrointestinal Endoscopy Clinics of North America*, *32*(1), 147–162. https://doi.org/10.1016/j.giec.2021.08.004

Kosztolányi, G. (2008). György Kosztolányi Risk assessment 3. RISK ASSESSMENT. *EJIFCC*, *19*, 13–21.

Kousi, M., & Katsanis, N. (2015). Genetic modifiers and oligogenic inheritance. *Cold Spring Harbor Perspectives in Medicine*, *5*(6), 1–22. https://doi.org/10.1101/cshperspect.a017145

Lakhani, S. R. (1999). The pathology of familial breast cancer. Morphological aspects. *Breast Cancer Research*, *1*(1), 31–35. https://doi.org/10.1186/bcr10

Lalloo, F., & Evans, D. G. (2012). Familial Breast Cancer. *Clinical Genetics*, *82*(2), 105–114. https://doi.org/10.1111/j.1399-0004.2012.01859.x

Laurén, 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 Pathologica et Microbiologica Scandinavica*, *64*, 31–49. https://doi.org/10.1111/apm.1965.64.1.31

Lauren, P., & Nevalainen, T. (1993). Epidemiology of intestinal and difuse types of gastric carcinoma: A time-trend study in Finland with comparison between studies from high- and low-risk areas. *Cancer Journal for Clinicians*, *73*(5), 1533–1533. https://doi.org/10.1002/1097-0142(19940301)73:5<1533::AID-CNCR2820730535>3.0.CO;2-J

Lee, C. Y. C., Olivier, A., Honing, J., Lydon, A. M., Richardson, S., O'Donovan, M., Tischkowitz, M., Fitzgerald, R. C., & di Pietro, M. (2023). Endoscopic surveillance with systematic random biopsy for the early diagnosis of hereditary diffuse gastric cancer: a prospective 16-year longitudinal cohort study. *The Lancet Oncology*, *24*(1), 107–116. https://doi.org/10.1016/S1470-2045(22)00700-8

Lee, K., Krempely, K., Roberts, M. E., Anderson, M. J., Carneiro, F., Chao, E., Dixon, K., Figueiredo, J., Ghosh, R., Kaurah, P., Kesserwan, C., Landrith, T., Li, S., & Arjen, R. (2018). *the Analysis of Germline CDH1 Sequence Variants*.

39(11), 1553–1568. https://doi.org/10.1002/humu.23650.Specifications

Lerner, B. A., & Llor, X. (2020). Genetic Gastric Cancer Risk Syndromes. *Current Treatment Options in Gastroenterology*, *18*(4), 604–615. https://doi.org/10.1007/s11938-020-00312-z

Li, Y., Liu, X., Chen, D., Liu, Y., Ma, Y., Liu, Q., Sun, W., & Li, J. (2021). Generation of an induced pluripotent stem cell line SDUBMSi005-A from a patient with double primary gastric and colon carcinoma. *Stem Cell Research*, *53*, 102253. https://doi.org/10.1016/j.scr.2021.102253

Li, Y., Zhikai, Z., Fuhai, M., Liyan, X., & Yantao, T. (2020). Gastric Signet Ring Cell Carcinoma : Current Management and Future Challenges. *Cancer Management and Research*, 7973–7981.

Libânio, D., Rodrigues, J. R., Bento, M. J., Ebigbo, A., Messman, H., Verhoeven, R. H. A., Van Damme, N., Bisschops, R., Spaander, M. C. W., & Dinis-Ribeiro, M. (2022). Gastric cancer incidence and mortality trends 2007–2016 in three European countries. *Endoscopy*, *54*(7), 644–652. https://doi.org/10.1055/a-1673-1118

Lim, H. J., Zhuang, L., & Fitzgerald, R. C. (2023). Current advances in understanding the molecular profile of hereditary diffuse gastric cancer and its clinical implications. *Journal of Experimental and Clinical Cancer Research*, *42*(1), 1–12. https://doi.org/10.1186/S13046-023-02622-3/TABLES/3

Linhares, M., Pinto, C. M., Libânio, D., Teixeira, M. R., Dinis-Ribeiro, M., & Brandão, C. (2023). Gastric Cancer: A Practical Review on Management of Individuals with Hereditary or Familial Risk for Gastric Cancer. *GE Portuguese Journal of Gastroenterology*, *30*(4), 253–266. https://doi.org/10.1159/000527208

Lobo, S., Benusiglio, P. R., Coulet, F., Boussemart, L., Golmard, L., Spier, I., Hüneburg, R., Aretz, S., Colas, C., & Oliveira, C. (2021). European Journal of Medical Genetics Cancer predisposition and germline CTNNA1 variants. *European Journal of Medical Genetics*, *64*(10), 104316. https://doi.org/10.1016/j.ejmg.2021.104316

López, M. J., Carbajal, J., Alfaro, A. L., Saravia, L. G., Zanabria, D., Araujo, J. M., Quispe, L., Zevallos, A., Buleje, J. L., Cho, C. E., Sarmiento, M., Pinto, J. A., &

Fajardo, W. (2023). Characteristics of gastric cancer around the world. *Critical Reviews in Oncology/Hematology*, *181*(October 2022). https://doi.org/10.1016/j.critrevonc.2022.103841

Łukasiewicz, S., Czeczelewski, M., Forma, A., Jacek, B., Robert, S., & Stanisławek, A. (2021). Breast Cancer—Epidemiology, Risk Factors, Classification, Prognostic Markers, and Current Treatment Strategies— An Updated Review. *Cancer Management and Research*, *13*(4287), 1–30.

Luveta, J., Parks, R. M., Heery, D. M., Cheung, K.-L., & Johnston, S. J. (2020). Invasive Lobular Breast Cancer as a Distinct Disease: Implications for Therapeutic Strategy. *Oncology and Therapy*, *8*(1), 1–11. https://doi.org/10.1007/s40487-019-00105-0

Lv, L., Liang, X., Wu, D., Wang, F., Zhang, Y., Cang, H., Deng, X., & Li, M. (2021). Is cardia cancer a special type of gastric cancer? A differential analysis of early cardia cancer and non-cardia cancer. *Journal of Cancer*, *12*(8), 2385–2394. https://doi.org/10.7150/JCA.51433

Machado, J. C., Oliveira, C., Carvalho, R., Soares, P., Berx, G., Caldas, C., & Seruca, R. (2001). *E-cadherin gene (CDH1) promoter methylation as the second hit in sporadic di □ use gastric carcinoma*. 7–10.

Majewski, I. J., Kluijt, I., Cats, A., Scerri, T. S., De Jong, D., Kluin, R. J. C., Hansford, S., Hogervorst, F. B. L., Bosma, A. J., Hofland, I., Winter, M., Huntsman, D., Jonkers, J., Bahlo, M., & Bernards, R. (2013). An αE-catenin (CTNNA1) mutation in hereditary diffuse gastric cancer. *Journal of Pathology*, *229*(4), 621–629. https://doi.org/10.1002/path.4152

Makki, J. (2015). Diversity of breast carcinoma: Histological subtypes and clinical relevance. *Clinical Medicine Insights: Pathology*, *8*(1), 23–31. https://doi.org/10.4137/CPath.s31563

Mathew, A., Rajagopal, P. S., Villgran, V., Sandhu, G. S., Jankowitz, R. C., Jacob, M., Rosenzweig, M., Oesterreich, S., & Brufsky, A. (2017). Distinct Pattern of Metastases in Patients with Invasive Lobular Carcinoma of the Breast. *Geburtshilfe Und Frauenheilkunde*, *77*(6), 660–666. https://doi.org/10.1055/s-0043-109374

McCart Reed, A. E., Kalinowski, L., Simpson, P. T., & Lakhani, S. R. (2021). Invasive lobular carcinoma of the breast: the increasing importance of this special

subtype. *Breast Cancer Research*, *23*(1), 1–16. https://doi.org/10.1186/s13058-020-01384-6

Melina Arnold, Jin Young Park, M. Constanza Camargo, Nuno Lunet, David Forman, I. S. (2020). Is gastric cancer becoming a rare disease? A global assessment of predicted incidence trends to 2035. *Gut*, *69*(5), 823–829. https://doi.org/10.1136/gutjnl-2019-320234.Is

Michaut, M., Chin, S. F., Majewski, I., Severson, T. M., Bismeijer, T., De Koning, L., Peeters, J. K., Schouten, P. C., Rueda, O. M., Bosma, A. J., Tarrant, F., Fan, Y., He, B., Xue, Z., Mittempergher, L., Kluin, R. J. C., Heijmans, J., Snel, M., Pereira, B., … Bernards, R. (2016). Integration of genomic, transcriptomic and proteomic data identifies two biologically distinct subtypes of invasive lobular breast cancer. *Scientific Reports*, *6*(January), 1–13. https://doi.org/10.1038/srep18517

Monster, J., Kemp, L., Martijn, G., & Van Der Post, R. (2022). BBA - Reviews on Cancer Diffuse gastric cancer : Emerging mechanisms of tumor initiation and progression. *BBA - Reviews on Cancer*, *1877*(3), 188719. https://doi.org/10.1016/j.bbcan.2022.188719

Monster, J. L., Kemp, L. J. S., Gloerich, M., & van der Post, R. S. (2022). Diffuse gastric cancer: Emerging mechanisms of tumor initiation and progression. *Biochimica et Biophysica Acta - Reviews on Cancer*, *1877*(3), 188719. https://doi.org/10.1016/j.bbcan.2022.188719

Muir, J., Aronson, M., Esplen, M. J., Pollett, A., & Swallow, C. J. (2016). Prophylactic Total Gastrectomy: a Prospective Cohort Study of Long-Term Impact on Quality of Life. *Journal of Gastrointestinal Surgery*, *20*(12), 1950–1958. https://doi.org/10.1007/s11605-016-3287-8

Nagtegaal, I. D., Odze, R. D., Klimstra, D., Paradis, V., Rugge, M., Schirmacher, P., Washington, K. M., Carneiro, F., & Cree, I. A. (2020). The 2019 WHO classification of tumours of the digestive system. *Histopathology*, *76*(2), 182–188. https://doi.org/10.1111/his.13975

Nanki, K., Toshimitsu, K., Takano, A., Fujii, M., Shimokawa, M., Ohta, Y., Matano, M., Seino, T., Nishikori, S., Ishikawa, K., Kawasaki, K., Togasaki, K., Takahashi, S., Sukawa, Y., Ishida, H., Sugimoto, S., Kawakubo, H., Kim, J., Kitagawa, Y., … Sato, T. (2018). Divergent Routes toward Wnt and R-spondin

Niche Independency during Human Gastric Carcinogenesis. *Cell*, *174*(4), 856–869.e17. https://doi.org/10.1016/j.cell.2018.07.027

Nascimento, R. G. do, & Otoni, K. M. (2020). Histological and molecular classification of breast cancer: what do we know? *Mastology*, *30*, 1–8. https://doi.org/10.29289/25945394202020200024

Neuhausen, S. L. (2000). Founder populations and their uses for breast cancer genetics. *Breast Cancer Research*, *2*(2), 77–81. https://doi.org/10.1186/bcr36

Obeagu, E., & Obeagu, G. (2023). Breast Cancer Breast Cancer. *Journal of the Royal Society of Medicine*, *70*(8), 515–517.

Obermair, F., Rammer, M., Burghofer, J., Malli, T., Schossig, A., Wimmer, K., Kranewitter, W., Mayrbaeurl, B., Duba, H. C., & Webersinke, G. (2019). Cleft lip/palate and hereditary diffuse gastric cancer: report of a family harboring a CDH1 c.687 + 1G > A germline mutation and review of the literature. *Familial Cancer*, *18*(2), 253–260. https://doi.org/10.1007/s10689-018-0111-5

Oliveira, C., Pinheiro, H., Figueiredo, J., Seruca, R., & Carneiro, F. (2015a). Familial gastric cancer : genetic susceptibility , pathology , and implications for management. *Lancet Oncology*, *16*(2), e60–e70. https://doi.org/10.1016/S1470-2045(14)71016-2

Oliveira, C., Pinheiro, H., Figueiredo, J., Seruca, R., & Carneiro, F. (2015b). Familial gastric cancer: Genetic susceptibility, pathology, and implications for management. *The Lancet Oncology*, *16*(2), e60–e70. https://doi.org/10.1016/S1470-2045(14)71016-2

Oliveira, C., Senz, J., Kaurah, P. ., & Pinheiro, H. (2009). *Germline CDH1 deletions in hereditary diffuse gastric cancer families*. *18*(9), 1545–1555. https://doi.org/10.1093/hmg/ddp046

Oliveira, C., Seruca, R., & Carneiro, F. (2009). Hereditary gastric cancer. *Best Practice and Research: Clinical Gastroenterology*, *23*(2), 147–157. https://doi.org/10.1016/j.bpg.2009.02.003

Oliveira, C., Seruca, R., Hoogerbrugge, N., Ligtenberg, M., & Carneiro, F. (2013). Clinical utility gene card for: Hereditary diffuse gastric cancer (HDGC).

European Journal of Human Genetics, *21*(8), 891. https://doi.org/10.1038/ejhg.2012.247

Ooki, A. (2022). *The dawn of precision medicine in diffuse-type gastric cancer*. *X*, 1–25. https://doi.org/10.1177/https

Orrantia-Borunda E., Anchondo-Nuñez P, Acuña-Aguilar LE, Gómez-Valles FO, & Ramírez-Valdespino CA. (2022). Chapter 3: Subtypes of Breast Cancer. In *Breast Cancer*.

Pareja, F., Ferrando, L., Lee, S. S. K., Beca, F., Selenica, P., Brown, D. N., Farmanbar, A., Da Cruz Paula, A., Vahdatinia, M., Zhang, H., Zoppoli, G., Wen, H. Y., Brogi, E., Robson, M. E., Razavi, P., Chandarlapaty, S., Weigelt, B., & Reis-Filho, J. S. (2020). The genomic landscape of metastatic histologic special types of invasive breast cancer. *Npj Breast Cancer*, *6*(1), 1–10. https://doi.org/10.1038/s41523-020-00195-4

Pharoah, P. D. P., 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*(6), 1348–1353. https://doi.org/10.1053/gast.2001.29611

Pitrez, P. R., Monteiro, L. M., Borgogno, O., Nissan, X., Mertens, J., & Ferreira, L. (2024). Cellular reprogramming as a tool to model human aging in a dish. *Nature Communications*, *15*(1). https://doi.org/10.1038/s41467-024-46004-5

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*(13), 1–11. https://doi.org/10.3390/cancers13133132

Rahit, K. M. T. H., & Tarailo-Graovac, M. (2020). Genetic Modifiers and Rare Mendelian Disease. *Genes*, *11*(3). https://doi.org/10.3390/genes11030239

Richard, F., Majjaj, S., Venet, D., Rothe, F., Pingitore, J., Boeckx, B., Marchio, C., Clatot, F., Bertucci, F., Mariani, O., Galant, C., van den Eynden, G., Salgado, R., Biganzoli, E., Lambrechts, D., Vincent-Salomon, A., Pruneri, G., Larsimont, D., Sotiriou, C., & Desmedt, C. (2020). Characterization of stromal tumor-infiltrating lymphocytes and genomic alterations in metastatic lobular breast

cancer. *Clinical Cancer Research*, *26*(23), 6254–6265. https://doi.org/10.1158/1078-0432.CCR-20-2268

Richards, S., Aziz, N., Bale, S., Bick, D., & Das, S. (2015). ACMG Standards and Guidelines 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. *Genetics in Medicine*, *17*(5), 405–424. https://doi.org/10.1038/gim.2015.30

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 Oncology*, *5*(9), 1325–1331. https://doi.org/10.1001/jamaoncol.2019.1208

Rodenburg, R. J. (2018). The functional genomics laboratory: functional validation of genetic variants. *Journal of Inherited Metabolic Disease*, *41*(3), 297–307. https://doi.org/10.1007/s10545-018-0146-7

São José, C., Garcia-Pelaez, J., Ferreira, M., Arrieta, O., André, A., Martins, N., Solís, S., Martínez-Benítez, B., Ordóñez-Sánchez, M. L., Rodríguez-Torres, M., Sommer, A. K., te Paske, I. B. A. W., Caldas, C., Tischkowitz, M., Tusié, M. T. T., Aretz, S., Capella, G., Castedo, S., de Voer, R. M., … Oliveira, C. (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*, *26*(5), 653–666. https://doi.org/10.1007/s10120-023-01395-0

Sarhangi, N., Hajjari, S., & Hedyari, S. F. (2022). Breast Cancer In the Era of Precision Medicine. *Molecular Biology Reports*, *49*, 10023–10037. https://doi.org/10.37723/jumdc.v14i1.823

Sawyer, E., Roylance, R., Petridis, C., Brook, M. N., Nowinski, S., Papouli, E., Fletcher, O., Pinder, S., Hanby, A., Kohut, K., Gorman, P., Caneppele, M., Peto, J., dos Santos Silva, I., Johnson, N., Swann, R., Dwek, M., Perkins, K. A., Gillett, C., … Garcia-Closas, M. (2014). Genetic Predisposition to In Situ and Invasive Lobular Carcinoma of the Breast. *PLoS Genetics*, *10*(4). https://doi.org/10.1371/journal.pgen.1004285

Scalzi, A. M. De, Bonanni, B., Galimberti, V., Pravettoni, G., & Corso, G. (2019). *E-*

cadherin germline mutations in M aori population. *28*, 1291–1294.

Scully, O. J., Bay, B., Yip, G., & Yu, Y. (2012). *Breast Cancer Metastasis*. *320*, 311–320.

Selvanathan, A., Nixon, C. Y., Zhu, Y., Scietti, L., Forneris, F., Murray, C., Buckley, M. F., Cox, T. C., & Roscioli, T. (2020). CDH1 Mutation Distribution and Type Suggests. *Genes*, *11*(391).

Sokolova, A., Johnstone, K. J., McCart Reed, A. E., Simpson, P. T., & Lakhani, S. R. (2023). Hereditary breast cancer: syndromes, tumour pathology and molecular testing. *Histopathology*, *82*(1), 70–82. https://doi.org/10.1111/his.14808

Song, M., Camargo, M. C., Weinstein, S. J., Best, A. F., Männistö, S., Albanes, D., & Rabkin, C. S. (2018). Family history of cancer in first-degree relatives and risk of gastric cancer and its precursors in a Western population. *Gastric Cancer*, *21*(5), 729–737. https://doi.org/10.1007/s10120-018-0807-0

Spargo, T. P., Opie-Martin, S., Bowles, H., Lewis, C. M., Iacoangeli, A., & Al-Chalabi, A. (2022). Calculating variant penetrance from family history of disease and average family size in population-scale data. *Genome Medicine*, *14*(1), 1–13. https://doi.org/10.1186/s13073-022-01142-7

Sugihara, H., Hattori, T., Fukuda, M., & Fujita, S. (1987). A Cell proliferation and differentiation in intramucosal and advanced signet ring cell carcinomas of the human stomach. *Virchows Archis Archi*.

Sung, H., Ferlay, J., Siegel, R. L., Laversanne, M., Soerjomataram, I., Jemal, A., & Bray, F. (2021a). *Global Cancer Statistics 2020 : GLOBOCAN Estimates of Incidence and Mortality Worldwide for 36 Cancers in 185 Countries*. *71*(3), 209–249. https://doi.org/10.3322/caac.21660

Sung, H., Ferlay, J., Siegel, R. L., Laversanne, M., Soerjomataram, I., Jemal, A., & Bray, F. (2021b). Global Cancer Statistics 2020: GLOBOCAN Estimates of Incidence and Mortality Worldwide for 36 Cancers in 185 Countries. *CA: A Cancer Journal for Clinicians*, *71*(3), 209–249. https://doi.org/10.3322/caac.21660

Takeichi, M. (2018). Multiple functions of a -catenin beyond cell adhesion regulation. *Current Opinion in Cell Biology*, *54*, 24–29.

https://doi.org/10.1016/j.ceb.2018.02.014

Tarighati, E., Keivan, H., & Mahani, H. (2023). A review of prognostic and predictive biomarkers in breast cancer. *Clinical and Experimental Medicine*, *23*(1), 1–16. https://doi.org/10.1007/s10238-021-00781-1

Thrift, A. P., Wenker, T. N., & El-Serag, H. B. (2023). Global burden of gastric cancer: epidemiological trends, risk factors, screening and prevention. *Nature Reviews Clinical Oncology*, *20*(5), 338–349. https://doi.org/10.1038/s41571-023-00747-0

Tsang, J. Y. S., & Tse, G. M. (2020). Molecular Classification of Breast Cancer. *Advances in Anatomic Pathology*, *27*(1), 27–35. https://doi.org/10.1097/PAP.0000000000000232

Turner, H., & Jackson, L. (2020). Evidence for penetrance in patients without a family history of disease: a systematic review. *European Journal of Human Genetics*, *28*(5), 539–550. https://doi.org/10.1038/s41431-019-0556-5

van Dieren, Jolanda, Bisseling, T. M., Liudmila, K., van der Post, C. R. S., & Corso, G. (2023). Endoscopic Surveillance and Pathology of Biopsies in CDH1, CTNNA1, and HDGC-Like Families. In *Hereditary Gastric and Breast Cancer Syndrome*. https://doi.org/10.1007/978-3-031-21317-5

van Velzen, M. J. M., Braemer, M., Nieuwenhuijzen, G. A. P., van Sandick, J. W., Siersema, P. D., Ruurda, J. P., Verheij, M., Spaander, M. C. W., Beerepoot, L. V., Haj Mohammad, N., van Laarhoven, H. W. M., & Verhoeven, R. H. A. (2023). Incidence, Stage, Treatment, and Survival of Noncardia Gastric Cancer. *JAMA Network Open*, *6*(8), e2330018. https://doi.org/10.1001/jamanetworkopen.2023.30018

Veen, E. M. Van, Evans, D. G., Harkness, E. F., Byers, H. J., Ellingford, J. M., Woodward, E. R., Bowers, N. L., Wallace, A. J., Howell, S. J., Howell, A., Lalloo, F., Newman, W. G., & Smith, M. J. (2022). Extended gene panel testing in lobular breast cancer. *Familial Cancer*, *21*(2), 129–136. https://doi.org/10.1007/s10689-021-00241-5

Wang, K., Yuen, S. T., Xu, J., Lee, S. P., Yan, H. H. N., Shi, S. T., Siu, H. C., Deng, S., Chu, K. M., Law, S., Chan, K. H., Chan, A. S. Y., Tsui, W. Y., Ho, S. L., Chan, A. K. W., Man, J. L. K., Foglizzo, V., Ng, M. K., Chan, A. S., … Leung, S. Y. (2014). Whole-genome sequencing and comprehensive molecular profiling identify

new driver mutations in gastric cancer. *Nature Publishing Group*, *46*(6). https://doi.org/10.1038/ng.2983

Weigelt, B., Geyer, F. C., & Reis-Filho, J. S. (2010). Histological types of breast cancer: How special are they? *Molecular Oncology*, *4*(3), 192–208. https://doi.org/10.1016/j.molonc.2010.04.004

Whiffin, N., Minikel, E., Walsh, R., O'Donnell-Luria, A. H., Karczewski, K., Ing, A. Y., Barton, P. J. R., Funke, B., Cook, S. A., Macarthur, D., & Ware, J. S. (2017). Using high-resolution variant frequencies to empower clinical genome interpretation. *Genetics in Medicine*, *19*(10), 1151–1158. https://doi.org/10.1038/gim.2017.26

WHO, & Organization, W. H. (n.d.). *WHO Breast Cancer 2021*. Retrieved January 26, 2023, from https://www.who.int/newsroom/fact-sheets/detail/breast-cancer

Wijnhoven, B. P. L., Dinjens, W. N. M., & Pignatelli, M. (2000). *E-cadherin ± catenin cell ± cell adhesion complex and human cancer*. 992–1005.

Worster, E., Liu, X., Richardson, S., Hardwick, R. H., Dwerryhouse, S., Caldas, C., & Fitzgerald, R. C. (2014). The impact of prophylactic total gastrectomy on health-related quality of life: A prospective cohort study. *Annals of Surgery*, *260*(1), 87–93. https://doi.org/10.1097/SLA.0000000000000446

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*. 838–843. https://doi.org/10.1136/jmedgenet-2019-105991

Yang, W. J., Zhao, H. P., Yu, Y., Wang, J. H., Guo, L., Liu, J. Y., Pu, J., & Lv, J. (2023). Updates on global epidemiology, risk and prognostic factors of gastric cancer. *World Journal of Gastroenterology*, *29*(16), 2452–2468. https://doi.org/10.3748/wjg.v29.i16.2452

Yao, Q., Qi, X., & Xie, S. H. (2020). Sex difference in the incidence of cardia and non-cardia gastric cancer in the United States, 1992–2014. *BMC Gastroenterology*, *20*(1), 1–7. https://doi.org/10.1186/s12876-020-01551-1

Zheng, H. C., Li, X. H., Hara, T., Masuda, S., Yang, X. H., Guan, Y. F., & Takano, Y.

(2008). Mixed-type gastric carcinomas exhibit more aggressive features and indicate the histogenesis of carcinomas. *Virchows Archiv*, *452*(5), 525–534. https://doi.org/10.1007/s00428-007-0572-7

Zhu, M., & Zhao, S. (2007). Candidate gene identification approach: Progress and challenges. *International Journal of Biological Sciences*, *3*(7), 420–427. https://doi.org/10.7150/ijbs.3.420

Chapter II| Rationale and Aims

Diffuse Gastric Lobular Breast Cancer Syndrome (DGLBCS) is a deadly cancer-syndrome, mainly associated with E-cadherin (*CDH1*) germline alterations. Despite *CDH1* pathogenic variants are frequently found in DGLBCS patients, a non-negligible proportion of carriers never develops clinical disease. The molecular mechanisms to explain why some *CDH1* carriers develop cancer very early, while others remain disease-free throughout their lives, are completely unknown. Therefore, it is impossible to predict if and when will a *CDH1* carrier be affected. However, it is becoming clear that disease penetrance greatly differs within families. Additionally, DGLBCS-like families share clinical presentations with DGLBCS, but lack molecular diagnosis, which prevents identifying at-risk individuals for disease prevention.

Herein, we hypothesized that: **I)** Genes causing other tumour risk syndromes or frequently mutated in sporadic DGCs are good candidates to explain DGLBCS-like families; **II)** Lifetime risk-assessment would be more realistic if performed using families carrying the same causative variant; **III)** Incomplete penetrance and variable disease spectrum are modulated by genetic modifiers.

As a reflex of the generalized lifetime risk estimations, current clinical guidelines offer the same surveillance and risk-reduction measures to all *CDH1* carriers (independently of their family history or causal variant), in a one-fits-all approach. Most importantly, prophylactic gastrectomy and mastectomy from healthy *CDH1* carriers is still the most efficient measure to prevent disease onset and is widely used, limiting the study of disease initiation. Therefore, human-derived cell models, like induced pluripotent stem cell lines (iPSCs), offer a way to replicate patient-specific variants, genetic background, and specific tissues for research, overcoming patients' lack of material.

In this project, we aimed to identify novel causes for DGLBC-like families, providing new genotype-phenotype correlations in *CDH1*-DGLBC, and lifetime-risk estimations for the largest cohort of *CDH1* carriers sharing a founder effect from Northern Portugal. Since in this context, both genetic and environmental factors are controlled, we stand at a privileged scenario for identifying genetic modifiers. We established a well characterized biobank of iPSCs and other biological materials

from DGLBCS families, analysed their genomes and transcriptomes to identify expression profiles of disease predisposition or disease protection. Recognizing specific lifetime risks and identifying disease modifiers, can have outstanding implications for the management of patients, by better defining the age for intensive surveillance and hopefully stratifying high-risk patients for prophylactic surgeries.

In this work, we addressed the following specific aims:

1) Identify DGLBCS candidate genes

- To perform a systematic review across DGLBCS-like families and FIGC families to pinpoint and identify candidate genes contributing to disease predisposition. |***Paper 1, Chapter III***
- To perform a systematic review on the somatic landscape of Diffuse Gastric Cancer (DGC) and its distinct subtypes |***Paper 2, Chapter III***

2) Characterize *CDH1* germline alterations and phenotypes associated with DGLBCS

- To demonstrate the functional pathogenic impact of the *CDH1* c.1901 C>T variant, and characterize the tumour spectrum and age of onset of several Portuguese families carrying this variant |***Paper 3, Chapter III***
- To perform a genotype-phenotype correlation study in the largest European cohort of *CDH1* carriers |***Paper 4, Chapter III***

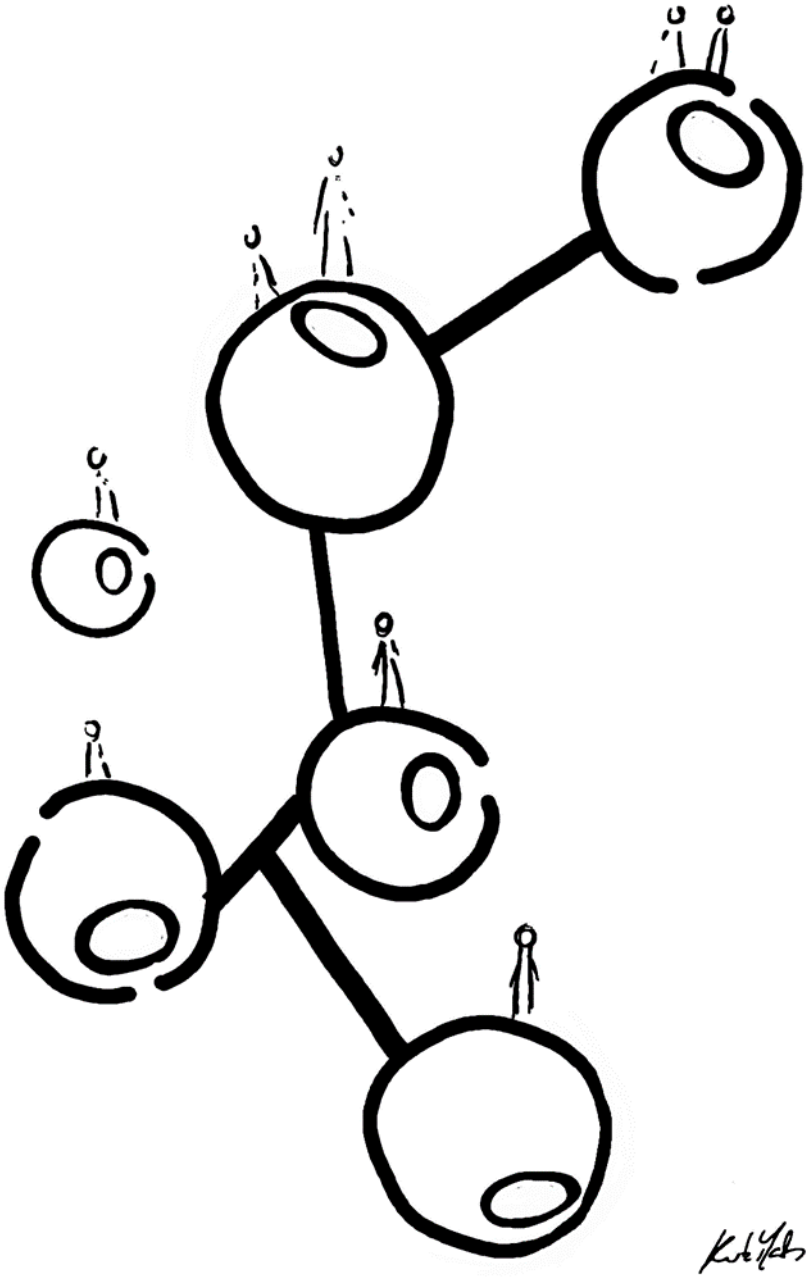
3) Calculate variant-specific lifetime risk in *CDH1* carriers and identify diagnostic signatures

- To perform a variant-specific penetrance study in the largest cohort with a *CDH1* founder effect |***Paper 5, Chapter III***
- To explore whether combinations of genomic and/or transcriptomic alterations allow the identification of carriers susceptible to disease development |***Paper 6, Chapter III***

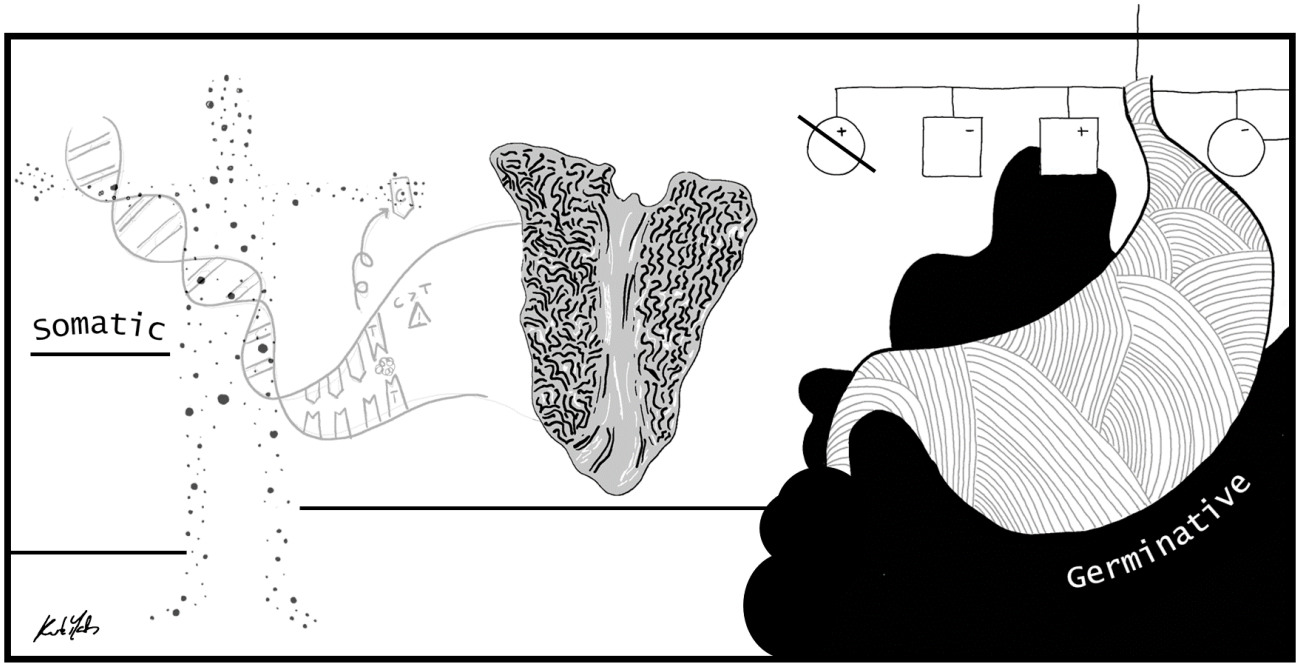
4) Establish a biobank for DGLBCS families

- To reprogram peripheral blood from Affected Carriers, Asymptomatic Carriers and Non-Carriers into iPSCs, and extract DNA and/or RNA from different sources (blood, peripheral blood mononuclear cells, formalin-fixed tissues).

Chapter III| Results



1. Search for DGLBCS candidate genes

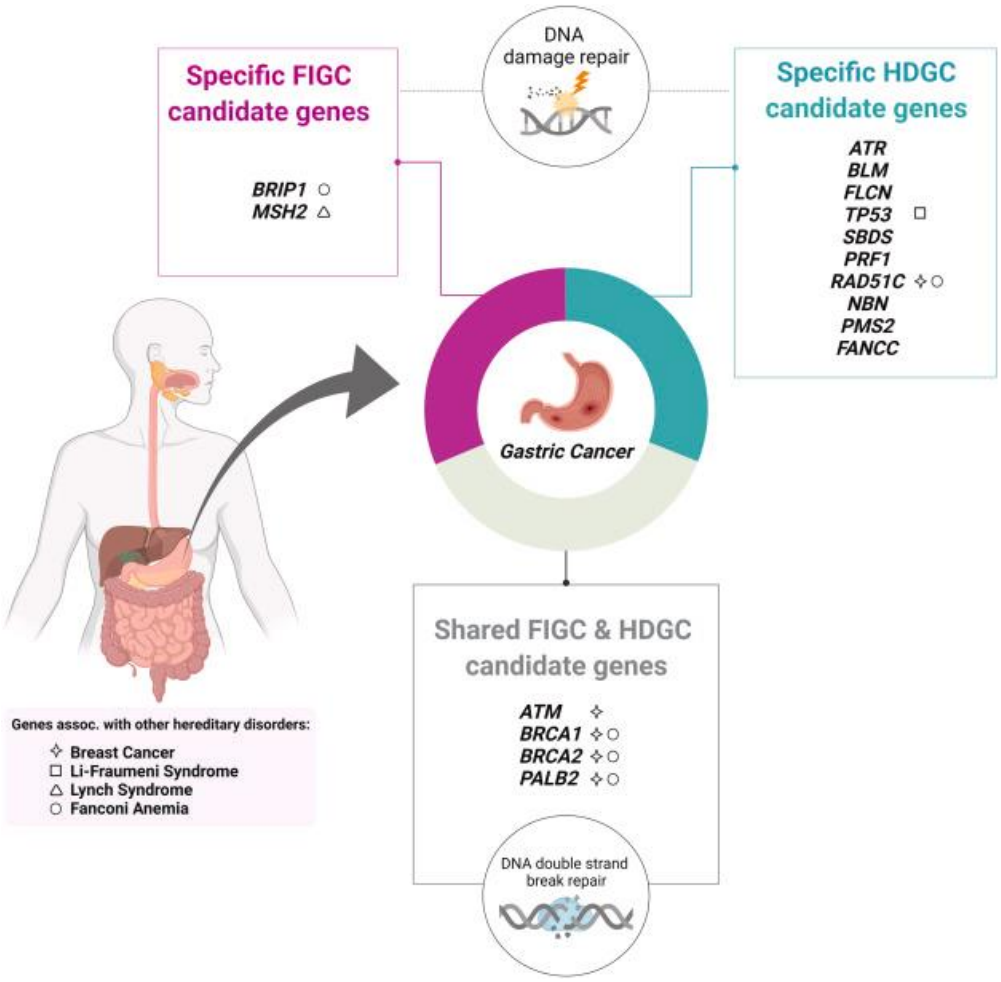


Paper 1: Gastric cancer genetic predisposition and clinical presentations: Established heritable causes and potential candidate genes.

Paper 2: Histological and mutational profile of diffuse gastric cancer: current knowledge and future challenges

1.1. Paper 1, Chapter III

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



Author statement:
Rita Barbosa-Matos was responsible for the conceptualization, investigation, methodology, image design, formal analysis, writing (original draft), review and editing.

This work provides a comprehensive overview of the established heritable causes and potential candidate genes associated with gastric cancer predisposition and clinical presentations. We explore the germline landscape of individuals with DGLBCS-like criteria, presenting phenotypes and family histories compatible with DGLBCS but lacking a *CDH1* or *CTNNA1* germline alteration. The list of DGC candidates genes highlights potential genetic factors in the predisposition and development of gastric cancer.

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

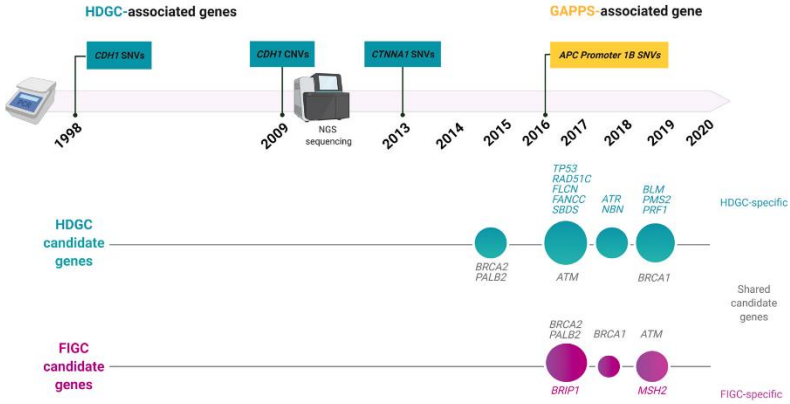


Fig. 1. Genetic timeline of *bona fide* and candidate genes for hereditary diffuse gastric cancer (HDGC), gastric adenocarcinoma and proximal polyposis of the stomach (GAPPS) and familial intestinal gastric cancer (FIGC). SNVs, single-nucleotide variants; CNVs, copy number variants.

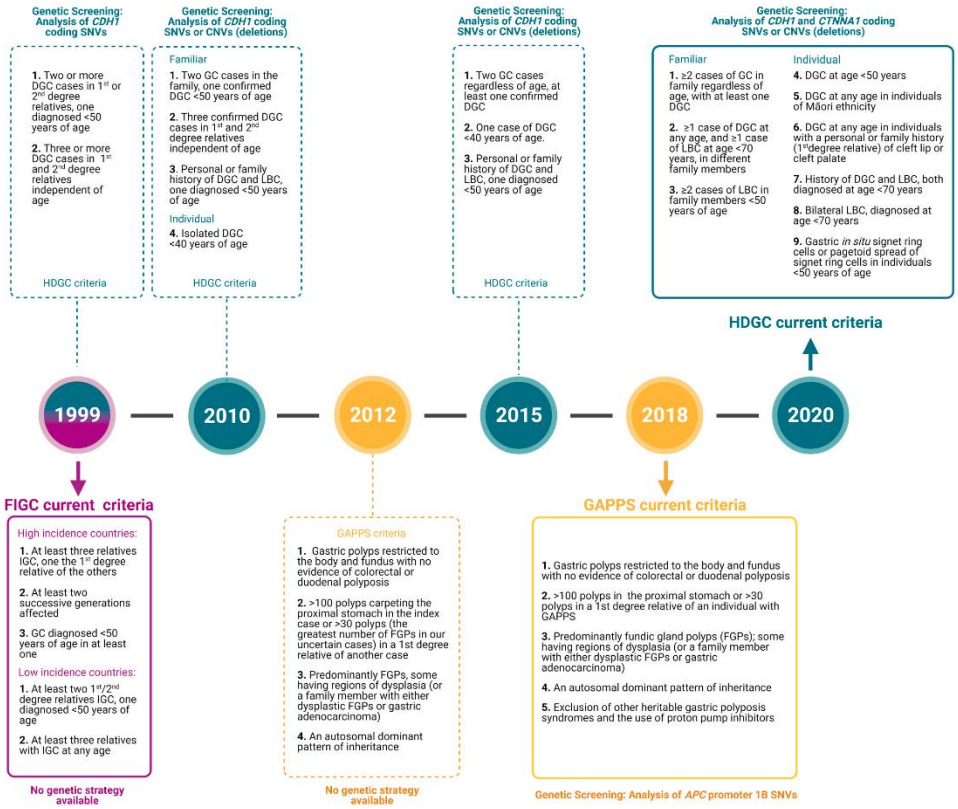


Fig. 2. Evolution of clinical criteria and genetic screening guidelines for the main GC-TRS: HDGC, GAPPS and FIGC. GC, gastric cancer; DGC, diffuse gastric cancer; LBC, lobular breast cancer; FGP(s), Fundic Gland Polyp(s); IGC, intestinal gastric cancer.

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

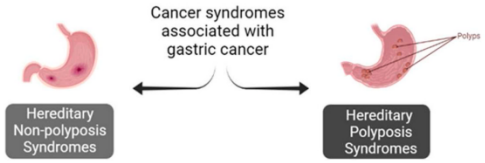
Knowledge on *CTNNA1*-associated HDGC is still very scarce, therefore most of what has been realized for *CDH1*-associated HDGC is yet to be discovered in *CTNNA1*-HDGC. *CTNNA1* carriers are predisposed to DGC, but whether LBC and cleft lip with or without cleft palate are also part of the disease spectrum 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 ≥ 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 criteria (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*)

Table 3 (*continued*)

Gene	mRNA position HGVS	*aa* position HGVS	Molecular type	Primary syndrome	Alternative syndromes	Spectrum	Probands information	Relatives information	ACMG criteria	ACMG classification	References (PMID)
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. Ǿ 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.

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 (Fig. 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 (%):

$$\frac{N^{o}ofHDGCfamilies\ bearing P/LP variants}{N^{o}offamilies\ screened\ fulfilling\ HDGC\ criteria}$$

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

$$\frac{N^{o}ofFIGCfamilies\ bearing P/LP variants}{N^{o}offamilies\ screened\ fulfilling\ FIGC\ criteria}$$

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 (Fig. 3, Table 3). *ATM*, *BRCA1*, *BRCA2,* and *PALB2* variants were present in both HDGC and FIGC (Fig. 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).

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; (Antoniou et al., 2014; Tischkowitz and Xia, 2010), and is considered a HBOC-associated gene (National Comprehensive Cancer Network, 2021a; Velázquez et al., 2019; Yang et al., 2016). *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 can 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., 2021). *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) (Huang et al., 2018; Lu et al., 2015). 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

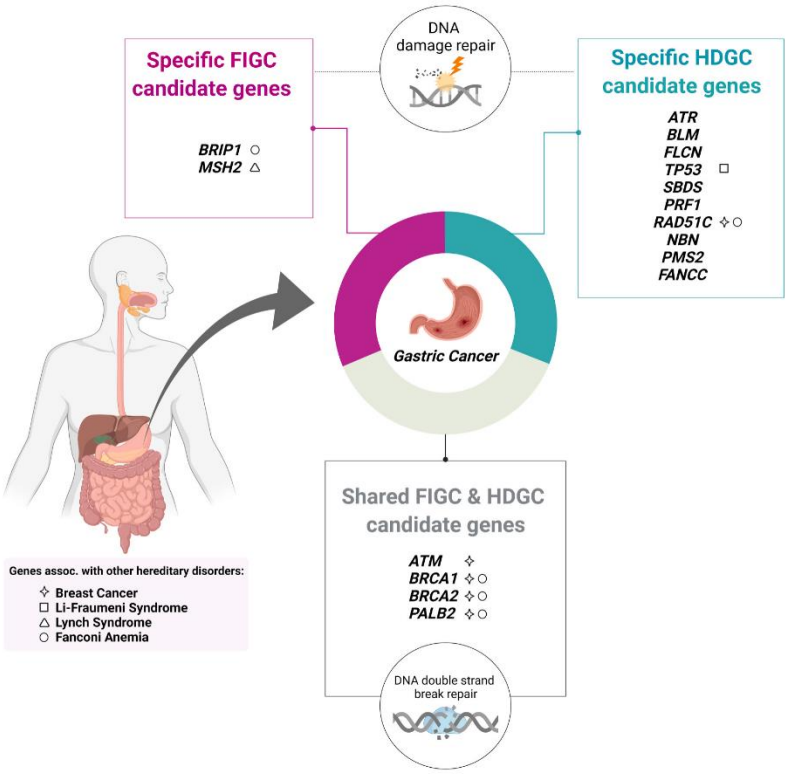


Fig. 3. List of candidate genes retrieved from our literature review and its associated biological processes.

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 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 (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 Nº 604370, 612555, 114480), and associated with autosomal recessive Fanconi Anaemia (OMIM Nº 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 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	Frecquency 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	
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	Frecquency 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 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	
		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 Garcia-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., Carneiro, F., Guilford, P., Blair, V., Chung, D.C., Norton, J., Ragunath, K., Van Krieken, J.H., Dwerryhouse, S., Caldas, C., 2010. Hereditary diffuse gastric cancer: updated consensus guidelines for clinical management and directions for future research. J. Med. Genet. 47, 436–444.

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., Stahlová, 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.

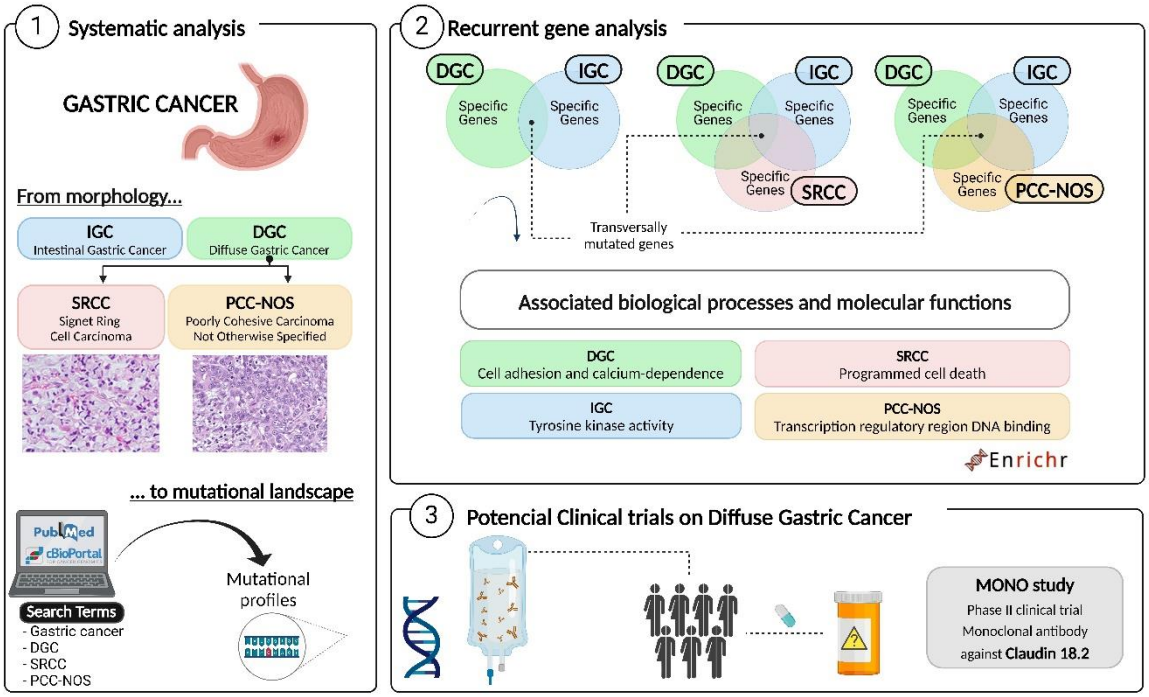
Setia, N., Wang, C.X., Lager, A., Maron, S., Shroff, S., Arndt, N., Peterson, B., Kupfer, S.S., Ma, C., Misdraji, J., Catenacci, D., Hart, J., 2020. Morphologic and molecular analysis of early-onset gastric cancer. Cancer 127, 103–114.
Shapiro, L., Fannon, A.M., Kwong, P.D., Thompson, A., Lehmann, M.S., 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.

Supplementary Material: Available through the link **:**

https://www.sciencedirect.com/science/article/pii/S1769721221002676?via%3Dihub#appsec1

1.2. Paper 2, Chapter III

Histological and mutational profile of diffuse gastric cancer: current knowledge and future challenges



Author statement:
Rita Barbosa-Matos was responsible for the investigation, methodology, conceptualization, formal analysis, writing (original draft), review and editing.

This publication emphasizes the distinction between diffuse (DGC) and intestinal gastric cancer (IGC), shedding light on their unique histological and molecular profiles. We explored the somatic mutational landscape of the DGC and its subtypes, providing valuable insights into its pathogenesis. We identified the most affected genes in DGC tumours, creating a specific mutational profile to analyse its molecular pathways, biological functions, and potential therapeutic targets.

Overall, the findings underscore the complexity of DGC and the necessity for a deeper understanding of its histological and mutational characteristics to advance personalized treatment strategies and improve patient outcomes.

Check for updates

Molecular Oncology

FEBS PRESS

SYSTEMATIC REVIEW

Histological and mutational profile of diffuse gastric cancer: current knowledge and future challenges

José Garcia-Pelaez[1,2,3], Rita Barbosa-Matos[1,2,4], Irene Gullo[1,2,5,6], Fátima Carneiro[1,2,5,6] and Carla Oliveira[1,2,5]

1 i3S – Instituto de Investigação e Inovação em Saúde da Universidade do Porto,Portugal
2 IPATIMUP – Institute of Molecular Pathology and Immunology, University of Porto, Portugal
3 Doctoral Programme on Biomedicine, Faculty of Medicine, University of Porto, Portugal
4 Doctoral Programme on Cellular and Molecular Biotechnology Applied to Health Sciences (BiotechHealth), ICBAS – Institute of Biomedical Sciences Abel Salazar, University of Porto, Portugal
5 Department of Pathology, FMUP - Faculty of Medicine of the University of Porto, Portugal
6 Department of Pathology, CHUSJ – Centro Hospitalar Universitário São João, Porto, Portugal

Keywords
diffuse gastric cancer; driver genes; intestinal gastric cancer; molecular classification; poorly cohesive carcinomas; signet-ring cell carcinoma

Correspondence
F. Carneiro, Ipatimup, Rua Júlio Amaral de Carvalho 45, 4200-135 Porto, Portugal
Fax: +351 225570799
Tel: +351 225570700
E-mail: fcarneiro@ipatimup.pt
C. Oliveira, i3S, Rua Alfredo Allen, 208, 4200-135 Porto, Portugal
Fax: +351 225570799
Tel: + 351 220408800
E-mail: carlaol@i3s.up.pt

José Garcia-Pelaez and Rita Barbosa-Matos contributed equally to the work as first authors
Fátima Carneiro and Carla Oliveira contributed equally to the work as senior authors

(Received 30 September 2020, revised 23 February 2021, accepted 12 March 2021, available online 2 May 2021)

doi:10.1002/1878-0261.12948

Gastric cancer (GC) pathogenesis is complex and heterogeneous, reflecting morphological, molecular and genetic diversity. Diffuse gastric cancer (DGC) and intestinal gastric cancer (IGC) are the major histological types. GC may be sporadic or hereditary; sporadic GC is related to environmental and genetic low-risk factors and hereditary GC is caused by inherited high-risk mutations, so far identified only for the diffuse histotype. DGC phenotypic heterogeneity challenges the current understanding of molecular mechanisms underlying carcinogenesis. The definition of a DGC-specific mutational profile remains controversial, possibly reflecting the heterogeneity of DGC-related histological subtypes [signet-ring cell carcinoma (SRCC) and poorly cohesive carcinoma not otherwise specified (PCC-NOS)]. Indeed, DGC and DGC-related subtypes may present specific mutational profiles underlying the particularly aggressive behaviour and dismal prognosis of DGC vs IGC and PCC-NOS vs SRCC. In this systematic review, we revised the histological presentations, molecular classifications and approved therapies for gastric cancer, with a focus on DGC. We then analysed results from the most relevant studies, reporting mutational analysis data specifying mutational frequencies, and their relationship with DGC and IGC histological types, and with specific DGC subtypes (SRCC and PCC-NOS). We aimed at identifying histology-associated mutational profiles with an emphasis in DGC and its subtypes (DGC vs IGC; sporadic vs hereditary DGC; and SRCC vs PCC-NOS). We further used these mutational profiles to identify the most commonly affected molecular pathways and biological functions, and explored the clinical trials directed specifically to patients with DGC. This systematic analysis is expected to expose a DGC-specific molecular profile and shed light into potential targets for therapeutic intervention, which are currently missing.

Abbreviations
DGC, diffuse gastric cancer; IGC, intestinal gastric cancer; PCC-NOS, poorly cohesive carcinoma not otherwise specified; SRCC, signet-ring cell carcinoma.

This is an open access article under the terms of the Creative Commons Attribution License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

1. Introduction – a general overview of gastric cancer morphology and mutational landscape

Gastric cancer (GC) is very heterogeneous from the morphological, genetic and molecular standpoints. The molecular heterogeneity is mirrored in the great variability of morphological phenotypes [1]. In this work, we revised the histological features, molecular classifications and approved therapies for GC, with a focus on DGC. We then explored a set of studies that reported mutational analysis data specifying mutational frequencies and their relationship with DGC and IGC histological types, as well as with specific subtypes of DGC, namely SRCC and poorly cohesive carcinoma not otherwise specified (PCC-NOS), to fill a gap in the current knowledge.

1.1. Main histopathological classifications of gastric cancer

A large number of histopathological classifications have been proposed over time [2–8]. The most commonly used gastric cancer classification is, in the Western countries, the one proposed by Laurén [3] and the World Health Organization (WHO) classification [8]. In Eastern countries, the most commonly used gastric cancer classification is the one issued by the Japanese Gastric Cancer Association (JGCA) [7], which is very similar to the WHO classification, 5th edition [9].

The Laurén classification [3] distinguishes two major types, intestinal gastric cancer (IGC) and diffuse gastric cancer (DGC). The former is constituted by tubular or papillary structures, while the latter is characterized by poorly cohesive and infiltrative tumour cells that may or may not have a signet-ring cell (SRC) morphology. Tumours presenting both intestinal and diffuse components are termed mixed carcinomas. Other carcinomas that do not fit in one of these subtypes are placed in the indeterminate category. Despite dating back to 1965, Laurén classification is still widely accepted and used, as it distinguishes subtypes with distinct epidemiologic settings, clinicopathologic profiles and biological behaviours. These subtypes also correspond to distinct tumour spreading patterns: while IGC tends to metastasize haematogenously to the liver, DGC usually disseminates through peritoneal surfaces. Mixed gastric cancer shows a poorer prognosis compared with IGC or DGC types [10,11] and a dual metastatic pattern (haematogenous metastases and peritoneal dissemination with lymph node metastases) [12], probably because of the cumulative adverse effect of the two components within a single tumour.

The classification of the World Health Organization (WHO), 5th edition (2019), [9] recognizes five major types of gastric adenocarcinoma (tubular, papillary, mucinous, poorly cohesive and mixed), as well as other rarer subtypes (gastric carcinoma with lymphoid stroma, hepatoid adenocarcinoma and related entities, micropapillary adenocarcinoma, gastric adenocarcinoma of fundic gland type, etc.).

1.2. The morphological heterogeneity of DGC

The DGC subtype in Laurén classification roughly corresponds to the poorly cohesive carcinomas (PCC) in the 2019 WHO classification [8].

DGC/PCC may show great morphological variability, between different tumours from distinct patients, but also within different areas of the same tumour (intertumour and intratumour morphological heterogeneity). The classical picture of DGC/PCC displays a tumour composed of poorly cohesive signet-ring cells (SRCs) (Fig. 1, left panel). SRCs are defined by the presence of an abundant mucin vacuole filling the cytoplasm and pushing the nucleus to the cell periphery [13]. Especially in early lesions (i.e., GCs limited to the mucosa and submucosa), different types of SRCs may be observed, which are distinguished by the nature of the cytoplasmic mucin [14,15]. The description of different SRC types may be considered merely

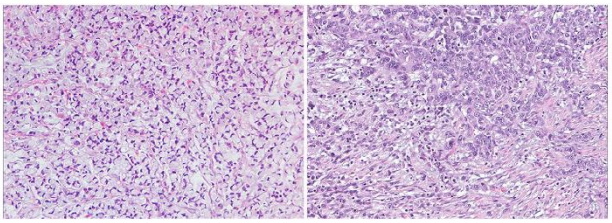


Fig. 1. Representative Images of a SRCC case (left) and a PCC-NOS case (right) highlighting the heterogeneity present in sporadic DGC.

18780261, 2021, 11, Downloaded from https://febs.onlinelibrary.wiley.com/doi/10.1002/1878-0261.12948 by Cochrane Portugal, Wiley Online Library on [31/01/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

academic, as it does not appear to influence the tumour prognosis. DGC/PCC may also show cells that do not present classic SRC morphology. In particular, advanced lesions infiltrating the gastric wall may show pleomorphic, bizarre and diffusely infiltrative neoplastic cells, sometimes with a lymphohistiocytic or sarcomatoid appearance, the latter representing the surrogate of epithelial-to-mesenchymal transition activation. The recognition of this subtype is of utmost importance, as it presents a poorer prognosis when compared to pure SRC carcinomas (SRCCs) [16]. Consistent with this observation, the new 2019 WHO classification stresses the importance of evaluating the morphological heterogeneity of DGC and distinguishing different PCC subtypes, based on the presence and quantity of SRCs [8]. Poorly cohesive carcinomas of the SRC type (SRCC) are composed predominantly or exclusively of SRCs and show a better prognosis, while non-SRC type, that is PCC-NOS, is composed (or show a predominant component) of poorly cohesive and cells, without a classic SRC morphology (Fig. 1, right panel).

1.3. Main molecular classifications of gastric cancer

The molecular and genomic analysis of large GC series has identified a large variety of oncogene activating mutations, tumour suppressor gene (TSG)-inactivating alterations, gene fusions, somatic copy number alterations (sCNAs) and other structural variations, as well as epigenetic and transcriptional changes [17]. Activating mutations in oncogenes such as *KRAS*, *BRAF* and *PIK3CA* have been reported in GC through single-gene analysis at least three decades ago [18–21] and remain important in GC using genome-wide strategies [22–24]. *TP53* is the most frequently mutated TSG in GC, often occurring in more than 50% of the cases, regardless of the experimental approach used, either target gene or genome-wide [22,25–27]. Somatic inactivating mutations in *CDH1* gene (a relevant TSG in GC) are known since the early 90's and remain a hallmark of the diffuse histological type [22,28–34].

Genome-wide strategies allowed to overcome the limitations of single-gene targeted approaches. Through next-generation sequencing (NGS) technologies, the knowledge on GC mutational burden and molecular signatures increased which culminated in the development of GC molecular classifications. These classifications based on genomics, transcriptomics and/or epigenomic profiles can be associated with the main GC histotypes (Table 1).

The first GC molecular classification was based exclusively on gene expression profiling and directly addressed differences in GC main histological types [35]. The second, although driven by gene expression profiling, was supported by gene-specific mutational analysis, chromosomal instability and DNA methylation analysis [36]. These studies also presented patients' response to treatment and disease outcomes according to molecular classification. In 2014, a game-changer GC molecular classification was published by The Cancer Genome Atlas (TCGA) consortium [22], based on a deep and integrative molecular profiling with array-based somatic copy number analysis, whole-exome sequencing, array-based DNA methylation profiling, messenger RNA sequencing (mRNA), microRNA (miRNA) sequencing and reverse-phase protein array (RPPA), microsatellite instability (MSI) testing and low-pass whole-genome sequencing. In 2015, the Asian Cancer Research Group (ACRG) generated comprehensive genomic and transcriptomic data sets from 300 GC cases, mainly from East Asia, producing the first GC classification based on integrated molecular data analysis and clinical outcomes [37]. In the two latter publications, there was an attempt to correlate the molecular subtypes with GC main histotypes, which did not provide a straightforward overlap (Table 1). Nevertheless, there is some correspondence between the two latter classifications, namely regarding tumours bearing mismatch repair deficiency (MSI), and those dominated by E-cadherin deficiency, genomic stability and epithelial-to-mesenchymal transition features (GS for TCGA and MSS/EMT for ACRG). The greater differences between TCGA and ACRG classifications relate to the lack of an EBV subgroup in the ACRG classification, which is molecularly defined in the TCGA study, and the heterogeneity of the CIN subgroup in TCGA, which becomes divided into two distinct subgroups in the ACRG study (Table 1).

1.4. Cost-effective strategies for molecular classification

Given the complexity and high cost of the experimental approaches needed to stratify GC according to TCGA and ACRG classifications, several studies proposed less costly approaches to reach similar endpoints (Fig. 2) [38–48].

In 2016, Setia *et al.* used 14 biomarkers in a GC cohort from the United States and Ahn *et al.* studied protein or mRNA expression of MLH1, E-cadherin, p53 and EBV in a South Korean population to derive five GC molecular subgroups and reproducing a combination of TCGA and ACRG classifications [38,39]. In 2016, *in situ* hybridization (ISH) or immunohistochemistry (IHC) were used to study MLH1, p53 and

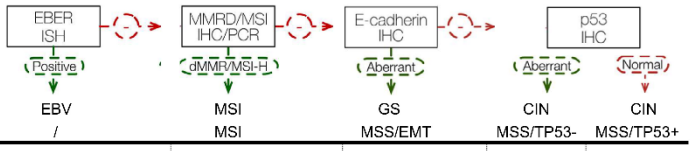
18780261, 2021, 11, Downloaded from https://febs.onlinelibrary.wiley.com/doi/10.1002/1878-0261.12948 by Cochrane Portugal, Wiley Online Library on [31/01/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

Table 1. Overview of gastric cancer molecular classifications and relationship with main Laurén histotypes. ACRG, Asian Cancer Research Group; CIMP, CpG island methylation phenotype; CIN, chromosomal instability; CNA, copy number alteration; CSC, cancer stem cell; EBV, Epstein–Barr virus; EMT, epithelial-to-mesenchymal transition; GC, gastric cancer; GS, genomically stable; MSI, microsatellite instability; MSS, microsatellite stable; TCGA, The Cancer Genome Atlas; TKR, tyrosine kinase receptors. Lauren histological subtypes are highlighted in green (diffuse GC) and light blue (intestinal GC).

Reference (sample size)	Methodology	Main features of molecular subtypes and relationship with Laurén histopathological classification				
[35] (n = 270)	Gene expression profiling			**G-DIF** (44%)	**G-INT** (56%)	
				Cell proliferation Fatty acid metabolism Diffuse GC	Cell adhesion Carbohydrate and protein metabolism Intestinal GC	
[36] (n = 248)	Gene expression profiling, *TP53* mutation, CNA and ;DNA methylation analysis			**MESENCHYMAL**	**PROLIFERATIVE**	**METABOLIC**
				Low *TP53* mutations Low E-cadherin mRNA CSC/EMT proprieties Diffuse GC	High *TP53* mutations Genomic instability Oncogene amplification Intestinal GC	Low *TP53* mutations Normal gastric mucosa gene expression Intestinal GC (gastric phenotype)
[22] (n = 295) – TCGA cohort	Array-based somatic copy number analysis, whole-exome seq, array-based DNA methylation profiling, mRNA seq, microRNA seq and reverse-phase protein array, MSI testing.	**EBV** (9%) EBV-CIMP *CDKN2A* silencing *PIK3A* mutations *PD-L1/2* amplification *JAK2* amplification	**MSI** (22%) Gastric-CIMP *MLH1* silencing *PIK3A* mutations *HER2/3* mutations *EGFR* mutations Intestinal GC	**GS** (20%) *CDH1* mutations *RHOA* mutations *CLDN18-ARHGAP* fusion (RhoA-GTPase) Diffuse GC	**CIN** (50%) High *TP53* mutations *TKR-RAS* amplification Amplification of cell-cycle mediators Intestinal GC	
[37] (n = 251) – ACRG cohort	Whole-genome seq, gene expression profiling, copy number analysis, targeted re-sequencing.	EBV+ cases included in MSS/TP53+	**MSI** (23%) *MLH1* loss Hypermutation (KRAS, ARID1A, PIK3A) Intestinal GC	**MSS/EMT** (15%) *CDH1* loss Diffuse GC	**MSS/TP53-** (36%) High *TP53* mutations Genomic instability Oncogene amplification Intestinal GC	**MSS/TP53+** (26%) Intestinal GC

Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/1878-0261.12945 by Cochrane Portugal, Wiley Online Library on [11/01/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

Algorithm strategy and molecular groups.



References (sample size)	%	%	%	%	%
[38] (n = 146)	4.5	16.4	20.5	48.6	9.6
[39] (n = 349)	7.4	6.9	15.2	49.0	21.5
[40] (n = 104)	6.7	16.3	/	38.5	38.5
[41] (n = 206)	/	23.5	6.0	17.0	53.5
[42] (n = 244)	9.1 (intestinal GC) 0.0 (diffuse GC)	10.2 (intestinal GC) 0.0 (diffuse GC)	6.0 (intestinal GC) 0.0 (diffuse GC)	55.4 (IGC) 19.2 (DGC)	28.0 (IGC) 80.8 (DGC)
[43] (n = 894)	8.8	8.6	13.2	28.1	41.3
[44] (n = 70)	2.9	7.2	11.4	61.4	10.0
[46] (n = 91)	13.1	12.1	16.5	17.8	38.5
[47] (n = 329)	12.2	5.2	32.8	49.5[1]	
[45] (n = 65)	9.2 (TCGA) /	23.1(TCGA) 21.5 (ACRG)	21.5 (TCGA) 20.0 (ACRG)	46.1 (TCGA) 43.1 (ACRG)	/ 15.4 (ACRG)
[48] (n = 256)	/	22.0	29.0	28.2[2]	20.8[3]

Fig. 2. Cost-effective strategies for molecular classification. [A] TCGA, The Cancer Genome Atlas [22]; [B] ACRG, Asian Cancer Research Group [37]; CIN, chromosomal instability; EBER ISH, EBV-encoded small RNA *in situ* hybridization; EBV, Epstein–Barr virus; EMT, epithelial-to-mesenchymal transition; GC, gastric cancer; GS, genomically stable; IHC, immunohistochemistry; MMRD, mismatch repair deficiency; MSI, microsatellite instability; MSS, microsatellite stability; PCR, polymerase chain reaction. [1]CIN subtype defined by the presence of aneuploidy, using DNA flow cytometry; [2]MSS/TP53- subtype defined by low p21 protein expression; [3]MSS/TP53+ subtype defined by high p21 protein expression.

EBV, but excluded the analysis of MSS/EMT (ACRG) or GS (TCGA) [40]. In 2018, GC patients were classified using IHC based on molecular biomarkers to derive four molecular subgroups, which excluded EBV, and supported this classification demonstrating a significant and independent correlation between IHC subtype and overall survival, clinical outcome, SRC phenotype, tumour grade and vessel invasion [41].

In parallel, in 2018, ISH and IHC were performed in a cohort mainly composed of IGC to refine previous GC classifications [42]. In 2019, 894 consecutive GC Korean patients were used to validate previous classifications, with IHC- and PCR-based approaches and confirming the association between molecular subtypes and patients' overall survival [43]. More recently, in 2020, an immunophenotypic classification, based on IHC and ISH, was used to evaluate MLH1, p53, HER2, E-cadherin and EBV in whole sections from surgical GC specimens [44]. Another similar approach was used to apply the IHC/ISH molecular algorithm to a cohort of GC patients from Chile [46]. In this study, the authors used targeted NGS sequencing to characterize the four molecular subtypes, identifying *FGFR2* and *KRAS* gene amplification as potential actionable targets in the EMT-like subgroup [46]. Wang *et al.* [45] studied a GC Chinese cohort to conclude that it would be possible to reproduce the TCGA and ACRG molecular classifications only by using IHC. In the two most recent cost-effective strategies for GC molecular classification, Zhao *et al.* and Tsai *et al.* used alternative surrogate biomarkers to define MSS/TP53 subgroups and CIN subtype, based on p21 protein expression and DNA content assessed by DNA flow cytometry, respectively [47,48].

1.5. Clinical outcomes, molecular features, prognosis, and approved and promising therapies for GC molecular subgroups

The research effort that followed the publication of the TCGA and ACRG molecular classifications led not only to the development of cost-effective strategies

18780261, 2021, 11, Downloaded from https://febs.onlinelibrary.wiley.com/doi/10.1002/1878-0261.12948 by Cochrane Portugal, Wiley Online Library on [31/01/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

to stratify GC patients according to these classifications (Fig. 2), but also to validate the molecular components from different molecular subgroups. This further allowed correlating molecular subgroups or alterations with clinical outcomes and response to therapy (Table 2). In Table 2, the possible correlations between TCGA and ACRG subgroups are presented, supported by shared molecular features. This information allows rethinking the application of former approved therapies and novel promising ones.

One of the most evident gaps in GC therapy is the lack of targeted therapies for DGC patients, which mainly reside in the TCGA GS group and in the MSS/EMT ACRG group (Table 2). This group of patients, which represents nearly 20% of all GC and those with the poorest prognosis, is currently treated with conventional chemotherapy but responds poorly [49] (Table 2). Immune checkpoint inhibitors did not prove as an alternative in these patients, as MSS/EMT cases from the ACRG group also showed poor treatment response [37]. Promising alternative therapies such as anti-CLDN18.2 antibody, targeting the *CLDN18-ARHGAP26* gene fusion, which is enriched in DGC cases [22,50,51] and intraperitoneal administration of anticancer agents [52], are just emerging treatments for this molecular subgroup, but results are still pending (Table 2).

The remaining GC molecular subgroups are less problematic regarding prognosis and available approved therapies or emerging ones (Table 2). The subgroups EBV-positive from TCGA and MSI from TCGA and ACRG show the best prognosis and are more likely to benefit from immune checkpoints inhibitors. In fact, MSI tumours are already selected for anti-PD1 treatment [53–55]. Patients bearing EBV-positive tumours are part of the ACRG MSS/TP53+ subgroup, present intermediate prognosis and are also often PD-L1-positive, turning them also into good candidates for immune checkpoint inhibition. Indeed, a phase II study on pembrolizumab showed the ORR to be not only 85.7% in MSI-H gastric cancers, but also more important to be 100% in EBV-related GC [56]. The TCGA CIN and ACRG MSS/TP53- subgroups present intermediate prognosis and the best survival benefit from adjuvant chemotherapy [37]. Given the high frequency of oncogene amplifications, tumours from this subgroup may also be treated with the approved anti-HER2 therapy [57], and promising treatments include inhibitors of the RTK-RAS pathway, DNA repair molecules and ATR (Table 2).

Diffuse gastric cancer has long been recognized as an independent GC histopathological entity; however, along the time, DGC-associated subtypes have been recognized and deeply characterized at the morphological and molecular levels. Despite this recognition, the most two comprehensive GC molecular classifications (TCGA and ACRG) have been blind to these subtypes, and DGC as a whole, mainly cluster in the TCGA GS group and in the MSS/EMT ACRG group. These two molecular subgroups, besides sharing a dominant GC histotype, also share mutations in DGC-associated genes. Despite the increasing knowledge about GC histopathological and molecular subtypes, the integration of these two aspects and its clinical utility remains poorly explored.

In summary, DGC and its subtypes remain a major problem in terms of prognosis, response to conventional chemotherapy and a still arid ground for novel therapeutic interventions.

1.6. Aim

After recognizing a knowledge gap regarding molecular features potentially useful for clinical outcome prediction and therapy design in DGC, our aim was to identify histology-associated mutational profiles for this GC histotype and its subtypes. We further used these mutational profiles to identify their most affected molecular pathways and biological functions, and explored the clinical trials specifically designed for DGC patients. This systematic analysis is expected to expose a DGC-specific molecular profile and shed light into potential targets for therapeutic intervention, which are currently missing.

2. Methodology

We collected available literature from 1998 to 2020, in order to get an overall view of the morphological and molecular features of sporadic GC and its molecular classifications. Peer-reviewed published articles were searched with the help of 'PubMed' database using the following keywords in different combinations: 'mutational landscape, mutational signature, gastric cancer, diffuse-type gastric cancer, poorly cohesive carcinoma and signet-ring cell carcinoma'. Only peer-reviewed published articles written in English language, irrespective of their study design and type (case reports, case series, original research articles, reviews, systematic reviews, meta-analyses, cohort studies, clinical trials), were consulted, and in the context of the present study, we selected only articles with relevant data and best fitting our aim.

2.1. Data collection for recurrent gene analysis

For the recurrent gene analysis, published studies or public cohorts reporting mutational frequencies and

18780261, 2021, 11, Downloaded from https://febs.onlinelibrary.wiley.com/doi/10.1002/1878-0261.12948 by Cochrane Portugal, Wiley Online Library on [31/01/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

Table 2. Representative genomic subtypes, common molecular features, clinical outcomes and corresponding therapeutic approaches. NA, not available; The Cancer Genome Atlas (TCGA); Asian Cancer Research Group (ACRG); Genomically stable (GS); Microsatellite Stable/Epithelial-Mesenchymal transition (MSS/EMT); Microsatellite instable (MSI); Helicobacter pylori (H. pylori); Vascular endothelial growth factor receptor 2 tyrosine kinase (VEGFR2); Programmed-cell-death-1-receptor (PD-1); Programmed death-ligand 1 (PD-L1); Receptor Tyrosine Kinase (RTK); Human epidermal growth factor receptor 2 (HER2); Epidermal growth factor receptor(EGFR); Fibroblast Growth Factor Receptor 2 (FGFR2); Poly (ADP-ribose) polymerase (PARP); ataxia telangiectasia and Rad-3 related protein kinase (ATR); References: [A] [22]; [B] [37] [C] [49] [D] [83]; [E] [84]; [F] [85]; [G] [86]; [H] [87] (e-updated in 2019); [I] [52]; [J] [88]; [K] [89]; [L] [53]; [M] [90]; [N] [91]; [O] [54]; [O] [92]; [P] [57]; [O] [93]; [R] [56].

TCGA molecular classification [A]			ACRG molecular classification [B]			Shared features and therapeutic approach			
Subtype frequency main location predominant histology	Clinical outcome/ prognosis	Molecular features	Subtype frequency main location predominant histology	Clinical outcome/ prognosis	Molecular features	Shared molecular features	Approved therapeutic strategies	Promising targets/ strategies
GS 20% NA DGC	Diagnosed at early age. Worst overall survival. Less benefit from adjuvant chemotherapy [C]. Poor response to immune checkpoint inhibitors [D][E][F].	High frequency of *CDH1* (37%) and *RHOA* (15%) mutations. *CLDN18-ARHGAP* fusion. Elevated expression of angiogenesis-related pathways [A].	MSS/EMT 15% NA DGC	Poor prognosis and high recurrence frequency with tendency to develop at earlier age. Majority of cases diagnosed at stages II/III/V [B].	Loss of E-cadherin expression. Low mutation rate [B].	GS and MSS/EMT *CDH1* and *RHOA* mutations	Standard chemotherapy: combinations of fluoropyrimidines and platinum derivatives [G][H].	Cell junctions: CLDN18.2 antibody (phase III clinical trials: NCT03504397, NCT03653507); intraperitoneal administration of anticancer agents (phase III clinical trial: UMIN000005930 [I]; (phase II clinical trial NCT04034251); anti-VEGFR2 treatment using ramucirumab and apatinib [J]; [K] and other angiogenesis inhibitors (phase III clinical trial NCT02773524) [L][M]
MSI 22% antrum IGC	Diagnosed at older ages; associated with *H. pylori* infection and intestinal metaplasia [A]. Better prognosis	CpG island methylator phenotype (*MLH1* silencing): hypermutation phenotype frequently targeting genes such as *TP53*, *KRAS*,	MSI 23% antrum IGC	Good prognosis [B].	*MLH1* gene silencing: commonly mutated genes: *ARID1A*, *MTOR*, *KRAS*, *PIK3CA*, *ALK*, *PTEN*.	MSI Hypermutated and hypermethylated tumours. Both enriched in IGCs.	Immune checkpoint inhibitors: Anti-PD-1 [L][O][P][R].	Immune checkpoint inhibitors: PD-1/PD-L1 pathway: (phase III clinical trials: NCT02872116, NCT02746796, NCT03019588.

This article is protected by copyright. All rights reserved

Table 2. (Continued).

TCGA molecular classification [A]			ACRG molecular classification [B]			Shared features and therapeutic approach		
Subtype frequency main location predominant histology	Clinical outcome/ prognosis	Molecular features	Subtype frequency main location predominant histology	Clinical outcome/ prognosis	Molecular features	Shared molecular features	Approved therapeutic strategies	Promising targets/ strategies
	than the MSS subtype [C]	*ARID1A*, *PIK3CA*, *ERBB3*, *PTEN*, *CHRD* and *HLA-B* with nucleotide substitutions; and RNF43, B2 M and NF1 with insertions/ deletions [A][N].			Overexpression of PD-L1; T-cell infiltration [B].			NCT03615326, NCT03189719, NCT02625610, NCT04152889) [L][M].
EBV +			MSS/TP53+			EBV+ and MSS/ TP53+		
9% Gastric fundus or body Gastric cancer with lymphoid stroma	Good prognosis [C].	EBV infection; CpG island methylation; *CCDKN2A* silencing; *JAK2*, *CD274*, *PDCD1LG2*, *ERBB2*, *PD-L1 and PD-L2* amplification; mutations in *PIK3CA* (80%), *ARID1A* (55%) and *BCOR* (23%) [A].	26% NA IGC	Intermediate prognosis [B].	Mutations in *ARID1A*, *PIK3CA*, *SMAD4*, *APC* [B].	EBV infection; *PIK3CA* mutations	Standard chemotherapy: combinations of fluoropyrimidines and platinum derivatives [G][H].	Immune checkpoint inhibitors: PD-1/PD-L1 pathway: (phase III clinical trials NCT02872116, NCT02746796, NCT03019588, NCT03615326, NCT03189719, NCT02625610, NCT04152889) and (phase II clinical trials NCT03755440, NCT03257163 and NCT04202601) [L][M] [P][Q][R].

18780261, 2021, 11, Downloaded from https://febs.onlinelibrary.wiley.com/doi/10.1002/1878-0261.12948 by Cochrane Portugal, Wiley Online Library on [11/01/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

Table 2. (Continued).

TCGA molecular classification [A]			ACRG molecular classification [B]			Shared features and therapeutic approach		
Subtype frequency main location predominant histology	Clinical outcome/ prognosis	Molecular features	Subtype frequency main location predominant histology	Clinical outcome/ prognosis	Molecular features	Shared molecular features	Approved therapeutic strategies	Promising targets/ strategies
CIN			MSS/TP53-			CIN and MSS/ TP53-		
50% Gastro-oesophageal junction/cardia IGC	Poorer overall survival than EBV subtype but better than GS. Benefit from adjuvant chemotherapy [C]. Poor response to immune checkpoint inhibitors [E][F].	RTK-RAS pathway genes amplification namely *HER2*, *VEGFA* and *FGFR2*, *TP53* (71%) mutations [A] [O].	36% NA IGC	Intermediate prognosis and recurrence rates compared with the other subtypes. Survival benefit with adjuvant chemotherapy [B].	Genomic instability; *TP53* mutations; recurrent focal amplification of RTKs and cell cycle modulators; *ERBB2* (HER2) amplification[B].	High frequency of *TP53* mutations. Alterations in RTK-RAS pathway genes. Both enriched in IGCs	Standard chemotherapy: Combinations of fluoropyrimidines and platinum derivatives [G][H]. Trastuzumab for HER2-positive GCs [L][P].	RTK-RAS pathway: EGFR, HER2 and HER4 inhibition (phase I/II/II clinical trial NCT03130790); EGFR overexpression (phase III clinical trial NCT01813253); FGFR2 amplification (phase III clinical trial NCT0369452 and phase II clinical trial NCT03694522); anti-VEGFR2 treatment using ramucirumab and apatinib [J][K], and other angiogenesis inhibitors (phase III clinical trial NCT02773524); DNA repair: PARP inhibitors (phase II clinical trials NCT03427814, NCT02678182) TP53 mutants: ATR inhibitor (phase II clinical trial NCT03641313) [L][M].

18780261, 2021, 11, Downloaded from https://febs.onlinelibrary.wiley.com/doi/10.1002/1878-0261.12948 by Cochrane Portugal, Wiley Online Library on [31/01/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

their relationship with GC histological types and subtypes of interest (IGC, DGC, SRCC and PCC-NOS) were searched for, selected upon compliance with the following criteria and refined by manual curation:

1 Studies from CBIOPORTAL (https://www.cbioportal.org/) encompassing 'gastric/ stomach adenocarcinoma' cancer NGS data;
2 'PubMed' deposited peer-reviewed articles, published between 2014 and 2020, reporting on data produced by targeted sequencing and/or whole-exome sequencing and/or whole-genome sequencing, and according to different MESH term combinations:
   a (mutational landscape OR mutational signature) AND (diffuse-type gastric carcinoma OR diffuse-type gastric cancer OR gastric adenocarcinoma OR gastric cancer OR poorly cohesive carcinoma) NOT review.
   b (genomic variation) AND (diffuse-type gastric cancer) NOT review.
   c (target sequencing OR whole-exome sequencing OR whole-genome sequencing) AND (diffuse-type gastric cancer) NOT review.

Manual curation was further performed on all publications selected with the above criteria, aiming at finding those presenting histological classification for each tumour, either according to Laurén [3] for comparison between DGC and IGC, or specifically mentioning comparisons between DGC subtypes, and not restricted to Laurén classification, namely if mentioning SRCC and PCC-NOS. Moreover, to be selected for the analysis, each publication should also specify either the set of mutated genes per case, or the gene mutation frequency among histological types or subtypes, as well as the total number of cases analysed within the materials (main or supplementary).

Whenever gene mutation frequencies per histological type or subtypes were not provided, their calculation was performed based on the formula:

$$\text{Mutated frequency of gene X} = \frac{\text{N}^{\text{o}}\text{ of cases of a specific histological subtype mutated for gene X}}{\text{Total N}^{\text{o}}\text{ of cases of a specific histological subtype}}$$

According to criteria 1, seven different entries were identified; from these, only three data sets were used due to the availability of Laurén histological classification and specific gene mutation frequencies. According to criteria 2a, there were 162 results from which five publications were selected after manual curation. Criteria 2b retrieved 11 results, from which two had already been found with previous criteria and eight did not comply with our criteria after manual curation; therefore, only one publication was selected. Criteria 2c retrieved 29 results, from which four had already been found with previous criteria and 24 did not comply with our criteria after manual curation. Accordingly, a single new publication was selected based on criteria 2c. Overall, 3 data sets and 7 original articles were selected (Fig. 3).

All remaining articles not fulfilling the above criteria were excluded.

2.2. Criteria for recurrent gene analysis

To identify relevant mutated genes, we established two selection steps: gene recurrence and gene classification.

2.2.1. Gene Recurrence

Only genes found to be mutated in at least three studies/resources in a frequency higher than 5% were selected for further analysis (Fig. 3, Analysis I). Due to the limited number of studies available for the analysis of SRCC and PCC-NOS, all genes with an available mutation frequency were used for this part of the study (Fig. 3, Analysis II).

2.2.2. Gene Classification

Mutated genes, for each histological type or subtype, were allocated to three classes: specific mutated genes, transversally mutated genes and enriched mutated genes.

Specific mutated genes were those exclusive of a single histological type or subtype.

Transversally mutated genes were those for which the difference between histological types or subtypes was, on average, lower than twofold. If the mutation frequency of a gene was reported in one study only, hampering average calculation, the absolute mutation frequency was used.

Enriched mutated genes were those for which difference between histological types or subtypes was, on average, higher than twofold. If the mutation frequency of a gene was reported in one study only, hampering average calculation, the absolute mutation frequency was used.

18780261, 2021, 11, Downloaded from https://febs.onlinelibrary.wiley.com/doi/10.1002/1878-0261.12948 by Cochrane Portugal, Wiley Online Library on [31/01/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

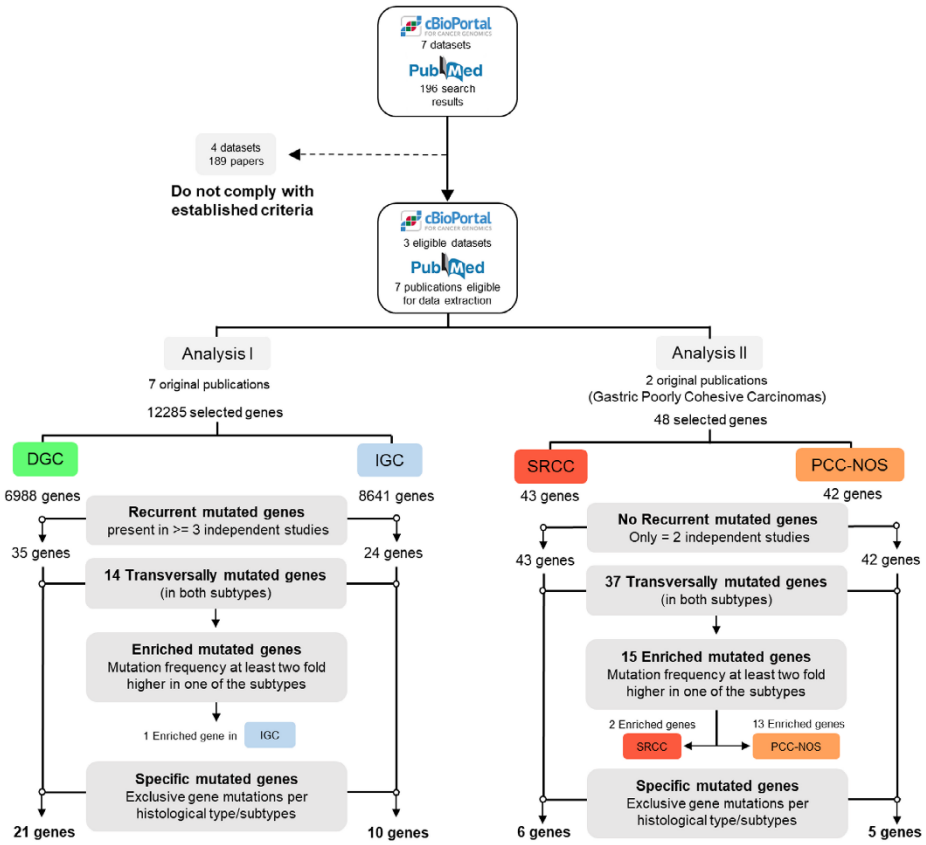


Fig. 3. Data collection and research strategy for recurrent gene analysis.

2.3. Identification of biological processes and molecular function associated with recurrent gene analysis

Specific mutated genes, transversally mutated genes and enriched mutated genes were used to feed Jvenn [58], an interactive diagram viewer, and allowed identifying genes overlapping between distinct GC histological types and subtypes through Venn Diagrams.

Recurrently mutated genes in DGC and IGC were compared to define the specific mutated genes for each histological type. The remaining DGC-associated genes were considered transversally DGC- and IGC-mutated genes, as no enrichment was seen in any of them. For the SRCC vs PCC-NOS analysis, specific mutated genes per subtype were obtained from a single report.

Further, the output of the Venn diagrams was used to analyse Gene Ontology (GO) terms, biological processes and molecular functions associated with each gene set using Enrichr [59].

2.4. Search for clinical trials directed to DGC

Relevant clinical trials directed specifically to DGC patients were selected by searching the website www.clinicaltrials.gov (/clinicaltrials.gov/), and using as

18780261, 2021, 11, Downloaded from https://febs.onlinelibrary.wiley.com/doi/10.1002/1878-0261.12948 by Cochrane Portugal, Wiley Online Library on [31/01/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

query terms 'diffuse gastric cancer', which retrieved 24 trials; 'signet-ring cell carcinoma' with 19 trials; and 'gastric poorly cohesive carcinoma' with no results. After identifying only those associated with GC, 12 studies were selected. Only two clinical trials (NCT01285557 – DIGEST; NCT03977220 – NORDICA) were specifically directed to DGC patients, and one was directed to patients bearing SRCC (NCT01717924 – ADCI002 Study). Beyond these, and given the promising results recently obtained using anti-claudin 18.2 antibodies, we manually added the MONO study where almost 50% of patients enrolled had DGC (NCT01197885 – MONO) [60]. In total, 13 clinical trials were selected since they were directed to DGC patients.

For these, a deeper analysis on therapeutic strategy, candidate drugs and molecular targets was performed.

2.5. Study limitations

There are several limitations in this study, namely the discrepancies between mutational profiles and frequencies among cohorts, which may be due to several factors: the methodological approach applied by different research groups (sequencing methods and pipeline analyses); the heterogeneity of GC (intra- and inter-tumoral); and the sample size of the cohorts under analysis (specifically in the studies focused on SRCC and PCC-NOS). Another issue is the lack of diversity in the geographical origin of the cohorts used in the selected studies for analysis, which reflects data available in the literature. However, the reproducibility of findings across the various molecular classifications (e.g. TCGA and ACRG) indicates that relevant GC-associated molecular defects likely override population differences. Our analysis was restricted to mutational data, leaving behind transcriptomics and epigenetics, which can provide valuable data. Copy number variation analysis would have been a strength in this study; however, there are very few publications listing CNVs per case or their relative frequency to allow an accurate analysis. In addition, recent data suggested that copy number gains tended to occur more commonly in IGC than in DGC that was the focus of this work [61]. In addition, the comparison of sporadic and hereditary DGC suffered from scarce data published so far on somatic profiles in HDGC. Notwithstanding, we believe that the information herein provided constitutes a good basis for further studies aiming at a better understanding of the molecular pathogenesis of DGC and its clinical implications regarding prognosis and therapy.

3. Results

Strategies for gene mutation detection evolved from single-gene analysis to the analysis of multigene panels, whole exomes or even whole genomes. This created the opportunity to derive multigene profiles that are valuable to correlate molecular defects with specific histological subtypes.

3.1. DGC vs IGC: distinctive mutational profiles

The data compilation performed herein and depicted in Fig. 4 shows the mutational landscape across DGC cohorts retrieved from nine independent studies, mainly Asian, and obtained with different technical approaches (four targeted sequencing; four WES, two WGS) (Data S1, Table S1). We next retrieved data from six independent studies of IGC, mostly also of Asian origin, having a similar degree of detailed information, depicted in Fig. 5. The mutational landscape of IGC obtained with different technological approaches (three targeted sequencing; three WES, two WGS) (Data S1, Table S1).

To derive the mutational landscape from DGC, we performed a recurrent gene analysis, as described in the methodology section that provided three mutated gene sets, both for DGC and for IGC: specific mutated genes, transversally mutated genes and enriched mutated genes.

3.2. DGC-specific mutated genes

This analysis confirmed *CDH1* genetic variants as the hallmark of the DGC histological type [31,33,34], as *CDH1* is the only gene found to be mutated in all DGC data sets and in none of the IGC data sets (specific DGC-mutated gene) (Figs 4 and 5). *CDH1* variants were found on average in 25% of DGC cases, ranging from 12% to 38% among studies (Fig. 4). Besides *CDH1*, another 20 genes were found to be mutated specifically in DGC, in at least three studies (Fig. 4). *TTN* gene mutations were found on average in 40% of DGC cases, and described in three studies, making this gene the top-ranking specific DGC-mutated gene, while *RHOA* and *CTNNB1* were found to be mutated in seven and five studies, on average in 10% and 7% of the cases, respectively. All other specific DGC-mutated genes were found in three or four studies (Fig. 4). Seven of these were mutated on average in > 12% of DGC. *FAT3* (15%) and *KMT2C* (12%) have a previously reported association with DGC; *CSMD1* (21%), *OBSCN* (15%), *RYR2* (13%) and *TAF1* (12%) have

18780261, 2021, 11, Downloaded from https://febs.onlinelibrary.wiley.com/doi/10.1002/1878-0261.12948 by Cochrane Portugal, Wiley Online Library on [31/01/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

Information source		TCGA Firehouse Legacy	CBIPORTAL OncoSG, 2018	PMID: 29622765	PMID: 31171626	PMID: 24816253	PMID: 24816255	PMID: 28873240	PMID: 25583476	PMID: 31164161	Average mutation frequency
Origin of samples		**Several population**	**China**	**Korea**	**USA, China, Japan**	**Hong Kong**	**Tokyo**	**Korea**	**China**	**China**	
Cohort Specification		DGC	DGC	DGC	DGC	DGC	DGC	Gastric Poorly Cohesive Carcinoma	DGC	DGC	
Number of samples		*n* = 72	*n* = 34	*n* = 23	*n* = 21	*n* = 29	*n* = 87	*n* = 91	*n* = 155	*n* = 50	
Genes		**Frequency of mutated samples per study**									
Specific DGC-mutated genes	*TTN*	39%	41%	39%	–	–	–	–	–	–	**40%**
	CSMD1	29%	21%	13%	–	–	–	–	–	–	21%
	OBSCN	23%	9%	13%	–	–	–	–	–	–	15%
	CDH1	21%	12%	30%	38%	28%	33%	20%	12%	30%	**25%**
	RYR2	16%	9%	13%	–	–	–	–	–	–	13%
	KMT2C	14%	6%	–	29%	–	1%	–	–	–	12%
	MUC6	14%	6%	–	–	7%	–	–	–	–	9%
	FAT3	11%	15%	18%	–	–	–	–	–	–	15%
	APOB	10%	9%	–	10%	–	–	–	–	–	9%
	CHL1	10%	9%	9%	–	–	–	–	–	–	9%
	SETBP1	10%	9%	–	5%	–	–	–	–	–	8%
	DOCK2	10%	6%	18%	–	–	–	–	–	–	11%
	THSD7B	7%	15%	–	–	14%	–	–	–	–	12%
	RHOA	7%	12%	9%	0%	10%	25%	17%	3%	–	10%
	CACNA1E	7%	12%	13%	–	–	–	–	–	–	11%
	ABCA13	7%	9%	13%	–	–	–	–	–	–	10%
	TAF1	7%	9%	–	19%	–	–	–	–	–	12%
	NCOR2	7%	6%	13%	–	–	–	–	–	–	9%
	TGFBR2	7%	6%	9%	–	–	–	–	–	–	7%
	CTNNB1	6%	6%	–	–	–	1%	22%	1%	–	7%
	ERBB3	6%	6%	–	–	–	–	11%	2%	–	6%
Transversally DGC- and IGC-mutated genes	*TP53*	37%	41%	18%	57%	45%	34%	41%	–	54%	**41%**
	FAT4	23%	15%	–	–	–	5%		–	–	14%
	ARID1A	23%	9%	–	–	10%	15%	20%	–	–	15%
	SPTA1	21%	3%	–	14%	7%	–		–	–	11%
	PIK3CA	19%	–	–	–	–	5%	22%	10%	–	14%
	GLI3	14%	9%	–	–	14%	–		–	–	12%
	LRP2	9%	9%	–	–	–	–		–	6%	8%
	RIMS2	9%	9%	23%	–	10%	–		–	–	13%
	DCLK1	9%	6%	–	–	10%	–		–	–	8%
	APC	7%	3%	–	–	–	2%	28%	6%	0%	8%
	KRAS	6%	–	–	10%	3%	1%	12%	3%	–	6%
	SMAD4	4%	6%	–	14%	10%	–	13%	3%	10%	9%
	ATM	4%	6%	–	10%	–	–	12%	–	14%	9%
	PTEN	4%	3%	–	19%	0%	3%	17%	–	10%	8%

Fig. 4. Mutational landscape of DGC subtype. Nine sources were used: two CBIOPORTAL data sets and seven references. The origin, number of samples, reported study identification and other cohorts' characteristics are indicated in the heading of the table. The mutated genes of each subtype were divided into two main classifications: transversally mutated and specific genes. The specific mutated genes for DGC were associated with the green colour and are listed in the upper part of the table. In the lower part of the table, the transversally mutated genes in both DGC and IGC are presented. Mutation frequency was represented in a colour scale. Green was appointed to genes with mutation frequencies between 0% and 9% (less frequently mutated). Yellow was used for mutation frequencies between 10% and 19%. Light orange to mutation frequencies between 20% and 39%. A variation between dark orange and red was associated with mutation frequencies increasing from 40% to a maximum of 57% (frequently mutated). The average mutation frequency calculated for each DGC recurrent gene is present in the right column, and a similar colour scale was applied to distinguish the genes with higher mutation frequencies.

18780261, 2021, 11, Downloaded from https://febs.onlinelibrary.wiley.com/doi/10.1002/1878-0261.12948 by Cochrane Portugal, Wiley Online Library on [11/01/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

Information source		CBIOPORTAL, UHK, 2011	CBIPORTAL OncoSG, 2018	PMID: 24816253	PMID: 24816255	PMID: 31164161	PMID: 25583476	Average mutation frequency
Origin of samples		Hong Kong	China	Hong Kong	Tokyo	China	China	
Cohort Specification		IGC	IGC	IGC	IGC	IGC	IGC	
Number of samples		*n* = 18	*n* = 75	*n* = 41	*n* = 51	*n* = 47	*n* = 139	
Genes		Frequency of mutated samples per study						
Specific IGC-mutated genes	*BRCA2*	17%	7%	–	–	4%	9%	9%
	ERBB2	11%	12%	12%	–	–	6%	10%
	AKAP9	11%	8%	–	–	26%	–	15%
	DLC1	11%	9%	–	8%	–	–	9%
	MACF1	11%	9%	–	8%	–	–	9%
	RNF43	11%	11%	7%	–	–	–	10%
	NRG1	6%	5%	–	–	–	6%	6%
	CTNNA2	6%	11%	15%	–	–	–	10%
	GSR	6%	9%	–	–	6%	–	7%
	PKHD1	6%	12%	–	6%	–	–	8%
Transversally DGC- and IGC-mutated genes	*TP53*	44%	49%	76%	43%	72%	–	**57%**
	PTEN	33%	1%	10%	2%	13%	–	12%
	ARID1A	28%	16%	10%	12%	–	–	16%
	SPTA1	22%	15%	12%	–	–	–	16%
	APC	17%	9%	–	10%	11%	9%	11%
	PIK3CA	17%	7%	–	14%	–	11%	12%
	FAT4	17%	8%	–	6%	–	–	10%
	*LRP2**	17%	13%	–	–	21%	–	17%
	DCLK1	11%	7%	15%	–	–	–	11%
	GLI3	11%	11%	10%	–	–	–	11%
	KRAS	6%	9%	12%	12%	–	7%	9%
	SMAD4	6%	5%	2%	–	11%	5%	6%
	ATM	6%	11%	–	–	9%	–	8%
	RIMS2	6%	13%	17%	–	–	–	12%

Fig. 5. Mutational landscape of IGC subtype. Six sources were used: two CBIOPORTAL data sets and four publications. The origin, number of samples, reported study identification and other cohorts' characteristics are indicated in the heading of the table. The mutated genes of each subtype were divided into two main classifications: transversally mutated and specific genes. The specific mutated genes for IGC were associated with the blue colour and are listed in the upper part of the Table. In the lower part of the table are presented the transversally mutated genes in both DGC and IGC. Mutation frequency was represented in a colour scale. Green was appointed to genes with mutation frequencies between 0% and 9% (less frequently mutated). Yellow was used for mutation frequencies between 10% and 19% and light orange for mutation frequencies between 20% and 39%. A variation between dark orange and red was associated with mutation frequencies increasing from 40% to a maximum of 76% (frequently mutated). The average mutation frequency calculated for each IGC recurrent gene is present in the right column, and a similar colour scale was applied to distinguish the genes with higher mutation frequencies. * LPR2 is enriched in IGC.

been previously associated with GC but not DGC, and *THSD7B* (12%) with no previous association with GC.

There are other specific DGC-mutated genes depicted in Fig. 4 that present on average < 12% of mutations that we used for downstream analyses.

3.3. IGC-specific mutated genes

We next focused on the mutational landscape of IGC from six studies, to identify the set of specific IGC-mutated genes (Fig. 5). No single gene was commonly mutated across all IGC data sets; however, *BRCA2* and *ERBB2* were specific IGC-mutated genes (9% and 10% on average, respectively), according to four out of six studies.

Additionally, *AKAP9*, *DLC1*, *MACF1*, *RNF43*, *NRG1*, *CTNNA2*, *GSR* and *PKHD1* were also identified as specific IGC-mutated genes. *AKAP9* (found in 3/6 studies) was the one with the highest average mutation frequency (15%).

18780261, 2021, 11, Downloaded from https://febs.onlinelibrary.wiley.com/doi/10.1002/1878-0261.12948 by Cochrane Portugal, Wiley Online Library on [31/01/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

3.4. Transversally DGC- and IGC- mutated genes

We then ranked the transversally DGC- and IGC-mutated genes across 3 studies or more (Figs 4 and 5). These were *TP53*, *FAT4*, *ARID1A*, *SPTA1*, *PIK3CA*, *GLI3*, *LRP2*, *RIMS2*, *DCLK1*, *APC*, *KRAS*, *SMAD4*, *ATM* and *PTEN* (Figs 4 and 5). From these 14 genes, a single gene was classified as enriched mutated gene in IGC (*LRP2*), as its mutation frequency was >two-fold higher in IGC (17%) as compared to DGC (8%) (Figs 4 and 5). All remaining genes displayed similar mutation frequencies (<twofold difference between histotypes) in both DGC and IGC.

TP53 and *ARID1A* genes were the most frequently transversally mutated genes across DGC (41% and 15%, respectively) and IGC (57% and 16%, respectively) data sets (Figs 4 and 5). Besides these two important GC driver genes, *APC* and *PTEN* tumour suppressor driver genes, as well as *KRAS* and *PIK3CA* driver oncogenes [62], were also transversally mutated in both GC histotypes with frequencies below 15% across data sets (Figs 4 and 5).

3.5. DGC vs IGC: biological processes and molecular functions associated with mutational profiles

The comparison of molecular functions and biological terms derived from the mutational landscapes of specific DGC genes vs specific IGC genes (Fig. 6) shows that the recurrence and high mutation frequency of *CDH1* and *RHOA* associate DGC with cell adhesion and calcium dependence. On the other hand, specific IGC-mutated genes are mainly associated with tyrosine

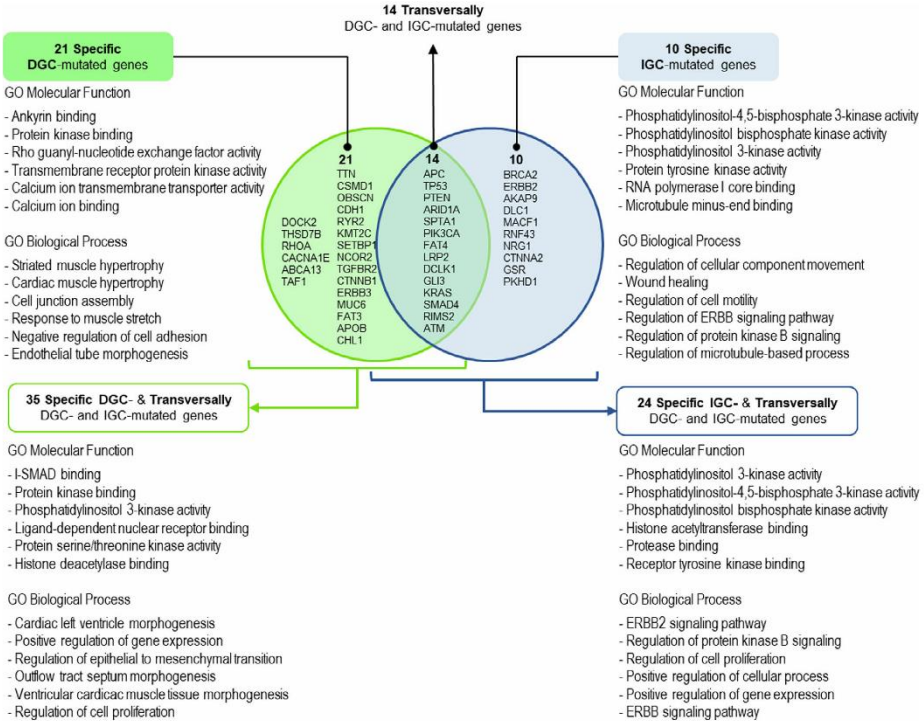


Fig. 6. Integrative analysis of specific- and transversally mutated genes in DGC or IGC. GO terms for molecular functions and biological processes are depicted for the specific genes lists and for the overall set of genes of each subtype (specific- and transversally mutated genes analysed together). The analysis generated using Enrichr and the associated GO terms number is available in Data S2.

18780261, 2021, 11, Downloaded from https://febs.onlinelibrary.wiley.com/doi/10.1002/1878-0261.12948 by Cochrane Portugal, Wiley Online Library on [31/01/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

kinase activity and related signalling pathways, encompassing some promising therapeutic targets (Table 2).

As tumour cells most likely bear histotype-specific mutations together with histotype nonspecific mutations, equivalently frequent, we combined the specific DGC-mutated gene set with the transversally DGC- and IGC-mutated gene set (all genes listed in Fig. 4) and compared the resulting gene list with its IGC counterpart (all genes in Fig. 5) regarding molecular functions and biological terms. This integrated analysis of specific and transversally mutated genes, which is depicted in Fig. 6, turned DGC closer to IGC, with proliferation and tyrosine kinase activity and associated signalling pathways appearing as an output, also in DGC. However, regulation of epithelial-to-mesenchymal transition was still a feature of DGC only. This analysis also revealed the involvement of histone modifications associated with both histological types, histone deacetylase related to DGC and histone acetyltransferase (HAT) to IGC. Details on all analyses are presented in Fig. 6 and Data S2.

3.6. SRCC vs PCC-NOS: distinctive mutational profiles

We then wanted to understand whether the mutational landscape of DGC is common across DGC histological subtypes, or if each subtype (SRCC and PCC-NOS) has its unique molecular profile.

For this, we compared the set of genes mutated in SRCC and in PCC-NOS, described in two Asian studies and obtained with different technical approaches (one targeted sequencing; one WES) (Data S1, Table S1).

3.7. SRCC vs PCC-NOS transversally mutated genes

There were 22 transversally mutated genes across SRCC and PCC-NOS, from which four genes (*CDH1*, *TP53*, *ATM* and *KRAS)* were found in two independent data sets per histological subtype (Fig. 7). *TP53* was the most frequently mutated gene in both DGC subtypes. The remaining transversally mutated genes in SRCC and PCC-NOS were found in a single study, and none of them was mutated >twofold when comparing SRCC and PCC-NOS (Fig. 7).

3.8. SRCC- vs PCC-NOS-enriched mutated genes

Only two genes (*CREBBP* and *MAP2K4*) were found enriched in SRCC (27% for both) as compared to PCC-NOS (10% for both) (Fig. 7). On the other hand, 13 mutated genes were found enriched in PCC-NOS as compared to SRCC. From these, *PTEN, SMAD4*, *SETD2*, *PIK3CA*, *BRAF* and *RPL5* mutations were found in at least 20% of PCC-NOS as compared to a maximum of 9% in SRCC (Fig. 7). Most PCC-NOS enriched mutated genes either had tumour suppressor activity or were associated with histone markers of transcription repression.

3.9. SRCC-specific mutated genes

Six genes (*CDKN2A*, *POLQ*, *SETBP1*, *SOX9*, *TNFAIP3* and *ZFHX3*) were found to be exclusive of SRCC (SRCC-specific mutated genes) (Fig. 7). While *CDKN2A* was found mutated in 27% of the cases, the remaining five genes were found mutated in < 10% of the SRCC.

3.10. PCC-NOS-specific mutated genes

Five genes (*RHOA*, *BRCA2*, *JAK3*, *BRCA1* and *MSH6*), were found exclusively mutated in PCC-NOS (PCC-NOS-specific mutated genes) (Fig. 7). While *RHOA*, *BRCA2* and *JAK3* were found mutated in > 20% of the cases, *BRCA1* and *MSH6* genes were found mutated in 10% of PCC-NOS cases.

This mutation frequency-focused analysis revealed more similarities than differences between SRCC and PCC-NOS, regarding the set of genes affected in both subtypes. However, some very specific features were also found per subtype, likely supporting their morphological differences.

3.11. Insights into similarities and differences between SRCC and PCC-NOS mutational profiles

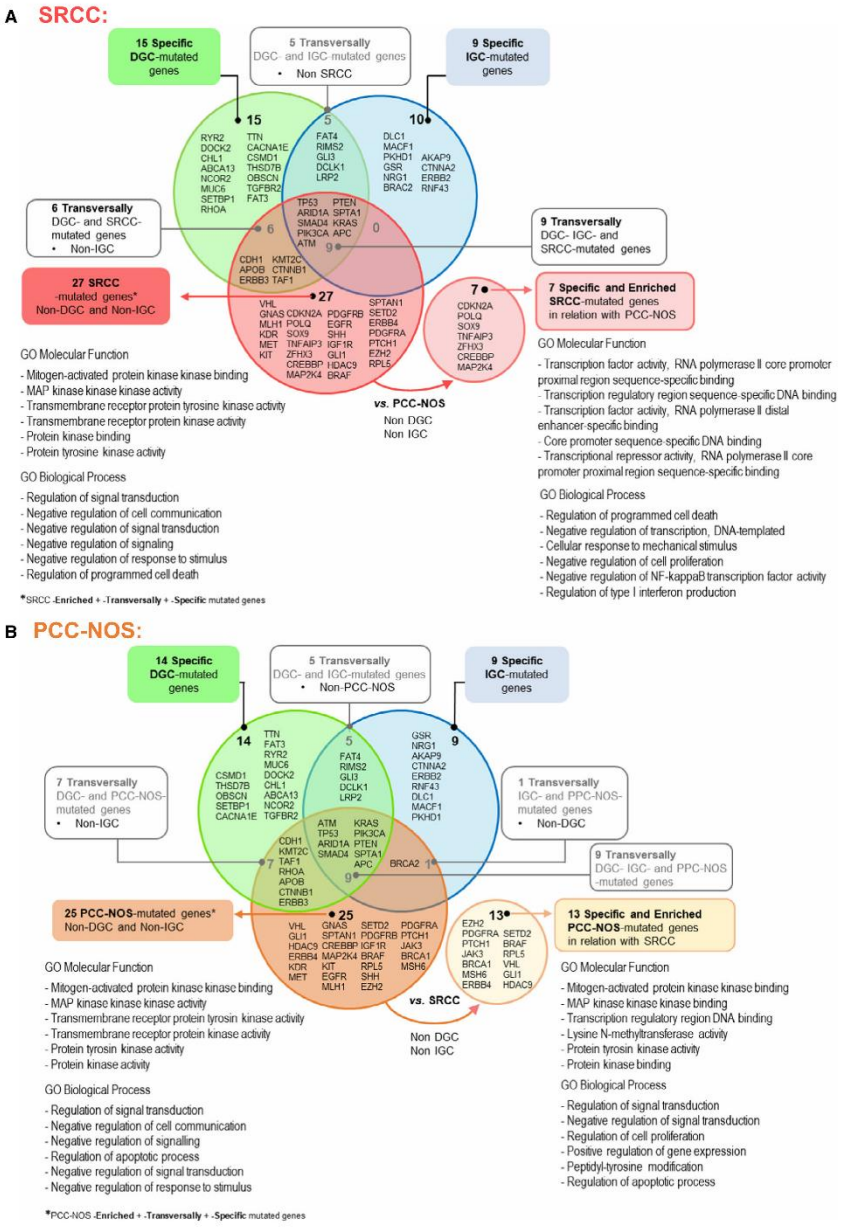
We compared the mutation landscapes of DGC, IGC, SRCC and PCC-NOS aiming to address particular mutational landscapes of SRCC and PCC-NOS. Nine genes were shared by DGC, IGC, PCC-NOS and

Fig. 7. Mutational landscape of SRCC and PCC-NOS. For this analysis, two original publications were used. The origin, number of samples, reported study identification and other cohorts' characteristics are indicated in the heading of the table. The mutated genes of each subtype were divided into three main classifications: enriched, transversally mutated and specific genes. Each mutation frequency was associated with a colour scale. Green was appointed to genes with mutation frequencies between 0% and 9% (less frequently mutated). Yellow was used for mutation frequencies between 10% and 19% and light orange for mutation frequencies between 20% and 39%. A variation between dark orange and red was associated with mutation frequencies increasing from 40% to a maximum of 66% (frequently mutated).

18780261, 2021, 11, Downloaded from https://febs.onlinelibrary.wiley.com/doi/10.1002/1878-0261.12948 by Cochrane Portugal, Wiley Online Library on [31/01/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

Information source		PMID: 28873240	PMID: 31171626	PMID: 28873240	PMID: 31171626	
Origin of samples		**Korea**	**USA, China, Japan**	**Korea**	**USA, China, Japan**	***Recurrence***
Number of samples		*n* = 32	*n* = 11	*n* = 59	*n* = 10	
Cohort Specification		**SRCC**		**PCC-NOS**		
Genes		**Frequency of mutated samples per study**				
Transversally mutated genes	*TP53*	25%	64%	66%	50%	Frequencies reported in both studies
	CDH1	16%	45%	22%	30%	
	ATM	6%	9%	15%	10%	
	KRAS	6%	9%	15%	10%	
PCC-NOS-**enriched** mutated genes	*PTEN*	6%	9%	22%	30%	
	SMAD4	3%	9%	19%	20%	
PCC-NOS-**specific** mutated genes	*RHOA*	0%	0%	25%	0%	
Transversally mutated genes	*APC*	19%	–	32%	–	Frequencies reported in a single study
	KIT	19%	–	29%	–	
	ARID1A	19%	–	27%	–	
	CTNNB1	19%	–	24%	–	
	EGFR	16%	–	29%	–	
	KDR	16%	–	20%	–	
	PDGFRB	16%	–	19%	–	
	ERBB3	13%	–	10%	–	
	SHH	13%	–	8%	–	
	IGF1R	9%	–	17%	–	
	KMT2C	–	36%	–	20%	
	MLH1	–	18%	–	30%	
	TAF1	–	18%	–	20%	
	SPTA1	–	18%	–	10%	
	APOB	–	9%	–	10%	
	GNAS	–	9%	–	10%	
	MET	–	9%	–	10%	
	SPTAN1	–	9%	–	10%	
PCC-NOS-**specific** mutated genes	*BRCA2*	–	0%	–	20%	
	JAK3	–	0%	–	20%	
	BRCA1	–	0%	–	10%	
	MSH6	–	0%	–	10%	
PCC-NOS-**enriched** mutated genes	*SETD2*	–	9%	–	30%	
	PIK3CA	9%	–	24%	–	
	BRAF	6%	–	24%	–	
	RPL5	–	9%	–	20%	
	VHL	9%	–	19%	–	
	GLI1	9%	–	19%	–	
	HDAC9	9%	–	19%	–	
	ERBB4	9%	–	19%	–	
	EZH2	6%	–	17%	–	
	PDGFRA	6%	–	17%	–	
	PTCH1	3%	–	17%	–	
SRCC-**specific** mutated genes	*CDKN2A*	–	27%	–	0%	
	POLQ	–	9%	–	0%	
	SETBP1	–	9%	–	0%	
	SOX9	–	9%	–	0%	
	TNFAIP3	–	9%	–	0%	
	ZFHX3	–	9%	–	0%	
SRCC-**enriched** mutated genes	*CREBBP*	–	27%	–	10%	
	MAP2K4	–	27%	–	10%	
Cohort Specification		**SRCC**		**PCC-NOS**		
Information source		PMID: 28873240	PMID: 31171626	PMID: 28873240	PMID: 31171626	

18780261, 2021, 11, Downloaded from https://febs.onlinelibrary.wiley.com/doi/10.1002/1878-0261.12948 by Cochrane Portugal, Wiley Online Library on [31/01/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License



18780261, 2021, 11, Downloaded from https://febs.onlinelibrary.wiley.com/doi/10.1002/1878-0261.12948 by Cochrane Portugal, Wiley Online Library on [31/01/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

Fig. 8. Specific and transversally mutated genes in SRCC (A) and PCC/NOS (B) in comparison with DGC and IGC. SRCC-associated genes were linked to a red colour, while PCC-NOS-associated genes are related to the orange colour. In both Fig. 8 A, B, the circle outside of the Venn diagram contains the specific genes for SRCC and PCC-NOS, respectively. GO terms for molecular functions and biological processes are depicted for the mutated genes associated with each subtype that were not shared with DGC nor IGC, as well for the SRCC/PCC-NOS-specific genes. The analysis generated using *Enrichr* and the associated GO terms number is available in Data S2.

SRCC. These are well-known cancer driver genes with crucial functions related to cell proliferation, migration, DNA repair or angiogenesis [63]. Some are well-known tumour suppressors such as *TP53*, *ATM*, *APC*, *ARID1A*, *SMAD4* and *PTEN* or oncogenes such as *PIK3CA* and *KRAS* (Fig. 8A,B).

SRCC shares a set of six mutated genes with PCC-NOS, which are also found in DGC but not in IGC (*CDH1, APOB, ERBB3, KMT2C, CTNNB1, TAF1*) (Fig. 8A,B). Among these six genes is the hallmark gene *CDH1* (Fig. 8A,B). There are no genes shared between SRCC and IGC, and *BRCA2* is the only mutated gene shared by PCC-NOS and IGC (Fig. 8B). These data likely identify the mutated genes that are common in DGC, SRCC and PCC-NOS, in keeping with the fact that the two latter are subtypes of the former. SRCC and PCC-NOS diverge in morphology and also in their mutational profile. We found seven genes that are SRCC-specific and enriched genes (*CDKN2A, POLQ, SOX9, TNFAIP3, ZFHX3, CREBBP and MAP2K4*) in relation to PCC-NOS (Fig. 8A). Both in the restricted set of seven specific and enriched SRCC-mutated genes and in a larger set of 27 SRCC-mutated genes (Fig. 8A), an association with programmed cell death is always identified among GO biological processes, representing a novel feature of SRCC. Moreover, the seven genes that differentiate SRCC from all other GC histotypes or subtypes are mainly associated with transcriptional activity (TOP6 ranking GO molecular functions) and cell proliferation and immune or inflammatory response (TOP6 GO biological processes) (Fig. 8A).

We found 25 mutated genes that differentiate PCC-NOS from DGC and IGC, and from these, 13 are PCC-NOS-specific and enriched genes (*EZH2, SETD2, PDGFRA, BRAF, PTCH1, RPL5, JAK3, VHL, BRCA1, GLI1, MSH6, HDAC9 and ERBB4*) in relation to SRCC (Fig. 8B). This combination of mutated genes likely represents a distinctive feature of PCC-NOS, which despite being mainly associated with MAPK pathway and protein kinase activity, as for the main GC histotypes, it is specifically associated with transcription regulatory region DNA binding (TOP6 ranking GO molecular functions). This feature brings PCC-NOS closer to SRCC (Fig. 8B).

In summary, this analysis supports a divergent evolution of SRCC and PCC-NOS mutation profiles and highlights particular molecular pathways and functions from each of these two DGC subtypes.

3.12. Therapies specifically directed to DGC do not consider DGC mutational landscape

We searched for ongoing and recently closed clinical trials specifically mentioning the inclusion of patients with DGC or DGC subtypes to understand whether any of these trials used as therapy target, the genes or pathways herein identified has preferentially involved in DGC and its subtypes.

We found 13 clinical trials using the search criteria described in the methodology (Table S2). In total, we found four trials directed to DGC patients, two completed (DIGEST and MONO) and two currently recruiting or about to start recruitment (NORDICA and ADCI002 Study) (Table 3).

The DIGEST phase III study (NCT01285557) compared the use of fluorouracil (5-FU) combined with cisplatin with the use of the latter plus S-1 in 361 patients with metastatic DGC (Table 3). This trial has been completed, and results showed that there was no significant difference in the outcome of the patients comparing both approaches [64].

The MONO study (NCT01197885), a phase II clinical trial, examined the outcome of zolbetuximab (monoclonal antibody against CLDN18.2: IMAB362) monotherapy in a series of patients with recurrent or refractory, locally advanced or metastatic and CLDN18.2-positive gastric, gastro-oesophageal junction, or oesophageal adenocarcinoma (Table 3). From the 54 patients enrolled in this trial, 22 were advanced DGC patients with CLDN18.2 expression in $\geq$ 50% of tumour cells. Zolbetuximab was well-tolerated and exhibited antitumour activity with a clinical benefit rate of 23% [60]; whether patients benefiting from this targeted therapy were DGC patients is unknown.

The ADCI002 phase II/III clinical trial (NCT01717924), that is currently recruiting, was designed to address the inherent chemoresistance of SRCC. It evaluates primary surgery vs primary chemotherapy followed by surgery, using epirubicin for topoisomerase II inhibition combined with cisplatin and 5-FU [65].

18780261, 2021, 11, Downloaded from https://febs.onlinelibrary.wiley.com/doi/10.1002/1878-0261.12948 by Cochrane Portugal, Wiley Online Library on [31/01/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

Table 3. Details of clinical trials registered at ClinicalTrials.gov directed to DGC patients.Green was used to highlight clinical trials that specifically mention DGC; red was used to highlight clinical trials that specifically mention SRCC; and white was used to highlight a specific clinical trial with promising results based on CLDN18.2 expression (Table 2).

ClinicalTrials.gov Identifier	Title of the clinical trial	Strategy	Molecular target	Status	Publications	Clinical trial outcome
NCT03977220	Nab-paclitaxel combined with S-1 treating diffuse type of stage gastric cancer as adjuvant setting (NORDICA)	Evaluate paclitaxel for microtubules stabilization and S-1 (oral derivative of 5-FU)	Microtubules and conventional chemotherapy targets	Not yet recruiting	No publications available	Not available
NCT01717924	Evaluation of Surgery vs Primary Chemotherapy in Resectable Signet-ring Cell Gastric Adenocarcinoma (ADCI002) (Phase II/III)	Compare primary surgery vs primary chemotherapy followed by surgery. epirubicin for topoisomerase II inhibition; cisplatin; 5-fluorouracil	Topoisomerase II and conventional chemotherapy targets	Recruiting	[65]	Not available
NCT01285557	Diffuse Gastric and Esophagogastric Junction Cancer S-1 Trial (DIGEST) (Phase III)	Evaluate the safety and efficacy of S-1 and cisplatin compared with 5-FU and cisplatin	Conventional chemotherapy targets	Completed	[64]	No significant difference has been found in the outcome of the patients comparing both approaches
NCT01197885	Efficacy and Safety Study of Multiple Doses of IMAB362 in Patients with Advanced Gastroesophageal Cancer (MONO) (Phase IIa)	Evaluate the safety and efficacy of IMAB362 used as monoclonal antibody against CLDN18.2 (single agent) in GC patients (22/54 DGC)	CLDN18.2	Completed	[60]	Zolbetuximab was well-tolerated and exhibited antitumour activity with a clinical benefit rate for patients of 23%; whether these were DGC patients was not disclosed

The NORDICA phase I study (NCT03977220), which is specifically directed to DGC patients, is not yet recruiting and aims at using microtubules stabilization agent (paclitaxel) combined with S-1 (Table 3).

None of the trials described above used as target for therapy, the specific mutations or mutation profiles and associated molecular functions herein found to differentiate DGC and its subtypes from each other and from IGC.

3.13. Sporadic DGC and hereditary DGC (HDGC): somatic mutational profiles

We tried to understand whether DGC arising in different settings (hereditary or sporadic) would share a similar mutational profile, and whether DGC arising in the HDGC context would mimic a mutational profile closer to that of SRCC.

The so-called HDGC syndrome is caused mainly by germline *CDH1* and rarely by *CTNNA1* pathogenic or likely pathogenic variants and is dominated by the development of early-onset DGC, mainly SRCC [66–68]. However, the somatic mutational landscape of DGC/SRCC arising in this setting is, to date, fairly unknown.

We could only find two studies reporting the somatic mutational landscape of HDGC tumours [67,69]. The germline causative defect in one family affected *CDH1* [69] and in the other family affected *CTNNA1* [67].

We compared the list of somatic mutated genes found in these HDGC cases with that found in sporadic DGC cases (Fig. 4). The mutational profile of the *CTNNA1* germline mutated tumour [67] encompassed somatic mutations in *RHOA*, a specific DGC and PCC-NOS gene (Fig. 7), and in *ARID1A* and *PIK3CA*, genes transversally mutated in both DGC and IGC (Fig. 7). This suggests that, independently of the cancer setting (hereditary vs diffuse), or histological type (DGC vs IGC), GC developing in this *CTNNA1* variant germline carrier presents a

18780261, 2021, 11, Downloaded from https://febs.onlinelibrary.wiley.com/doi/10.1002/1878-0261.12948 by Cochrane Portugal, Wiley Online Library on [31/01/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

nonspecific combination of somatic events. By contrast, for the *CDH1* germline mutated tumour, somatic mutations were described in *TTN* and *CDH1* by Funakoshi *et al* [69], which are both specific DGC-mutated genes (this study, Fig. 6); being *CDH1* was also considered transversally mutated in SRCC and PCC-NOS (Fig. 7).

To better understand whether there are real differences between the mutational profiles of sporadic and hereditary DGC, or whether HDGC cases are closer to a particular DGC subtype, additional studies need to be performed.

4. Discussion

In the present study, we started by revising GC histological and molecular classifications, highlighting associated relevant molecular landscapes, clinical outcomes and currently approved or promising therapies. The state of the art on all these aspects is presented in the introduction section. The main outcome of this literature revision was the identification of an asymmetry, regarding prognosis, therapy targets and available effective therapies (both conventional and targeted), in TCGA GS and ACRG MSS/EMT, mainly encompassing DGC, in comparison with other molecular subgroups.

To address this gap in knowledge, and shed light into potential targets for therapeutic intervention, we explored a set of studies reporting on multigene mutational analysis to identify DGC-associated mutational profiles, molecular pathways and biological functions. We chose this approach, as mutational profiles are rather stable genetic events, easy to detect, interpret and less costly as compared to RNA profiles, methylation patterns or other epigenomic approaches. On the other hand, several tumour-specific mutations constitute well-proven, highly sensitive and specific predictive biomarkers of response to selective targeted therapies and are now an indispensable laboratory tool for therapeutic decisions [70].

Our systematic analysis confirms that *TP53, ARID1A, SMAD4, PIK3CA, PTEN, KRAS, APC, ATM and SPTA1* are driver and central genes in gastric tumorigenesis, and transversal to the main GC histotypes and DGC subtypes [62].

Also, herein we identified specific sets of genes that define the main molecular pathways classically involved in either DGC or IGC. However, tumours emerge and evolve due to a mutational profile that can be shared with other tumour types. We could find that the combination of all recurrently mutated genes either in DGC or in IGC triggered the involvement of similar molecular pathways regardless of being histotype-specific or common to both histotypes. This may explain why conventional therapies, and even some approved targeted therapies, demonstrate clinical benefit regardless of the GC histological type.

We further showed that the most recurrently mutated and specific genes in DGC, across many studies, were *CDH1* and *RHOA*. This result is reflected in the findings from the TCGA GS and ACRG MSS/EMT subgroups, which are dominated by DGC. According to both studies, this subgroup presents the worst prognosis within GC molecular subtypes, resistance to conventional chemotherapy and so far, have no approved targeted therapies [22,37,49]. Interestingly, a recent study reported that late-onset DGCs presented higher mutation frequencies in *RHOA* and less frequent mutations in *CDH1* in comparison with early-onset DGCs [31,71].

Although none of the ongoing clinical trials directed to DGC patients use these two proteins as targets for therapy (Table 3), it is tempting to speculate whether their mutations would perform well as therapy targets or predictors of therapy responsiveness. E-cadherin inactivation is a highly unlikely druggable target mechanism, and it has been shown to cause multidrug resistance. By contrast, E-cadherin disruption is also known to induce upregulation of the anti-apoptotic protein BCL2, and activation of Rho-Rock pathway, which may be targetable (reviewed in Ref. [72]). Taking advantage of the pro-survival context in E-cadherin deficient DGC, treatment using pro-apoptotic drugs may be feasible. Indeed, preclinical treatment with pro-apoptotic BH3-only mimetics combined with chemotherapy or endocrine treatment regimens had a positive effect in breast cancer [73,74]. Preclinical studies also suggested that pharmacological inhibition of Rho-Rock signalling may be a clinically relevant target in cancer [75]. Other genes specifically mutated in DGC may also be interesting to consider when rethinking biomarkers for prognosis and therapy in patients with DGC. An example is the *TTN* gene, whose mutations were found specifically in DGC and on average in 40% of the cases. Although not previously related specifically to DGC, *TTN* was found as one of the top five hub genes of a specific co-expression module positively correlated with GC pathological tumour and lymph node stages [76]. In another study, *TTN* gene mutations were correlated with poor prognosis and predicted tumour mutation burden and immunotherapy efficacy in GC [77]. Other examples are *FAT3* and *KMT2C*, herein found mutated on average in 15% and 12% of DGC, respectively, and found previously associated with this cancer

18780261, 2021, 11, Downloaded from https://febs.onlinelibrary.wiley.com/doi/10.1002/1878-0261.12948 by Cochrane Portugal, Wiley Online Library on [31/01/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

histotype [76,78,79]. Namely, *KMT2C* loss was shown to promote epithelial-to-mesenchymal transition and was associated with poor overall survival.

Although our study has not addressed the role of epigenetic changes in DGC, these events are also considered significant carcinogenic factors, and epigenetic regulation has been claimed as a promising molecular target for therapy in GC precision oncology [17,80]. Herein, we have shown that histone modifications appear not only in the TOP6 molecular functions from DGC, but also in those from IGC, and when specific and transversally mutated genes are considered. This finding reinforces the need of studies focusing specifically on the epigenetics profiles of DGC and its subtypes, with the aim of finding potential targets for precision oncology.

The current systematic analysis also provides, to our knowledge the first combined analysis on SRCC and PCC-NOS mutational profiles, considering two independent studies. This revealed *CDH1*, *TP53*, *ATM* and *KRAS* mutations as recurrent and common events of both DGC subtypes, again supporting the role of GC driver genes, and *RHOA* mutations as a specificity of PCC-NOS [16]. In terms of more specific mutational landscapes, both sets of divergent mutated genes between overall DGC and SRCC and between overall DGC and PCC-NOS are associated with TRK pathways, cell–cell communication and apoptosis, although involving different genes. However, and interestingly, the set of genes that distinguish SRCC and PCC-NOS from DGC overall and from each other associates with transcriptional regulation and gene expression control, besides TRK pathways and apoptosis, particularly the SRCC subtype. Indeed, *CREBBP* and *MAP2K4* genes, whose mutations are enriched in SRCC as compared to PCC-NOS (Fig. 7), are involved in coupling chromatin remodelling to transcription factor recognition and transcription regulation, respectively [81,82]. A loss-of-function mutation in the *CBP/CREBBP* gene, which encodes a HAT, is common to many cancers, including GC [82], and *MAP2K4* loss of expression has been reported in gastric adenocarcinoma associated with poor survival [81]. Given that a great deal of transcriptional control involves epigenetic regulation, epigenetic-related treatments may hold promise for SRCC and PCC-NOS precision oncology [80], besides the above-mentioned possibilities described for DGC overall.

To the best of our knowledge, this is the first systematic analysis describing the distinctive mutational profiles of DGC and its subtypes, resourcing to a recurrent analysis, to address the gap in potential treatment options for this GC histotype. By exploring the literature regarding clinical trials directed to the particularities of DGC, we concluded that these are only a few and rarely specifically designed for DGC patients. This may explain the delay in finding treatment options for these patients. Because we were not able to find more promising results than those obtained for anti-Claudin18.2 targeted therapy, we herein highlight a set of recurrent mutant genes and associated pathways that may be explored for novel therapeutic designs in DGC and its subtypes. The interpretation of past clinical trials in the light of tumour morphology can also guide therapeutic development by identifying subsets of patients with better response to a certain treatment strategy. Exploring the mutational landscape of DGC subtypes can further identify new and more adequate druggable targets with therapeutic implications.

Finally, improving the knowledge on somatic molecular/genomic players in sporadic DGC may also be valuable to guide future treatments in the HDGC setting, which also remains poorly studied from the somatic standpoint and without chemotherapeutic options.

Acknowledgements

The authors acknowledge the support of the National Infrastructure 'GenomePT' – National Laboratory for Genome Sequencing and Analysis (POCI-01-0145-FEDER-022184). J.G-P acknowledges the support of Faculty of Medicine from University of Porto, specifically by the Doctoral Programme in Biomedicine and the Solve-RD Project (H2020-SC1-2017-Single-Stage-RTD) for his PhD fellowship. R.B-M. acknowledges the support of Institute of Biomedical Sciences Abel Salazar from University of Porto, specifically by the Doctoral Programme BiotechHealth and the FCT – Fundação para a Ciência e a Tecnologia for her PhD fellowship (ref. SFRH/BD/145132/2019).

This research and its authors were funded 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). This work was also financed by the projects NORTE-01-0145-FEDER-000003 (DOCnet) and NORTE-01-0145-FEDER-000029 (CANCER) – supported by Norte Portugal Regional Programme (NORTE 2020), under the PORTUGAL 2020 Partnership Agreement, through the

18780261, 2021, 11, Downloaded from https://febs.onlinelibrary.wiley.com/doi/10.1002/1878-0261.12948 by Cochrane Portugal, Wiley Online Library on [31/01/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

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.

Conflict of interest

The authors declare no conflict of interest.

Author contributions

CO and FC conceived and supervised the study; CO, RB-M and JG-P designed the study; RB-M, JG-P and IG acquired the data; RB-M, JG-P, IG, FC and CO analysed and interpreted the data; all authors drafted the manuscript; FC, IG and CO edited and reviewed the manuscript; CO and FC obtained funding. All authors critically revised the manuscript for important intellectual content. All authors have read and agreed to the published version of the manuscript.

References

1 Gullo I, Carneiro F, Oliveira C & Almeida GM (2018) Heterogeneity in gastric cancer: from pure morphology to molecular classifications. *Pathobiology* **85**, 50–63.
2 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.
3 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.
4 Ming SC (1977) Gastric carcinoma. A pathobiological classification. *Cancer* **39**, 2475–2485.
5 Nakamura K, Sugano H & Takagi K (1968) Carcinoma of the stomach in incipient phase: its histogenesis and histological appearances. *Gan* **59**, 251–258.
6 Solcia E, Klersy C, Mastracci L, Alberizzi P, Candusso M, Diegoli M, Tava F, Riboni R, Manca R & Luinetti O (2009) A combined histologic and molecular approach identifies three groups of gastric cancer with different prognosis. *Virchows Arch* **455**, 197–211.
7 Japanese Gastric Cancer Association (2017) Japanese Classification of Gastric Carcinoma, 15th edn. Kanehara Shuppan, Tokyo.
8 WHO Classification of Tumours Editorial Board (2019) Digestive system tumours, 5th edn. WHO classification of tumours series, Lyon: International Agency for Research on Cancer.
9 Carneiro F, Fukayama M, Grabsch H & Yasui WGA (2019) WHO Classification of Tumours Editorial Board. Digestive system tumours, pp. 85–95, WHO classification of tumours series, Lyon: International Agency for Research on Cancer.
10 Park HK, Lee KY, Yoo MW, Hwang TS & Han HS (2016) Mixed carcinoma as an independent prognostic factor in submucosal invasive gastric carcinoma. *J Korean Med Sci* **31**, 866–872.
11 Zheng HC, Li XH, Hara T, Masuda S, Yang XH, Guan YF & Takano Y (2008) Mixed-type gastric carcinomas exhibit more aggressive features and indicate the histogenesis of carcinomas. *Virchows Arch* **452**, 525–534.
12 Carneiro F (1997) Classification of gastric carcinomas. *Curr Diagn Pathol* **4**, 51–59.
13 Mariette C, Carneiro F, Grabsch HI, van der Post RS, Allum W & de Manzoni G (2019) Consensus on the pathological definition and classification of poorly cohesive gastric carcinoma. *Gastric Cancer* **22**, 1–9.
14 Sugihara H, Hattori T, Fukuda M & Fujita S (1987) Cell proliferation and differentiation in intramucosal and advanced signet ring cell carcinomas of the human stomach. *Virchows Arch A Pathol Anat Histopathol* **411**, 117–127.
15 Gullo I, van der Post RS & Carneiro F (2021) Recent advances in the pathology of heritable gastric cancer syndromes. *Histopathology* **78**, 125–147.
16 Kwon CH, Kim YK, Lee S, Kim A, Park HJ, Choi Y, Won YJ, Park DY & Lauwers GY (2018) Gastric poorly cohesive carcinoma: a correlative study of mutational signatures and prognostic significance based on histopathological subtypes. *Histopathology* **72**, 556–568.
17 Ooi WF, Xing M, Xu C, Yao X, Ramlee MK, Lim MC, Cao F, Lim K, Babu D, Poon LF *et al.* (2016) Epigenomic profiling of primary gastric adenocarcinoma reveals super-enhancer heterogeneity. *Nat Commun* **7**, 12983.
18 Miki H, Ohmori M, Perantoni AO & Enomoto T (1991) K-ras activation in gastric epithelial tumors in Japanese. *Cancer Lett* **58**, 107–113.
19 Oliveira C, Pinto M, Duval A, Brennetot C, Domingo E, Espín E, Armengol M, Yamamoto H, Hamelin R, Seruca R *et al.* (2003) BRAF mutations characterize colon but not gastric cancer with mismatch repair deficiency. *Oncogene* **22**, 9192–9196.
20 Samuels Y, Wang Z, Bardelli A, Silliman N, Ptak J, Szabo S, Yan H, Gazdar A, Powell SM, Riggins GJ *et al.* (2004) High frequency of mutations of the PIK3CA gene in human cancers. *Science* **304**, 554.
21 Velho S, Oliveira C, Ferreira A, Ferreira AC, Suriano G, Schwartz S Jr, Duval A, Carneiro F, Machado JC, Hamelin R *et al.* (2005) The prevalence of PIK3CA mutations in gastric and colon cancer. *Eur J Cancer* **41**, 1649–1654.

18780261, 2021, 11, Downloaded from https://febs.onlinelibrary.wiley.com/doi/10.1002/1878-0261.12948 by Cochrane Portugal, Wiley Online Library on [31/01/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

22 Cancer Genome Atlas Research Network (2014) Comprehensive molecular characterization of gastric adenocarcinoma. *Nature* **513**, 202–209.
23 Hu BL, Xie MZ, Li KZ, Li JL, Gui YC & Xu JW (2019) Genome-wide analysis to identify a novel distant metastasis-related gene signature predicting survival in patients with gastric cancer. *Biomed Pharmacother* **117**, 109159.
24 Pirini F, Noazin S, Jahuira-Arias MH, Rodriguez-Torres S, Friess L, Michailidi C, Cok J, Combe J, Vargas G, Prado W *et al.* (2017) Early detection of gastric cancer using global, genome-wide and IRF4, ELMO1, CLIP4 and MSC DNA methylation in endoscopic biopsies. *Oncotarget* **8**, 38501–38516.
25 Cui J, Yin Y, Ma Q, Wang G, Olman V, Zhang Y, Chou WC, Hong CS, Zhang C, Cao S *et al.* (2015) Comprehensive characterization of the genomic alterations in human gastric cancer. *Int J Cancer* **137**, 86–95.
26 Hanazono K, Natsugoe S, Stein HJ, Aikou T, Hoefler H & Siewert JR (2006) Distribution of p53 mutations in esophageal and gastric carcinomas and the relationship with p53 expression. *Oncol Rep* **15**, 821–824.
27 van Beek E, Hernandez JM, Goldman DA, Davis JL, McLaughlin K, Ripley RT, Kim TS, Tang LH, Hechtman JF, Zheng J *et al.* (2018) Rates of TP53 Mutation are Significantly Elevated in African American Patients with Gastric Cancer. *Ann Surg Oncol* **25**, 2027–2033.
28 Becker KF, Atkinson MJ, Reich U, Becker I, Nekarda H, Siewert JR & Höfler H (1994) E-cadherin gene mutations provide clues to diffuse type gastric carcinomas. *Cancer Res* **54**, 3845–3852.
29 Berx G, Becker KF, Höfler H & van Roy F (1998) Mutations of the human E-cadherin (CDH1) gene. *Hum Mutat* **12**, 226–237.
30 Bustos-Carpinteyro AR, Oliveira C, Sousa A, Oliveira P, Pinheiro H, Carvalho J, Magaña-Torres MT, Flores-Miramontes MG, Aguilar-Lemarroy A, Jave-Suárez LF *et al.* (2019) CDH1 somatic alterations in Mexican patients with diffuse and mixed sporadic gastric cancer. *BMC Cancer* **19**, 69.
31 Cho SY, Park JW, Liu Y, Park YS, Kim JH, Yang H, Um H, Ko WR, Lee BI, Kwon SY *et al.* (2017) Sporadic early-onset diffuse gastric cancers have high frequency of somatic CDH1 alterations, but low frequency of somatic RHOA mutations compared with late-onset cancers. *Gastroenterology* **153**, 536–549.e26.
32 Moridnia A, Tabatabaiefar MA, Zeinalian M, Minakari M, Kheirollahi M & Moghaddam NA (2019) Novel variants and copy number variation in CDH1 gene in Iranian patients with sporadic diffuse gastric cancer. *J Gastrointest Cancer* **50**, 420–427.
33 Nemtsova MV, Kalinkin AI, Kuznetsova EB, Bure IV, Alekseeva EA, Bykov II, Khorobrykh TV, Mikhaylenko DS, Tanas AS, Kutsev SI *et al.* (2020) Clinical relevance of somatic mutations in main driver genes detected in gastric cancer patients by next-generation DNA sequencing. *Sci Rep* **10**, 504.
34 Wang SC, Yeu Y, Hammer STG, Xiao S, Zhu M, Hong C, Clemenceau JR, Yoon LY, Nassour I, Shen J *et al.* (2020) Hispanic/Latino patients with gastric adenocarcinoma have distinct molecular profiles including a high rate of germline CDH1 variants. *Cancer Res* **80**, 2114–2124.
35 Tan IB, Ivanova T, Lim KH, Ong CW, Deng N, Lee J, Tan SH, Wu J, Lee MH, Ooi CH *et al.* (2011) Intrinsic subtypes of gastric cancer, based on gene expression pattern, predict survival and respond differently to chemotherapy. *Gastroenterology* **141**, 476–485.e11.
36 Lei Z, Tan IB, Das K, Deng N, Zouridis H, Pattison S, Chua C, Feng Z, Guan YK, Ooi CH *et al.* (2013) Identification of molecular subtypes of gastric cancer with different responses to PI3-kinase inhibitors and 5-fluorouracil. *Gastroenterology* **145**, 554–565.
37 Cristescu R, Lee J, Nebozhyn M, Kim KM, Ting JC, Wong SS, Liu J, Yue YG, Wang J, Yu K *et al.* (2015) Molecular analysis of gastric cancer identifies subtypes associated with distinct clinical outcomes. *Nat Med* **21**, 449–456.
38 Setia N, Agoston AT, Han HS, Mullen JT, Duda DG, Clark JW, Deshpande V, Mino-Kenudson M, Srivastava A, Lennerz JK *et al.* (2016) A protein and mRNA expression-based classification of gastric cancer. *Modern Pathol* **29**, 772–784.
39 Ahn S, Lee SJ, Kim Y, Kim A, Shin N, Choi KU, Lee CH, Huh GY, Kim KM, Setia N *et al.* (2017) High-throughput protein and mRNA expression-based classification of gastric cancers can identify clinically distinct subtypes, concordant with recent molecular classifications. *Am J Surg Pathol* **41**, 106–115.
40 Gonzalez RS, Messing S, Tu X, McMahon LA & Whitney-Miller CL (2016) Immunohistochemistry as a surrogate for molecular subtyping of gastric adenocarcinoma. *Hum Pathol* **56**, 16–21.
41 Diaz Del Arco C, Estrada Munoz L, Molina Roldan E, Ceron Nieto MA, Ortega Medina L, Gomez G, de Las Heras S & Fernandez Acenero MJ (2018) Immunohistochemical classification of gastric cancer based on new molecular biomarkers: a potential predictor of survival. *Virchows Arch* **473**, 687–695.
42 Birkman EM, Mansuri N, Kurki S, Algars A, Lintunen M, Ristamaki R, Sundstrom J & Carpen O (2018) Gastric cancer: immunohistochemical classification of molecular subtypes and their association with clinicopathological characteristics. *Virchows Arch* **472**, 369–382.

15780261, 2021, 11, Downloaded from https://febs.onlinelibrary.wiley.com/doi/10.1002/1878-0261.12948 by Cochrane Portugal, Wiley Online Library on [31/01/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

43 Koh J, Lee KW, Nam SK, Seo AN, Kim JW, Kim JW, Park DJ, Kim HH, Kim WH & Lee HS (2019) Development and validation of an easy-to-implement, practical algorithm for the identification of molecular subtypes of gastric cancer: prognostic and therapeutic implications. *Oncologist* **24**, e1321–e1330.

44 Di Pinto F, Armentano R, Arborea G, Schena N, Donghia R & Valentini AM (2020) Are immunohistochemical markers useful in phenotypic gastric cancer classification? *Oncology* **98**, 566–574.

45 Wang Q, Xie Q, Liu Y, Guo H, Ren Y, Li J & Zhao Q (2020) Clinical characteristics and prognostic significance of TCGA and ACRG classification in gastric cancer among the Chinese population. *Mol Med Rep* **22**, 828–840.

46 Pinto MP, Córdova-Delgado M, Retamal IN, Muñoz-Medel M, Bravo ML, Durán D, Villanueva F, Sanchez C, Acevedo F, Mondaca S *et al.* (2020) A molecular stratification of Chilean gastric cancer patients with potential clinical applicability. *Cancers* **12**, 1863.

47 Tsai JH, Jeng YM, Chen KH, Lee CH, Yuan CT & Liau JY (2020) An integrative morphomolecular classification system of gastric carcinoma with distinct clinical outcomes. *Am J Surg Pathol* **44**, 1017–1030.

48 Zhao C, Feng Z, He H, Zang D, Du H, Huang H, Du Y, He J, Zhou Y & Nie Y (2020) Protein expression-based classification of gastric cancer by immunohistochemistry of tissue microarray. *PLoS One* **15**, e0238836.

49 Sohn BH, Hwang JE, Jang HJ, Lee HS, Oh SC, Shim JJ, Lee KW, Kim EH, Yim SY, Lee SH *et al.* (2017) Clinical significance of four molecular subtypes of gastric cancer identified by the cancer genome atlas project. *Clin Cancer Res* **23**, 4441–4449.

50 Shu Y, Zhang W, Hou Q, Zhao L, Zhang S, Zhou J, Song X, Zhang Y, Jiang D, Chen X *et al.* (2018) Prognostic significance of frequent CLDN18-ARHGAP26/6 fusion in gastric signet-ring cell cancer. *Nat Commun* **9**, 2447.

51 Nakayama I, Shinozaki E, Sakata S, Yamamoto N, Fujisaki J, Muramatsu Y, Hirota T, Takeuchi K, Takahashi S, Yamaguchi K *et al.* (2019) Enrichment of CLDN18-ARHGAP fusion gene in gastric cancers in young adults. *Cancer Sci* **110**, 1352–1363.

52 Ishigami H, Fujiwara Y, Fukushima R, Nashimoto A, Yabusaki H, Imano M, Imamoto H, Kodera Y, Uenosono Y, Amagai K *et al.* (2018) Phase III trial comparing intraperitoneal and intravenous paclitaxel plus S-1 versus cisplatin plus S-1 in patients with gastric cancer with peritoneal metastasis: PHOENIX-GC trial. *J Clin Oncol* **36**, 1922–1929.

53 Selim JH, Shaheen S, Sheu WC & Hsueh CT (2019) Targeted and novel therapy in advanced gastric cancer. *Exp Hematol Oncol* **8**, 25.

54 Fashoyin-Aje L, Donoghue M, Chen H, He K, Veeraraghavan J, Goldberg KB, Keegan P, McKee AE & Pazdur R (2019) FDA approval summary: pembrolizumab for recurrent locally advanced or metastatic gastric or gastroesophageal junction adenocarcinoma expressing PD-L1. *Oncologist* **24**, 103–109.

55 Kang YK, Boku N, Satoh T, Ryu MH, Chao Y, Kato K, Chung HC, Chen JS, Muro K, Kang WK *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.

56 Kim ST, Cristescu R, Bass AJ, Kim KM, Odegaard JI, Kim K, Liu XQ, Sher X, Jung H, Lee M *et al.* (2018) Comprehensive molecular characterization of clinical responses to PD-1 inhibition in metastatic gastric cancer. *Nat Med* **24**, 1449–1458.

57 Bang YJ, Van Cutsem E, Feyereislova A, Chung HC, Shen L, Sawaki A, Lordick F, Ohtsu A, Omuro Y, Satoh T *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.

58 Bardou P, Mariette J, Escudié F, Djemiel C & Klopp C (2014) jvenn: an interactive Venn diagram viewer. *BMC Bioinformatics* **15**, 293.

59 Chen EY, Tan CM, Kou Y, Duan Q, Wang Z, Meirelles GV, Clark NR & Ma'ayan A (2013) Enrichr: interactive and collaborative HTML5 gene list enrichment analysis tool. *BMC Bioinformatics* **14**, 128.

60 Türeci O, Sahin U, Schulze-Bergkamen H, Zvirbule Z, Lordick F, Koeberle D, Thuss-Patience P, Ettrich T, Arnold D, Bassermann F *et al.* (2019) A multicentre, phase IIa study of zolbetuximab as a single agent in patients with recurrent or refractory advanced adenocarcinoma of the stomach or lower oesophagus: the MONO study. *Ann Oncol* **30**, 1487–1495.

61 Jin D-H, Park S-E, Lee J, Kim K-M, Kim S, Kim D-H & Park J (2015) Copy number gains at 8q24 and 20q11-q13 in gastric cancer are more common in intestinal-type than diffuse-type. *PLoS One* **10**, e0137657.

62 Chandrashekar P, Ahmadinejad N, Wang J, Sekulic A, Egan JB, Asmann YW, Kumar S, Maley C & Liu L (2019) Somatic selection distinguishes oncogenes and tumor suppressor genes. *Bioinformatics* **36**, 1712–1717.

63 Kandoth C, McLellan MD, Vandin F, Ye K, Niu B, Lu C, Xie M, Zhang Q, McMichael JF, Wyczalkowski MA *et al.* (2013) Mutational landscape and significance across 12 major cancer types. *Nature* **502**, 333–339.

64 Ajani JA, Abramov M, Bondarenko I, Shparyk Y, Gorbunova V, Hontsa A, Otchenash N, Alsina M, Lazarev S, Feliu J *et al.* (2017) A phase III trial

18780261, 2021, 11, Downloaded from https://febs.onlinelibrary.wiley.com/doi/10.1002/1878-0261.12948 by Cochrane Portugal, Wiley Online Library on [31/01/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

comparing oral S-1/cisplatin and intravenous 5-fluorouracil/cisplatin in patients with untreated diffuse gastric cancer. *Ann Oncol* **28**, 2142–2148.
65 Piessen G, Messager M, Le Malicot K, Robb WB, Di Fiore F, Guilbert M, Moreau M, Christophe V, Adenis A & Mariette C (2013) Phase II/III multicentre randomised controlled trial evaluating a strategy of primary surgery and adjuvant chemotherapy versus peri-operative chemotherapy for resectable gastric signet ring cell adenocarcinomas - PRODIGE 19 - FFCD1103 - ADCI002. *BMC Cancer* **13**, 281.
66 Guilford P, Hopkins J, Harraway J, McLeod M, McLeod N, Harawira P, Taite H, Scoular R, Miller A & Reeve AE (1998) E-cadherin germline mutations in familial gastric cancer. *Nature* **392**, 402–405.
67 Majewski IJ, Kluijt I, Cats A, Scerri TS, de Jong D, Kluin RJ, Hansford S, Hogervorst FB, Bosma AJ, Hofland I *et al.* (2013) An α-E-catenin (CTNNA1) mutation in hereditary diffuse gastric cancer. *J Pathol* **229**, 621–629.
68 Blair VR, McLeod M, Carneiro F, Coit DG, D'Addario JL, van Dieren JM, Harris KL, Hoogerbrugge N, Oliveira C, van der Post RS *et al.* (2020) Hereditary diffuse gastric cancer: updated clinical practice guidelines. *Lancet Oncol* **21**, e386–e397.
69 Funakoshi T, Miyamoto S, Kakiuchi N, Nikaido M, Setoyama T, Yokoyama A, Horimatsu T, Yamada A, Torishima M, Kosugi S(2019) Genetic analysis of a case of Helicobacter pylori-uninfected intramucosal gastric cancer in a family with hereditary diffuse gastric cancer. *Gastric Cancer* **22**, 892–898.
70 Malone ER, Oliva M, Sabatini PJB, Stockley TL & Siu LL (2020) Molecular profiling for precision cancer therapies. *Genome Med* **12**, 8.
71 Ishikawa M, Hayashi H, Sakamoto N, Tanaka S & Nishihara H (2019) Cancer gene profiling explores the possible precision medicine for diffuse-type gastric adenocarcinoma. *Med Oncol* **37**, 10.
72 Christgen M & Derksen P (2015) Lobular breast cancer: molecular basis, mouse and cellular models. *Breast Cancer Res* **17**, 16.
73 Vaillant F, Merino D, Lee L, Breslin K, Pal B, Ritchie ME, Smyth GK, Christie M, Phillipson LJ, Burns CJ *et al.* (2013) Targeting BCL-2 with the BH3 mimetic ABT-199 in estrogen receptor-positive breast cancer. *Cancer Cell* **24**, 120–129.
74 Oakes SR, Vaillant F, Lim E, Lee L, Breslin K, Feleppa F, Deb S, Ritchie ME, Takano E, Ward T *et al.* (2011) Sensitization of BCL-2–expressing breast tumors to chemotherapy by the BH3 mimetic ABT-737, *Proc Natl Acad Sci USA* **109**, 2766–2771.
75 Ying H, Biroc SL, Li W-W, Alicke B, Xuan J-A, Pagila R, Ohashi Y, Okada T, Kamata Y & Dinter H (2006) The Rho kinase inhibitor fasudil inhibits tumor progression in human and rat tumor models. *Mole Cancer Therap* **5**, 2158–2164.
76 Wang H, Shen L, Li Y & Lv J (2020) Integrated characterisation of cancer genes identifies key molecular biomarkers in stomach adenocarcinoma. *J Clin Pathol* **73**, 579–586.
77 Yang Y, Zhang J, Chen Y, Xu R, Zhao Q & Guo W (2020) MUC4, MUC16, and TTN genes mutation correlated with prognosis, and predicted tumor mutation burden and immunotherapy efficacy in gastric cancer and pan-cancer. *Clin Transl Med* **10**, e155.
78 Luo T, Zhao J, Lu Z, Bi J, Pang T, Cui H, Yang B, Li W, Wang Y, Wu S *et al.* (2018) Characterization of long non-coding RNAs and MEF2C-AS1 identified as a novel biomarker in diffuse gastric cancer. *Transl Oncol* **11**, 1080–1089.
79 Cho S-J, Yoon C, Lee JH, Chang KK, Lin J-X, Kim Y-H, Kook M-C, Aksoy BA, Park DJ, Ashktorab H *et al.* (2018) *KMT2C* mutations in diffuse-type gastric adenocarcinoma promote epithelial-to-mesenchymal transition. *Clin Cancer Res* **24**, 6556–6569.
80 Ho SWT & Tan P (2019) Dissection of gastric cancer heterogeneity for precision oncology. *Cancer Sci* **110**, 3405–3414.
81 Cunningham SC, Kamangar F, Kim MP, Hammoud S, Haque R, Iacobuzio-Donahue C, Ashfaq R, Kern SE, Maitra A, Heitmiller RE *et al.* (2006) MKK4 status predicts survival after resection of gastric adenocarcinoma. *Arch Surg* **141**, 1095–1099.discussion 1100.
82 Ogiwara H, Sasaki M, Mitachi T, Oike T, Higuchi S, Tominaga Y & Kohno T (2016) Targeting p300 addiction in CBP-deficient cancers causes synthetic lethality by apoptotic cell death due to abrogation of MYC expression. *Cancer Discov* **6**, 430–445.
83 Garattini SK, Basile D, Cattaneo M, Fanotto V, Ongaro E, Bonotto M, Negri FV, Berenato R, Ermacora P, Cardellino GG *et al.* (2017) Molecular classifications of gastric cancers: Novel insights and possible future applications. *World J Gastrointest Oncol* **9**, 194–208.
84 Sundar R, Huang KK, Qamra A, Kim KM, Kim ST, Kang WK, Tan ALK, Lee J & Tan P (2019) Epigenomic promoter alterations predict for benefit from immune checkpoint inhibition in metastatic gastric cancer. *Ann Oncol* **30**, 424–430.
85 Refolo MG, Lotesoriere C, Messa C, Caruso MG & D'Alessandro R (2020) Integrated immune gene expression signature and molecular classification in gastric cancer: new insights. *J Leukoc Biol* **108**, 633–646.
86 Orditura M, Galizia G, Sforza V, Gambardella V, Fabozzi A, Laterza MM, Andreozzi F, Ventriglia J, Savastano B, Mabilia A *et al.* (2014) Treatment of gastric cancer. *World J Gastroenterol* **20**, 1635–1649.
87 Smyth EC, Verheij M, Allum W, Cunningham D, Cervantes A & Arnold D (2016) Gastric cancer: ESMO

18780261, 2021, 11, Downloaded from https://febs.onlinelibrary.wiley.com/doi/10.1002/1878-0261.12948 by Cochrane Portugal, Wiley Online Library on [31/01/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

Clinical Practice Guidelines for diagnosis, treatment and follow-up. *Ann Oncol* **27**, v38–v49.
88 Wilke H, Muro K, Van Cutsem E, Oh SC, Bodoky G, Shimada Y, Hironaka S, Sugimoto N, Lipatov O, Kim TY *et al.* (2014) Ramucirumab plus paclitaxel versus placebo plus paclitaxel in patients with previously treated advanced gastric or gastro-oesophageal junction adenocarcinoma (RAINBOW): a double-blind, randomised phase 3 trial. *Lancet Oncol* **15**, 1224–1235.
89 Li J, Qin S, Xu J, Guo W, Xiong J, Bai Y, Sun G, Yang Y, Wang L, Xu N *et al.* (2013) Apatinib for chemotherapy-refractory advanced metastatic gastric cancer: results from a randomized, placebo-controlled, parallel-arm, phase II trial. *J Clin Oncol* **31**, 3219–3225.
90 Cartwright E, Athauda A & Chau I (2020) Emerging precision therapies for gastric cancer. *Exp Rev Precis Med Drug Dev* **5**, 299–311.
91 Riquelme I, Saavedra K, Espinoza JA, Weber H, García P, Nervi B, Garrido M, Corvalán AH, Roa JC & Bizama C (2015) Molecular classification of gastric cancer: towards a pathway-driven targeted therapy. *Oncotarget* **6**, 24750–24779.
92 Tirino G, Pompella L, Petrillo A, Laterza MM, Pappalardo A, Caterino M, Orditura M, Ciardiello F, Galizia G & De Vita F (2018) What's New in Gastric Cancer: the therapeutic implications of molecular classifications and future perspectives. *Int J Mol Sci* **19**, 2659.
93 Ignatova E, Seriak D, Fedyanin M, Tryakin A, Pokataev I, Menshikova S, Vakhabova Y, Smirnova K, Tjulandin S & Ajani JA (2020) Epstein-Barr virus-associated gastric cancer: disease that requires special approach. *Gastric Cancer* **23**, 951–960.

Supporting information

Additional supporting information may be found online in the Supporting Information section at the end of the article.

Data S1. Full data on gene mutation frequencies obtained from the selected studies (sheet 1), reporting all mutated genes found for DGC cases (sheet 2), IGC cases (sheet 3) and SRCC/PCC-NOS (sheet 4).

Data S2. GO TERMS, Biological Processes and Molecular Functions retrieved by Enrichr for each genes list: Transversally + Specific DGC/IGC, Specific DGC/IGC genes, Transversally + Enriched + Specific SRCC genes, Specific SRCC genes, Transversally + Enriched + Specific PCC-NOS genes, and Specific PCC-NOS genes.

Table S1. Sample size and details of sequencing methodologies used in the selected publications for analysis.

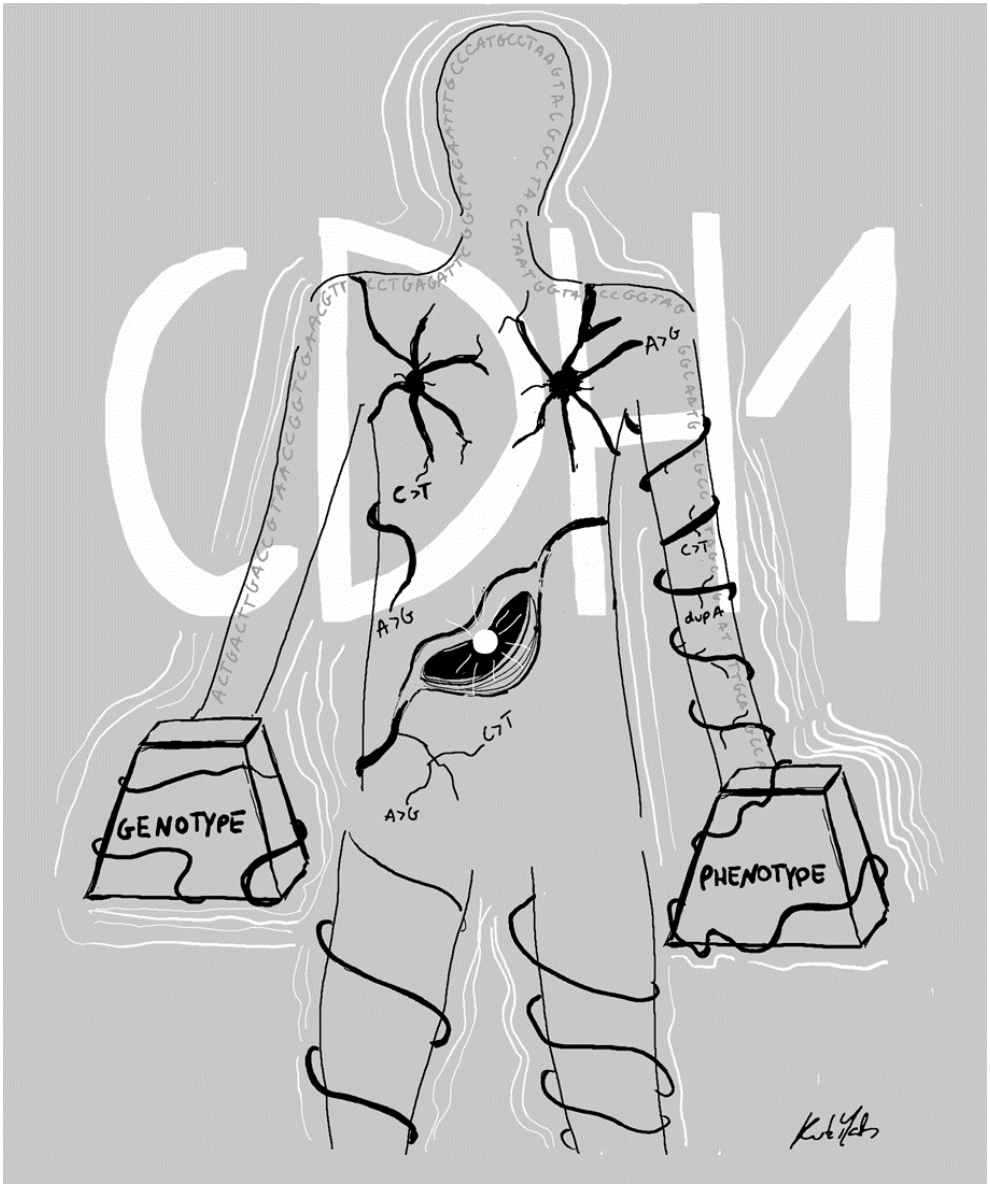
Table S2. Details of Clinical trial registered at ClinicalTrials.gov directed to gastric cancer and often including DGC patients.

18780261, 2021, 11, Downloaded from https://febs.onlinelibrary.wiley.com/doi/10.1002/1878-0261.12948 by Cochrane Portugal, Wiley Online Library on [31/01/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

Supplementary Material: Available through the link:

https://febs.onlinelibrary.wiley.com/doi/full/10.1002/1878-0261.12948

2. *CDH1* germline alterations and phenotypes associated with DGLBCS

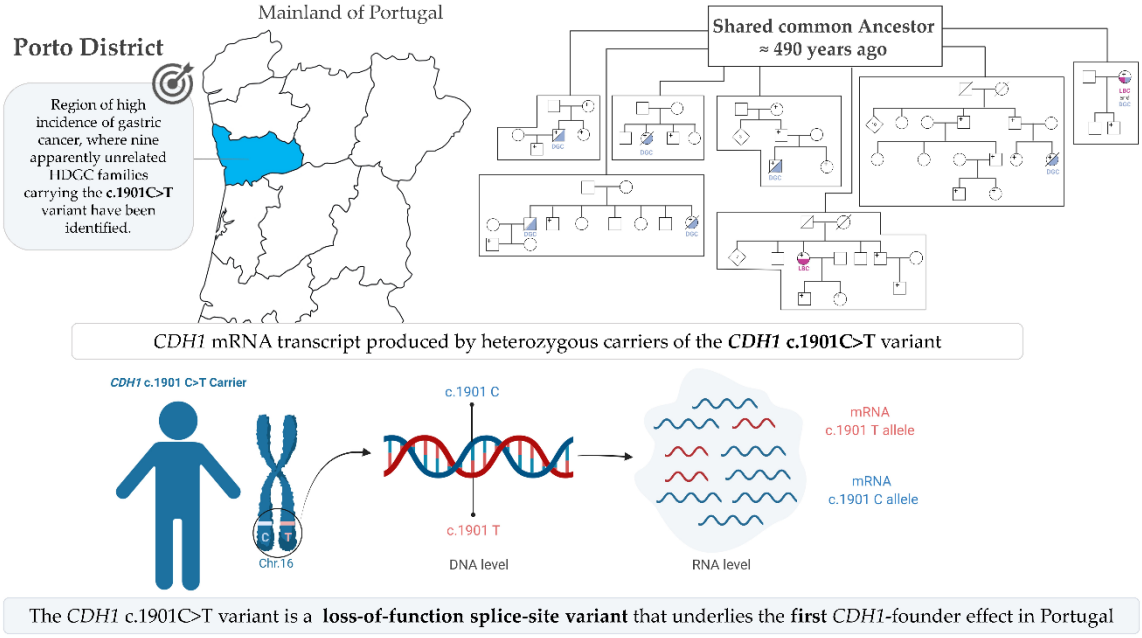


Paper 3: The *CDH1* c.1901C>T variant: a founder variant in the Portuguese population with severe impact in mRNA splicing

Paper 4: 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

2.1. Paper 3, Chapter III

The *CDH1* c.1901C>T variant: a founder variant in the Portuguese population with severe impact in mRNA splicing



Author statement:

Rita Barbosa-Matos contributed to the conceptualization, investigation, data collection, methodology, figure design, writing (original draft), reviewing, and editing.

This work identifies that the *CDH1* c.1901C>T variant present in several families from the Portuguese population, has a significant impact on mRNA transcription, causing the loss-of-function of the mutated allele. Therefore, this work contributed to the understanding of its pathogenicity and clinical implications since this variant, previously classified as missense, is herein defined as truncating. Furthermore, the research explains the unusually high prevalence of DGC cases in unrelated families due to the presence of a *CDH1* founder variant, providing valuable insights into the importance of population-specific founder variants in clinical practice and genetic counseling.

cancers

MDPI

Article

The *CDH1* c.1901C>T Variant: A Founder Variant in the Portuguese Population with Severe Impact in mRNA Splicing

Rita Barbosa-Matos [1,2,3,†], Rafaela Leal Silva [1,2,4,†], Luzia Garrido [5,†], Ana Cerqueira Aguiar [1,2,6], José Garcia-Pelaez [1,2,7], Ana André [1,2], Susana Seixas [1,2], Sónia Passos Sousa [1,2], Luísa Ferro [5], Lúcia Vilarinho [5], Irene Gullo [1,2,5,8], Vitor Devezas [5], Renata Oliveira [5], Susana Fernandes [8], Susy Cabral Costa [5,8], André Magalhães [5], Manuela Baptista [5], Fátima Carneiro [1,2,5,8], Hugo Pinheiro [1,2,9], Sérgio Castedo [1,2,5,8] and Carla Oliveira [1,2,8,*]

1 i3S, Instituto de Investigação e Inovação em Saúde, Rua Alfredo Allen 208, 4200-135 Porto, Portugal; amatos@ipatimup.pt (R.B.-M.); slealrafaela@gmail.com (R.L.S.); ana.c.aguiar@chedv.min-saude.pt (A.C.A.); jpelaez@ipatimup.pt (J.G.-P.); aandre@ipatimup.pt (A.A.); sseixas@ipatimup.pt (S.S.); ssousa@ipatimup.pt (S.P.S.); irene.gullo12@gmail.com (I.G.); fcarneiro@ipatimup.pt (F.C.); hpinheiro@ipatimup.pt (H.P.); scastedo@ipatimup.pt (S.C.)
2 Ipatimup, Instituto de Patologia e Imunologia Molecular da Universidade do Porto, Rua Júlio Amaral de Carvalho 45, 4200-804 Porto, Portugal
3 Doctoral Programme on Cellular and Molecular Biotechnology Applied to Health Sciences (BiotechHealth), ICBAS-Institute of Biomedical Sciences Abel Salazar, University of Porto, 4050-313 Porto, Portugal
4 PhD Programme on Health Data Science, Faculty of Medicine, University of Porto, 4200-319 Porto, Portugal
5 CHUSJ, Centro Hospitar e Universitário de São João, Alameda Prof. Hernâni Monteiro, 4200-319 Porto, Portugal; luzia.garrido@chsj.min-saude.pt (L.G.); luisa.ferro@chsj.min-saude.pt (L.F.); lucia.vilarinho@chsj.min-saude.pt (L.V.); vitor.devezas7@gmail.com (V.D.); renata.oliveira@chsj.min-saude.pt (R.O.); susycosta@sapo.pt (S.C.C.); amag1976@gmail.com (A.M.); manuela.batista@chsj.min-saude.pt (M.B.)
6 Department of Clinical Pathology, Centro Hospitalar Entre Douro e Vouga, Santa Maria da Feira, 4520-211 Santa Maria da Feira, Portugal
7 Doctoral Programme on Biomedicine, Faculty of Medicine, University of Porto, 4200-319 Porto, Portugal
8 Faculty of Medicine of the University of Porto, Alameda Prof. Hernâni Monteiro, 4200-319 Porto, Portugal; sf@med.up.pt
9 Department of Internal Medicine, Centro Hospitalar Tâmega e Sousa, 4564-007 Penafiel, Portugal
* Correspondence: carlaol@i3s.up.pt; Tel.: +351-220408800; Fax: +351-225570799
† The authors have equal contributions.

check for updates

Citation: 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.; et al. The *CDH1* c.1901C>T Variant: A Founder Variant in the Portuguese Population with Severe Impact in mRNA Splicing. *Cancers* **2021**, *13*, 4464. https://doi.org/10.3390/cancers13174464

Academic Editor: Thomas P. Conrads

Received: 30 July 2021
Accepted: 30 August 2021
Published: 4 September 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/).

Simple Summary: An unexpectedly high number of early-onset diffuse gastric and lobular breast cancer in apparently unrelated families carrying the same *CDH1* c.1901C>T variant (formerly known as missense p.A634V) in Northern Portugal suggested a founder effect in this region. We demonstrated that c.1901C>T is a truncating variant triggered by cryptic splicing, calculated its mutational age, and characterized the tumour spectrum and age of onset in affected families.

Abstract: Hereditary diffuse gastric cancer (HDGC) caused by *CDH1* variants predisposes to early-onset diffuse gastric (DGC) and lobular breast cancer (LBC). In Northern Portugal, the unusually high number of HDGC cases in unrelated families carrying the c.1901C>T variant (formerly known as p.A634V) suggested this as a *CDH1*-founder variant. We aimed to demonstrate that c.1901C>T is a bona fide truncating variant inducing cryptic splicing, to calculate the timing of a potential founder effect, and to characterize tumour spectrum and age of onset in carrying families. The impact in splicing was proven by using carriers' RNA for PCR-cloning sequencing and allelic expression imbalance analysis with SNaPshot. Carriers and noncarriers were haplotyped for 12 polymorphic markers, and the decay of haplotype sharing (DHS) method was used to estimate the time to the most common ancestor of c.1901C>T. Clinical information from 58 carriers was collected and analysed. We validated the cryptic splice site within *CDH1*-exon 12, which was preferred over the canonical one in 100% of sequenced clones. Cryptic splicing induced an out-of-frame 37bp deletion in exon 12, premature truncation (*p.Ala634ProfsTer7*), and consequently RNA mediated decay. The haplotypes carrying the c.1901C>T variant were found to share a common ancestral estimated at 490 years

(95% Confidence Interval 445–10,900). Among 58 carriers (27 males (M)–31 females (F); 13–83 years), DGC occurred in 11 (18.9%; 4M–7F; average age 33 ± 12) and LBC in 6 females (19.4%; average age 50 ± 8). Herein, we demonstrated that the c.1901C>T variant is a loss-of-function splice-site variant that underlies the first *CDH1*-founder effect in Portugal. Knowledge on this founder effect will drive genetic testing of this specific variant in HDGC families in this geographical region and allow intrafamilial penetrance analysis and better estimation of variant-associated tumour risks, disease age of onset, and spectrum.

Keywords: HDGC; *CDH1*; founder effect; missense variants; cryptic splicing

1. Introduction

Evidence for familial predisposition to gastric cancer (GC) is present in approximately 10% of all GC cases, with 1–3% proven to be hereditary [1]. The first heritable syndrome recognized as linked to GC predisposition is known as hereditary diffuse gastric cancer (HDGC) [2]. This autosomal dominant tumour risk syndrome predisposes to diffuse gastric cancer (DGC) in both sexes and to lobular breast cancer (LBC) in females [3]. The definition of HDGC evolved along the years, as well as the clinical criteria driving genetic testing. Currently, *CDH1* germline variants, namely single nucleotide variants and large deletions, are its best characterized predisposing factor [2,4,5].

Clinical practice guidelines were defined to identify at-risk individuals for DGC and/or LBC, and were recently updated in 2020 (Supplemental Table S1) [5]. The present genetic screening criteria are less restrictive than the previous ones regarding age limits and consider personal or family history of cleft lip and/or cleft palate as evidence if co-occurring with DGC [3,5–7].

The disease risk associated with causal *CDH1* germline variants considerably varies, depending on the selected study, which exposes the limited understanding of disease penetrance in HDGC families [8]. Several factors can influence risk calculations, namely family ascertainment, the type of *CDH1* variants analysed, and the GC incidence in the home country of these families [5]. The bias in most reported studies has mainly reflected the inclusion of families that comply with clinical criteria and present high prevalence of DGC and/or LBC [8–11]. When considering both unascertained and HDGC families together, predicted disease risks decrease, as demonstrated by Roberts et al [5,8].

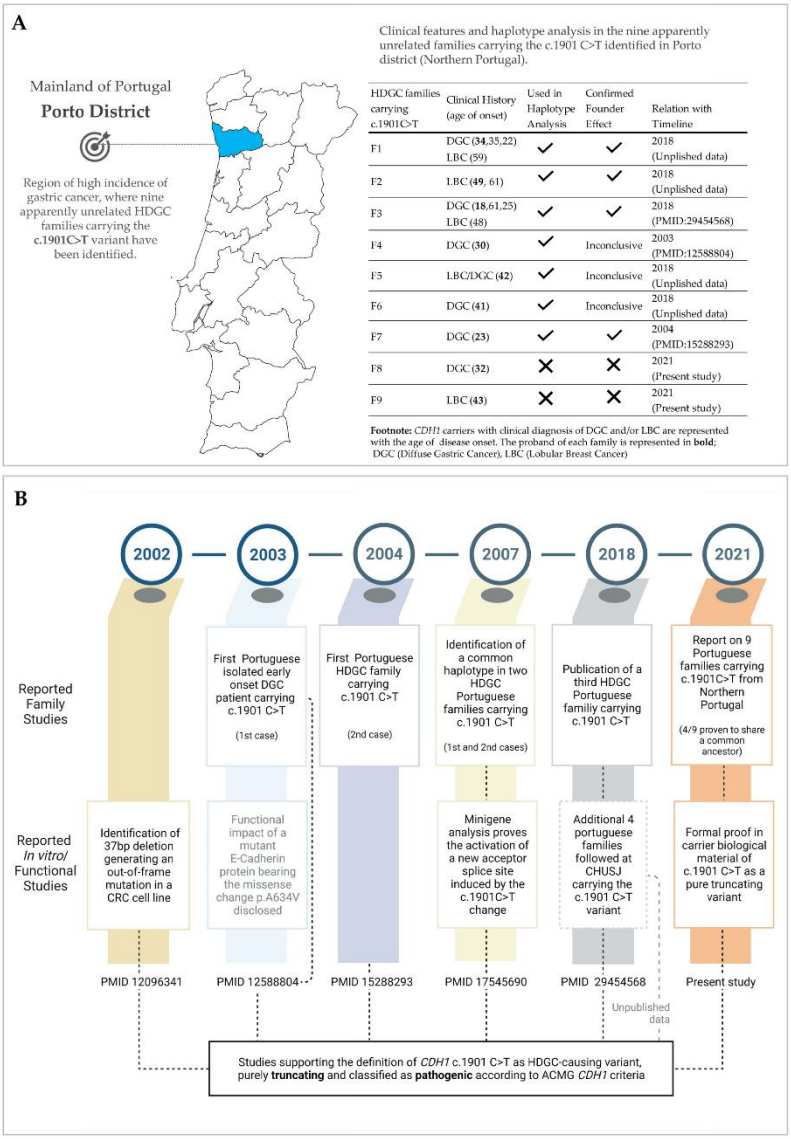
According to the ACMG/AMP criteria, only *CDH1* variants classified as pathogenic or likely pathogenic are considered actionable in the HDGC context [5,12]. For management of HDGC patients carrying actionable *CDH1* variants, prophylactic removal of the whole stomach to prevent DGC is recommended. The risk for LBC can be managed through bilateral risk-reducing mastectomy [5]. *CDH1* variant carriers may also undergo DGC surveillance by annual endoscopy with multiple biopsies and/or annual breast magnetic resonance imaging (MRI) if unfit or if refusing prophylactic measures [5].

Variants of unknown significance are a challenge for HDGC management, and *CDH1* missense variants are a particular hurdle. Despite being frequently detected in families with HDGC criteria, their actionability and ability to predispose to disease is hard to predict [12,13]. Some *CDH1* variants, initially considered to be missense, were later proven be protein-truncating, with clear functional consequences due to cryptic splicing activation, and therefore suggested as pathogenic or likely pathogenic [12,14]. This evidence reflects the importance of deepening the study of missense variants, particularly their potential to induce cryptic splicing and promote premature truncation. This will improve variant classification and have a clear impact in prediction of variant actionability.

In certain regions of Northern Portugal, more specifically in the district of Porto (an administrative division of the Portuguese territory currently subdivided in 18 counties—see map in Figure 1A, the economic growth, infrastructure development, and industrial expansion achieved in the last century have led to a recent populational increment and

the settlement of large communities in this area, which established therein businesses and families. Several Porto counties, namely those located on the northeastern side, are former rural regions where local populations were originally reduced in size and geographically isolated. These features might have led to significant shifts in variant frequency (genetic drift). If local populations were originally reduced in size and isolated, these features might have led to significant shifts in variant frequency, predisposing to a phenomenon known as a founder effect. By definition, this phenomenon is associated with a reduction in genetic diversity where colonization occurs associated with a population decrease, migration, or isolation [15]. Therefore, a founder effect might explain the clustering, in the district of Porto, of several apparently unrelated families carrying the *CDH1* c.1901C>T variant (all living in neighbouring parishes located in the northeast of Porto in Northern Portugal). This explanation would be particularly compelling if the variant's mutation age precedes the timing of the expansion of this population (the end of the 19th century to the beginning of the 21st century). However, if this hypothesis holds true, it is important to highlight that the recent population rise could have also promoted the spread and frequency increase of a previously rare and deleterious variant [16,17]. The high incidence of GC in Northern Portugal [18,19] and the privileged industrial and economic situation of this region, which allowed inhabitants to settle and succeed, further support a possible clustering of *CDH1* germline variant carriers predisposed to early-onset DGC and/or LBC in this region [14,20].

The c.1901C>T variant, located in *CDH1* exon 12 (Chr16: 68822190, GRCh38) was first described by Vécsey-Semjén et al. as a somatic out-of-frame mutation identified in a colon cancer cell line [21]. Previous in silico predictions and in vitro studies (cell lines and minigene assays) reported the c.1901C>T change to create cryptic-splicing and premature truncation [14,21] (Figure 1B). However, the outcome of the missense variant (p.A634V) overexpression in *CDH1*-negative cell lines is still widely used to support its deleteriousness [22]. To clarify this, we aimed at demonstrating, by using RNA obtained from HDGC germline carriers, that the *CDH1* c.1901C>T variant produces exclusively premature truncation codon (PTC)-bearing transcripts. This provides additional evidence to support the classification of the c.1901C>T variant as pathogenic (in ClinVar, currently classified as pathogenic/likely pathogenic—VCV000012244.5). This knowledge will improve clinical management provided to carriers of this variant and their families. Herein, by following families and individuals carrying the *CDH1* c.1901C>T variant, we characterized the first *CDH1*-related founder effect in Northern Portugal, traced its origin, and portraited tumour spectrum and age of onset in carrying families.



Clinical features and haplotype analysis in the nine apparently unrelated families carrying the c.1901 C>T identified in Porto district (Northern Portugal).

HDGC families carrying c.1901C>T	Clinical History (age of onset)	Used in Haplotype Analysis	Confirmed Founder Effect	Relation with Timeline
F1	DGC (**34**,35,22) LBC (59)	✓	✓	2018 (Unplished data)
F2	LBC (**49**, 61)	✓	✓	2018 (Unplished data)
F3	DGC (**18**,61,25) LBC (48)	✓	✓	2018 (PMID:29454568)
F4	DGC (**30**)	✓	Inconclusive	2003 (PMID:12588804)
F5	LBC/DGC (**42**)	✓	Inconclusive	2018 (Unplished data)
F6	DGC (**41**)	✓	Inconclusive	2018 (Unplished data)
F7	DGC (**23**)	✓	✓	2004 (PMID:15288293)
F8	DGC (**32**)	✗	✗	2021 (Present study)
F9	LBC (**43**)	✗	✗	2021 (Present study)

Footnote: *CDH1* carriers with clinical diagnosis of DGC and/or LBC are represented with the age of disease onset. The proband of each family is represented in **bold**; DGC (Diffuse Gastric Cancer), LBC (Lobular Breast Cancer)

Figure 1. Timeline of the discovery of a *CDH1*-related founder effect in Northern Portugal: the *CDH1* c.1901C>T variant. (**A**) Map of the mainland of Portugal highlighting the Porto district in Northern Portugal, a region of high incidence of GC where the nine c.1901C>T variant carriers' families were identified; and clinical features of carrier families (right hand side table). (**B**) Timeline of discoveries related to the c.1901C>T *CDH1* variant: from pathogenicity to carrier families. CRC, colorectal carcinoma; DGC, diffuse gastric cancer; HDGC, hereditary diffuse gastric cancer; CHUSJ, Centro Hospitalar Universitário de S. João; ACMG/AMP, American College of Medical Genetics and Genomics and the Association for Molecular Pathology *CDH1* criteria [12].

2. Materials and Methods

2.1. Study Design and Data Collection

The collection of clinical data and blood samples from patients and families was approved by the appropriate Ethics Committee from each of the centres participating in this work, i3S/IPATIMUP and Centro Hospital Universitário de São João (CHUSJ) in Porto, Portugal. Family histories and clinical data were obtained with informed consents and herein pseudoanonymized.

2.2. Biological Material (DNA and RNA) Collection from CDH1 c.1901C>T Variant Bearing Families

Genomic DNA was extracted from 200 µL whole blood using NZY Tissue gDNA Isolation Kit (NZYTech, Lisboa, Portugal) according to the manufacturer's instructions. The fragment containing the *CDH1* c.1901C>T variant was amplified by standard PCR (NZYTaq DNA polymerase, NZYTech, Lisboa, Portugal) with the following primers: forward: 5′-CACCCGGTTCCATCTACCTTT and reverse: 5′-CCTTTCCAACCCCTCCCTACT at an annealing temperature of 60 °C. Direct sequencing of the PCR products was performed by direct Sanger sequencing (BigDye® Terminator v3.1 Cycle Sequencing Kit, ThermoFisher, Waltham, MA, USA) using an automated sequencer (ABI PRISM® 3130 Genetic Analyzer, Applied Biosystems).

RNA was isolated from peripheral blood mononuclear cells (PBMCs). PBMCs were isolated using Ficoll® Paque Plus (Sigma-Aldrich, St. Louis, MO, USA) according to the manufacturer's instructions. The pellet was resuspended in 300 µL of Lysis/Binding Buffer from the kit mirVana™ miRNA Isolation Kit (Life Technologies, Carlsbad, CA, USA). This kit was used for total RNA isolation according to the manufacturer's instructions. RNA concentration was measured with Qubit™ RNA HS Assay Kit (Life Technologies, Carlsbad, CA, USA).

2.3. cDNA Synthesis and PCR Amplification

Reverse transcriptase (RT)-PCR was performed with 10 ng of input RNA using the SuperScript™ VILO™ cDNA Synthesis Kit (Life Technologies, Carlsbad, CA, USA). The reaction was performed using the following conditions: 5 × RT reaction mix, 10 × enzyme mix, 10 ng of RNA in a final volume of 10 µL, for 30 min at 42 °C and 5 min at 85 °C. The sequence between *CDH1* exon 11 and 13 was amplified with Qiagen® Multiplex PCR Kit (Qiagen, Hilden, Germany) according to the manufacturer's instructions and with the primers and conditions mentioned in Supplemental Table S2. Amplification was confirmed by electrophoresis in a 2% agarose gel, and PCR products were purified with the GE Kit according to "Protocol for purification of DNA from solution or an enzymatic reaction" and eluted in a final volume of 20 µL of DNase/RNase free water (Invitrogen, Carlsbad, CA, USA). The DNA concentration of the purified product was measured with NanoDrop 2000c Spectrophotometer (Thermo Scientific, Wilmington, NC, USA).

2.4. Cloning and Colony PCR

PCR products were cloned into pCR® 2.1-TOPO® vector, TOPO® TA Cloning® Kit, (Invitrogen, Carlsbad, CA, USA) according to the manufacturer's instructions and using Mach1-T1 competent bacteria. The colony PCR with Universal M13 primers was performed in 80 colonies with Qiagen® Multiplex PCR Kit (Qiagen, Hilden, Germany) according to the manufacturer's instructions. M13 forward and reverse primers were provided by the TOPO® TA Cloning® Kit, (Invitrogene, Carlsbad, CA, USA) and used in the reaction at 0.4 µm (primers and PCR conditions are described in Supplemental Table S2). Amplification was confirmed by electrophoresis of the products in a 2% agarose gel, and the selected amplified products were sequenced using the primers and conditions described in Supplemental Table S2.

2.5. CDH1 Allelic Specific Expression (ASE) by SNaPshot

Confirmation of the heterozygous state at the single nucleotide polymorphism (SNP) rs1801552 was performed with polymerase chain reaction (PCR) sequencing in germline DNA. cDNA and gDNA from rs1801552 C/T F2 (family proband) were amplified with SNP flanking primers. All primers and PCR conditions are available in Supplemental Table S2. Purified PCR products were submitted to SNaPshot reaction using SNaPshot Multiplex Kit (Applied Biosystems, Foster City, CA, USA) and a single-base-extension (SBE)-specific primer following manufacturer's instructions. Final products were purified with FastAP™ Thermosensitive Alkaline Phosphatase (Thermo Scientific, Wilmington, NC, USA) and run in the ABI PRISM® 3130xl Genetic Analyzer (Applied Biosystems, Foster City, CA, USA). Peak Scanner™ software (ThermoFisher, Waltham, MA, USA) was used for peak analysis.

2.6. Histopathological Analysis

Four-micrometre sections were used for haematoxylin and eosin (HE) staining and immunohistochemistry (IHC). IHC for E-cadherin (clone 4A2C7, ThermoFisher, Waltham, MA, USA) was performed with Ventana BenchMark XT (Ventana Medical Systems, AZ, USA) automated immunostainer according to the manufacturer's guidelines. Tissue sections were deparaffinized and rehydrated. After antigen retrieval, sections were incubated with a primary antibody against E-cadherin, and 3,3′-diaminobenzidine (DAB) was used as a chromogen. Finally, the slides were counterstained with haematoxylin, and coverslips were placed.

2.7. CDH1 c.1901C>T Variant Classification According to ACMG Guidelines

The actionability of *CDH1* c.1901C>T variant in the context of HDGC was evaluated using the ACMG/AMP Variant Curation Guidelines for the Analysis of Germline *CDH1* Sequence Variants [12]. These rules consider the variant's impact at the RNA and protein level, relevant in vitro studies, frequency in families fulfilling HDGC criteria, the presence of *de novo* variants in LBC/DGC cases, and frequency in control populations. This allows classifying *CDH1* variants into one of five categories: pathogenic and likely pathogenic (actionable in HDGC), VUS (variant unknown significance), likely benign, and benign.

2.8. Identification of the Recombination Point in 16q22.1 Chromosome's Region, Haplotyping of Different Polymorphic Markers

Thirty-eight DNA samples provided by CHUSJ were PCR amplified for the polymorphic markers (upstream and downstream of *CDH1*). Primers used and expected amplicons are described in Supplemental Table S3. For microsatellite analyses, 1 µL PCR products were mixed with 9.5 µL Hi-Di™ Formamide (Applied Biosystems, Foster City, CA, USA) and 0.5 µL of molecular size standard dye GeneScan™ 120 LIZ™ or GeneScan™ 500 ROX™ (Applied Biosystems, Foster City, CA, USA) depending on the molecular size of the amplified fragment. The mixture was run in BI PRISM® 3130xl Genetic Analyzer (Applied Biosystems, Foster City, CA, USA) and analysed with Peak Scanner™ software (ThermoFisher Waltham, MA, USA). For SNP analysis, the PCR products were purified by FastAP™ thermosensitive alkaline phosphatase and exonuclease I (Thermo Scientific, Wilmington, NC, USA), following the manufacturer's instructions, and used for Sanger sequencing with the BigDye® Terminator Cycle v3.1 (Applied Biosystems, Foster City, CA, USA), according to the manufacturer's instructions. The purified products were run in ABI PRISM® 3130xl Genetic Analyzer (Applied Biosystems, Foster City, CA, USA), and the analysis was performed using Mutation Surveyor® software (SoftGenetics, State College, PA, USA).

2.9. Characterization and Estimation of Age of the CDH1 c.1901C>T Variant

The time to most recent common ancestor (TMRCA) of *CDH1* c.1901C>T variant was estimated by the Decay of Haplotype Sharing MAPping (DHSMAP) software [23]. Four families carrying the c.1901C>T variant were characterized for 12 polymorphic mark-

ers (2 single nucleotide polymorphisms (SNPs) and 10 short tandem repeats (STRs)) located at different distances from *CDH1* and spanning a region of 19.2 Mb (16:62304172 to 16:81414947 GRCh38-hg38 assembly). All haplotypes were inferred according to their segregation within families (Supplemental Table S4). Since this work was based exclusively on family data, we limited identical copies of the same segregating haplotype to single chromosomes, as previously done [24]. Based on this condition, two different runs were performed with DHSMAP: a more conservative analysis (A) was first assumed, which allowed only the inclusion of single *CDH1* c.1901C>T chromosomes, one per each family (F1, F2, F3, and F7); while a second analysis (B) permitted the inclusion of two derived haplotypes resulting from mutation or recombination events at the D16S514 marker in families F7 (F7-A carrier) and F2 (F2-F carrier), respectively. Moreover, since for these individuals the D16S514 marker allele in phase with the *CDH1* c.1901C>T variant could not be accurately inferred from family segregation, two alternative haplotype configurations were considered in DHSMAP analysis. The genetic distance (23.2 cM for the 19.2 Mb region surrounding *CDH1*) used in DHSMAP for c.1901C>T calculations was extracted from chr.16 recombination maps generated for the Iberian sample of the 1000 Genomes project Phase I (http://www.internationalgenome.org/data#download; NCBI FTP Site: http://ftp-trace.ncbi.nih.gov/1000genomes/ftp/; http://ftp-trace.ncbi.nih.gov/1000genomes/ftp/technical/working/20130507_omni_recombination_rates/, accessed on 4 May 2018). Mutation rates were assumed as 0.001 for microsatellites [25] and 2.5×10^{-8} for SNPs [26]. In total, haplotypes from the 4 mutated chromosomes, 2 combinations of descendent affected haplotypes, and 18 control chromosomes were used for TMRCA estimation.

3. Results

3.1. The CDH1 c.1901C>T Variant Generates Cryptic Splicing within Exon 12, Leading to Premature Truncation and Decreased CDH1 RNA Levels

To evaluate the consequence of the *CDH1* c.1901C>T variant, we isolated RNA from PBMCs from patients F2 and F4-A, both carriers of the variant (methodological strategy depicted in Figure 2A). F2 was a female patient and the Index case of family F2 (partial pedigree—Supplemental Figure S1). She was diagnosed with multiple ($n = 5$) LBC foci at 49 years old and submitted to prophylactic gastrectomy at 51 years old. Histopathological analysis revealed three foci of signet ring cell carcinoma of the stomach (SRCC) invading the lamina propria (pT1a) of the stomach. At 53 years old, she underwent contralateral risk reduction mastectomy, which revealed lobular intraepithelial neoplasia. Individual F4-A is the son of the Index in family F4 (partial pedigree—Supplemental Figure S1). He was submitted to prophylactic total gastrectomy at 19 years old. Histopathological analysis of the gastrectomy specimen revealed 24 SRCC foci.

Colony PCR of cDNA transcripts using primers flanking exons 12 and 13, where both the *CDH1* c.1901C>T variant and rs1801552 SNP lie, showed PCR products with two different sizes in both patients (Figure 2B and Supplemental Figure S2). While most colonies in both patients (86% and 85%) presented a product of the expected size (674 bp), herein called large product or LP, a smaller product was also observed in both patients (14% and 15%), herein called small product or SP (637 bp) (Figure 2B and Supplemental Figure S2). As predicted, sequencing of the LP band revealed a canonical *CDH1* mRNA fragment encompassing a single allele carrying both the wild-type (WT) c.1901C and the frequent SNP rs1801552C. Sequencing of the SP band showed a deletion of 37 bp spanning from the c.1900 position (in exon 12) to the initial nucleotide of exon 13 (Supplemental Figure S3). This result is compatible with the presence of a cryptic splice site at the position c.1900 of *CDH1* mRNA, lying immediately before the c.1901C>T variant position, that is preferred over the canonical splice site if the variant c.1901C>T is present.

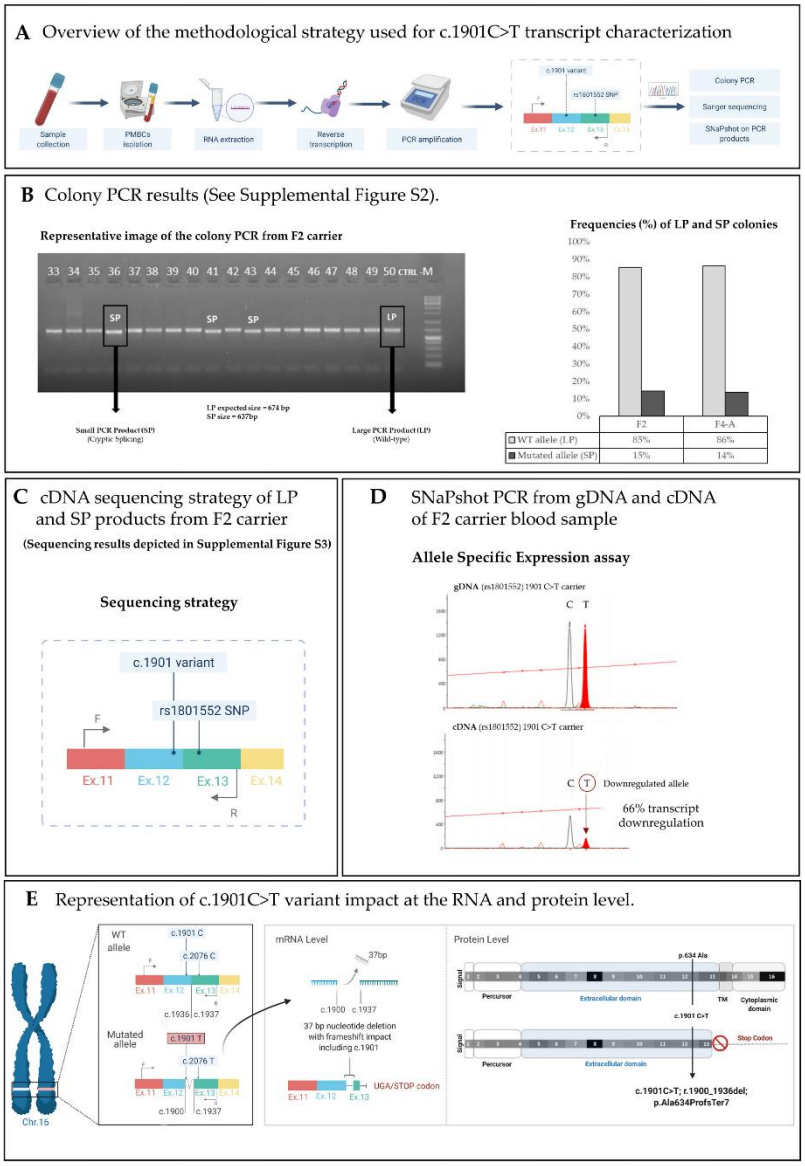


Figure 2. c.1901C>T *CDH1* variant generates a cryptic splice site and a transcript with premature truncation. (**A**) Overview of the methodological strategy used for c.1901C>T transcript characterization; (**B**) Electrophoresis of the colony PCR after cloning cDNA of 2 carriers; (**C**) cDNA sequencing strategy of LP and SP products from F2 carrier; (**D**) SNaPshot PCR from gDNA and cDNA of F2 carrier blood sample; (**E**) representation of c.1901C>T impact at the RNA and protein level. PMBCs, peripheral blood mononuclear cells; gDNA, genomic DNA; cDNA, complementary DNA; WT, wild type.

This cryptic splicing leads to the formation of an out-of-frame transcript that harbours a premature stop codon (PTC) 21 nucleotide after the c.1900 position. Additionally, all sequenced SP colonies displayed a T at SNP rs1801552 (position c.2076 of the *CDH1* coding sequence) (Figure 2C and Supplemental Figure S3). This showed that rs1801552C occurred in the same allele as the WT c.1901C, and that rs1801552T occurred in the same allele as the mutant c.1901T. The colony PCR also showed a lower abundance of the mutant c.1901T allele in comparison with the WT allele, which may suggest downregulation of the PTC-bearing mutant allele by nonsense-mediated mRNA decay. To determine if the two alleles were differentially expressed in variant carriers, we applied a primer extension analysis quantitative method (SNaPshot) in both gDNA and cDNA from PBMCs of F2's proband and used the heterozygous rs1801552 SNP as a marker. In gDNA, both C and T alleles were represented by equivalent peaks showing their heterozygosity. In contrast, in cDNA, the rs1801552C WT allele was consistently over-represented as compared to the rs1801552T mutant allele (Figure 2D). This transcript abundance analysis supported an active degradation of the PTC-bearing allele leading to its reduction by approximately 66%, which is consistent with the truncating nature of the *CDH1* c. 1901C>T variant (Figure 2D). The proposed mechanism is depicted in Figure 2E.

Further supporting the truncating nature of the *CDH1* c.1901C>T variant was the fact that E-cadherin protein expression was absent in both DGC and LBC tumour samples from families F1 and F2, as depicted in Figure 3.

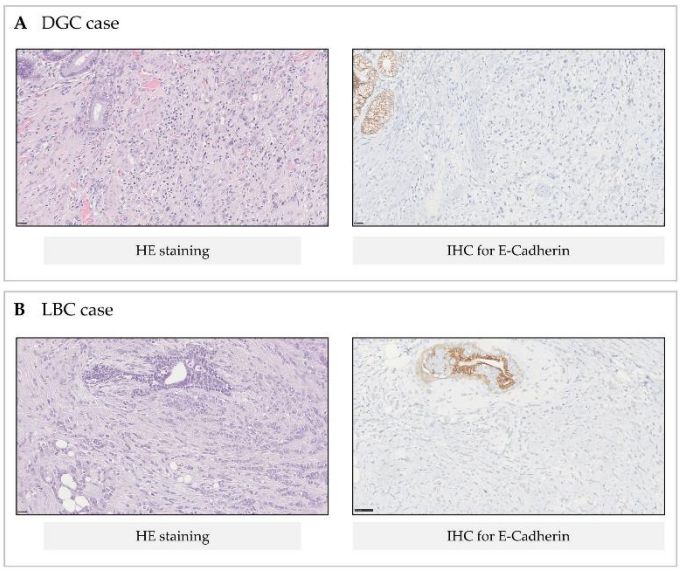


Figure 3. HE (haematoxylin and eosin) and IHC (immunohistochemistry) staining for DGC (F1 carrier) and LBC (F2 carrier) tumours from *CDH1* c.1901C>T variant carriers. (**A**) Poorly cohesive (diffuse) gastric cancer composed by pleomorphic and spindle cells infiltrating the submucosa. Note the presence of residual gastric glands (top right of the image). E-cadherin IHC shows loss of E-cadherin expression in tumour cells and preserved membranous expression in adjacent non-neoplastic epithelium. (**B**) Classic lobular breast cancer composed by uniform cells lacking cohesion, dispersed, and arranged in single-file linear cords. E-cadherin IHC shows loss of E-cadherin expression in tumour cells, while normal membranous expression is retained in non-neoplastic ducts (upper part of the image).

After confirming that *CDH1* c.1901C>T is a bona fide splice-site variant, we used ACMG/AMP guidelines for its classification (12). A PVS1_strong criterium was attributed, as RNA analysis has been previously performed in a cancer cell line (18) supporting our current data. We also considered for classification clinical information from 13 families fulfilling HDGC criteria and carrying this variant (reported to us by different labs), as well as two families fulfilling HDGC criteria and reported in the literature (14, 27). This analysis returned a PS4_strong score. Additionally, this variant was absent in control populations (GnomAD) (PM2_moderate), and it was demonstrated to segregate with DGC and LBC in multiple family members of several families (PP1_strong). Altogether, these criteria supported the classification of the *CDH1* c.1901C>T variant (VCV000012244.5) as pathogenic.

3.2. Age Estimation of CDH1 c.1901C>T Variant through Haplotyping and Identification of the Recombination Point in 16q22.1 Chromosome

From 1999 to 2018, we found seven apparently unrelated families or isolated early-onset DGC cases bearing the *CDH1* c.1901C>T variant, and from 2018 to 2021, we identified two additional HDGC families carrying the same variant (details provided in Figure 1). We could confirm that all nine families were originated from the same region of Northern Portugal (Porto district), although they did not know each other. Given the low likelihood that a very rare variant would have occurred by chance in the same geographical region, we set out to demonstrate that this was a founder effect. To this end, we studied 12 polymorphic markers spanning a region of 19.2 Mbp in four HDGC families with available gDNA (F1, F2, F3, and F7), including affected carriers, asymptomatic carriers, and noncarriers (Figure 4A and Supplemental Figure S1). Haplotype analysis and representative heredograms of two of these families (F1 and F2) are depicted with the results of the sequenced markers (Figure 4B,C). All remaining pedigrees with corresponding haplotypes are depicted in Supplemental Figure S1.

Four *CDH1* c.1901C>T chromosomes were identified (one per family) (Figure 4B). Families F1 and F2 were found to be identical across the entire 19.2 Mb/23.2 cM region studied. The other two chromosomes from F3 and F7 shared the same ancestral haplotype at shorter extensions; in F7, the affected chromosome carried the same marker alleles in a 6.8 Mb/5.1 cM region (D16S514 to D16S3067), and in F3, in a 19.1 Mb/18.4 cM region (D16S318 to D16S3098) region. Notably, in F7 and F2, two events of mutation or recombination were detected as modifying former haplotypes configurations at D16S514 marker (F7-A and F2-F) (Figure 4B,C and Supplemental Table S4).

Three other families (F4, F5, and F6) were also haplotyped, but only for 6/12 polymorphic markers because of a lack of material and a limited number of samples (Supplemental Figure S1), which hampered the complete haplotype analysis. Nevertheless, besides having in common the c.1901C>T variant, these families shared several STR alleles among themselves and with the other four families widely haplotyped as described above.

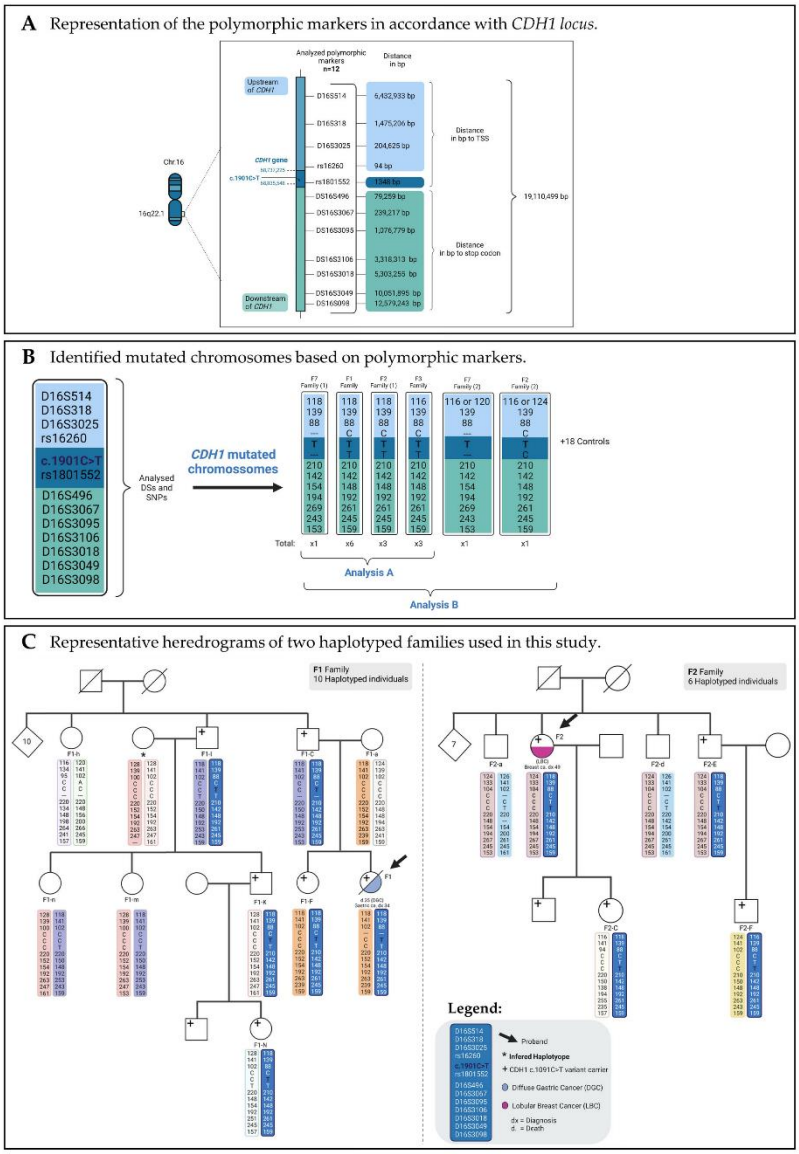


Figure 4. Polymorphic markers used and haplotype results for characterizing c.1901C>T mutated families: (**A**) Representation of the studied polymorphic markers in accordance with *CDH1* locus; (**B**) Identified *CDH1* mutated chromosomes used in the two different runs performed with DHSMAP for time to most recent common ancestor (TMRCA) estimation; (**C**) Selected heredograms of two representative families used in this study. TSS, transcription start site; bp, base pairs; STRs, short tandem repeats.

3.3. CDH1 c.1901C>T a Founder Variant with Approximately 500 Years

We used two approaches to estimate the age of *CDH1* c.1901C>T variant (Figure 4B and Supplemental Table S4). In Analysis A, we performed a more conservative analysis, and a single affected chromosome per family was considered. In Analysis B, the previously identified chromosomes were used together with two derived haplotypes in DHSMAP. The TMRCA estimates converged in both analyses to 19.58 generations ago, or 489.64 years, given a generation of roughly 25 years. The results of both approaches diverged only in their 95% confidence intervals (CI), which were larger in the 4-chromosome analysis (Analysis A): 17.8–436 generations or 445–10,900 years. The inclusion of two derived haplotypes (Analysis B) had a 95% CI of 449.75–1492 years, if assuming a recombination event, and 449.75–1475.5 years if a mutational event has occurred instead. This means that the original carrier of this mutation, the founder, lived in the 16th century, and that the mutation has segregated in Northern Portuguese populations for nearly 500 years. At the moment, the distribution and frequency of this variant in the population is unpredictable.

3.4. Clinical Presentations in HDGC Families Carrying CDH1 c.1901C>T

Assuming, from the data described above, that all families from Northern Portugal carrying the very rare *CDH1* c.1901C>T truncating variant had a common ancestor, we next present the clinical data from all carriers together.

To date, 134 individuals from nine families underwent carrier testing for the *CDH1* c.1901C>T variant, and 43.3% (58/134) were proven carriers of this variant. Of the 58 carriers, 31 were females (53.4%). We depict the numbers of clinically diagnosed DGC cases in Table 1 and LBC cases in Table 2, as well as related numbers of carriers submitted to risk reduction surgeries and, of these, those revealing DGC and/or LBC foci during surveillance.

Table 1. Overview of stomach-related clinical presentations in *CDH1* c.1901C>T variant carriers from nine HDGC families.

	Total n° of Tested Individuals	Total n° of Carriers	DGC Cases	Mean Age of DGC Diagnosis (±SD)	RRG	RRG with Lesions
Males	60	27	4	27 ± 7	10	8
Females	74	31	7 *	37 ± 13	16	13 #
Total	134	58	11	33 ± 12	26	21

Footnote: * One out of seven females presented DGC and LBC; # from 13 females bearing lesions in RRG, 3 also presented LBC. DGC, diffuse gastric cancer; RRG, risk reduction gastrectomy; LBC, lobular breast cancer; Age of diagnosis is represented by the mean of the affected carrier's age ± SD.

Table 2. Overview of breast-related clinical presentations in *CDH1* c.1901C>T variant carriers from nine HDGC families.

Number of Females Tested	Total n° of Carriers	LBC Cases	Mean Age of LBC Diagnosis (± SD)	RRM	RRM with Lesions	Lesions in both RRG and RRM
74	31	6	50 ± 8	8	4 *	3 #

Footnote: * Two out of four female patients with RRM lesions were submitted to unilateral mastectomy, since they were previously diagnosed with LBC in the other breast; # all three patients were previously diagnosed with LBC and presented carcinoma foci in the RRM (on the other breast) and in the RRG. LBC, lobular breast cancer; RRM, risk reduction mastectomy; Age of diagnosis is represented by the mean of the affected carrier's age ± SD.

DGC was diagnosed in 11/58 (18.9%) variant carriers, four males (M) with an average age of onset of 27 ± 7 years old and seven females (F) with an average age of onset of 37 ± 13 years old. Five out of the 11 DGC patients (2M/3F) were diagnosed in the context of disease surveillance (endoscopic biopsy) (Table 1).

Twenty-six carriers (10M/16F) decided to undergo risk reduction gastrectomy, and in 21/26 (80.8%; 8M/13F), DGC foci were found in the gastrectomy specimen.

LBC, which only occurs in females (31/58 carriers), was diagnosed in 6/31 (19.4%) female carriers at an average age of onset of 50 ± 8 years old (Table 2). Two out of the

six LBC patients were diagnosed during disease surveillance in the context of the high-risk consultation. From the female carriers that performed prophylactic mastectomies, 4/8 (50%) presented LBC foci (Table 2). Importantly, two of these women were already previously diagnosed with LBC on the contralateral breast; consequently, only unilateral prophylactic mastectomy was performed. A third woman, also previously diagnosed with LBC, underwent tumourectomy before proceeding to bilateral mastectomy. Therefore, only one of the four female carriers had no previous lesions when bilateral risk reducing mastectomy was performed. Three female carriers were found with both DGC and LBC foci in risk reduction surgical specimens (Table 1).

We further analysed the average age and age range of carriers upon disease onset, genetic test, or last surveillance, and separated the cohort (n = 58 carriers) in four groups: (1) carriers presenting clinical expression of DGC and/or LBC; (2) asymptomatic carriers with subclinical disease (foci found in RRG or RRM); (3) carriers without foci detected in RRG and RRM specimens; and (4) asymptomatic carriers not submitted to risk reduction surgery. The average age of onset of the first group, including all carriers with DGC and/or LBC (n = 16), was 39 ± 13.98 years old (age range: 18–61 years old). The average age at which subclinical disease was identified in the second group of asymptomatic carriers (n = 18), upon risk-reduction surgery, was 36 ± 14 years old (age range: 14–63). In the third group, carriers without subclinical disease (n = 4), the average age at risk reduction surgery was 40 ± 14.9 years old (age range: 23–59). The remaining carriers were also asymptomatic and decided to avoid RRG and/or RRM (n = 20). In this last group, follow-up was carried out for DGC in both genders and for LBC in females, starting at an average age of 44 ± 22.7 years old (age range: 18–83), and to date none have developed clinical disease. These data support a high variability in age of disease onset and incomplete penetrance in these families. c.1901C>T is, therefore, a deleterious but low-penetrant variant, probably because of co-segregation with either protective or disease-predisposing genetic modifiers.

4. Discussion

In this study, we used biological material from HDGC patients to formally prove that the single nucleotide variant in *CDH1* (c.1901C>T) they carry in their germline, earlier identified as the missense variant A634V (Figure 1), is indeed a bona fide splice-site variant (r.1900_1936del; p. Ala634ProfsTer7). This finding supports its classification as pathogenic according to ACMG/AMP guidelines for *CDH1* variant classification [12]. Like several other *CDH1* variants [12,14], the c.1901C>T induces cryptic splicing within *CDH1* exon 12, frameshifting, and premature truncation (Figure 2 and Supplemental Figure S3). This is believed to target the variant-bearing allele for degradation, most likely through the mRNA nonsense-mediated decay (NMD) pathway. NMD is a quality-control translational step that selectively degrades mRNAs harbouring premature termination (nonsense) codons [27,28]. Therefore, normal E-cadherin protein levels are sustained exclusively by the wild-type allele in normal tissues, until a second somatic hit occurs to abolish *CDH1* function in HDGC-target organs. As compared to transcripts generated by genuine missense variants, which produce E-cadherin molecules and retain residual activity, the complete loss of E-cadherin immunoexpression in DGC and LBC from carriers, accompanying the strong downregulation (at least 66% in relation to the wild-type allele) of PTC-bearing transcripts produced by the mutant allele, may explain the aggressive phenotypes seen in patients carrying the c.1901C>T germline variant [13,29]. We published on similar levels of down-regulation of PTC-carrying alleles, using RNA extracted either from peripheral blood lymphocytes (PBLs) or normal gastric mucosa, from patients carrying germline nonsense variants in *CDH1* from different families in two of our previous papers [28,30]. The lack of complete downregulation of the PTC-bearing allele by NMD may result from a delay in degradation after transcription. Indeed, as none of the nonsense variants studied resulted in transcription inactivation, the transcription level of the mutant allele is likely similar to that of the wild-type. After transcription, splicing and the pioneer round of translation

initiate, as well as degradation by NMD of PTC-carrying alleles [31]. This delay likely explains the residual detection of c.1901T variant-bearing transcripts in PBL samples from the c.1901C>T variant carriers. Regarding the impact that complete or incomplete down-regulation of the mutant allele may have for disease phenotype and severity, very little is known. However, it has been documented that certain genetic modifiers, and interindividual variability in NMD efficiency between patients carrying identical mutations, may lead to differences in the disease severity and clinical phenotype [32].

Although a previous study reported splicing impairment related to this variant in a colorectal cancer cell line [21] and using minigene analysis [14], the current report was the first to assess the impact of c.1901C>T in germline RNA from HDGC patients, supporting its reclassification as a truncating variant. This information should replace the fundaments for the variant's pathogenicity, currently available at the ClinVar entry VCV000012244.5, which mainly describe findings on the functional impact of a p.A634V missense-bearing protein.

Recently, Corso et al. performed an extensive literature review on *CDH1* missense variants described to occur in HDGC patients in order to assess their relevance and implications for clinical management of this disease [22]. This study, albeit one of great relevance, considered the c.1901C>T variant as a missense variant, and the functional consequence of the p.A634V mutant protein. This formally contaminated data analysis from pure missense variants with data referring to bona fide truncating variants.

Another important subject is the choice of functional studies to understand the impact of *CDH1* variants. Indeed, we and others have predicted a deleterious effect for the *CDH1* p.A643V missense variant, supposedly generated by the c.1901C>T variant. These predictions were based on overexpression of a cDNA construct with a single nucleotide variant, mimicking the c.1901C>T variant, in *CDH1*-negative cell lines [14,21,33]. However, given the data herein reported, such a missense-bearing transcript and mutant E-cadherin protein do not occur in nature. This brings into discussion the absolute need to assess the impact, in silico and at the RNA level, of variants classified as missense, prior to the design of the most appropriate set of functional analyses. This will be fundamental for variant interpretation and will certainly improve variant classification and variant carriers' management. The molecular reclassification of some missense variants may indeed change our knowledge on DGC and LBC frequencies by type of variant and, consequently, on disease penetrance estimation in HDGC.

Several publications have now demonstrated that a considerable fraction of variants initially assigned as missense are indeed splice-site variants. One such publication studied *NF1*, a gene causing neurofibromatosis type 1 [34]. After applying in silico predictive tools to 41 variants previously classified as missense variants, the authors demonstrated that over 45% of them were shown to affect splicing and claimed that such variants have been frequently underscored if not analysed in depth. These observations were further supported by studies using next-generation sequencing technology [32] and high-throughput splicing reporter assays [35] to unravel the effect on splicing of single nucleotide variants, including missense, in the von Willebrand factor gene and *POU1F1*, respectively. Altogether, there is a need for improvement of in silico tools to predict cryptic splicing, and for the development of guidelines for formal analysis of splicing defects induced by SNVs, particularly synonymous and missense variants. These will certainly improve the current knowledge of the molecular mechanisms causing disease and will facilitate variant interpretation.

The *CDH1* c.1901C>T variant was previously described in unrelated HDGC families in Portugal, England, and New Zealand [14,36–38]. Its unusually high frequency among early-onset DGC and LBC cancer cases, and families fulfilling HDGC criteria, in a specific region of Northern Portugal (Porto district) [18,19], led us to explore the occurrence of a potential founder effect.

Haplotyping data from four families carrying the c.1901 C>T variant showed carriers to share at least 1.8 Mbp of chromosome 16. Strikingly, families F1 and F2 shared the same haplotype in all the polymorphic markers studied (total of 19.2 Mbp, Figure 4B).

These results indicate that F1 and F2 carriers inherited and maintained the same genetic information from a past common ancestor. Carriers from family F7 differed from those from families F1 and F2 in five STR markers, showing that F7 family diverged from the most likely common ancestor present in F1 and F2 by a recombination event close to D16S3095 (Figure 4B). This was demonstrated first by a 6 bp difference in D16S3095 (154 bp in F7 vs. 148 bp in F1 and F2), corresponding to three mutational steps (three dinucleotide repeats) [39], and second because downstream markers also deviated from the common ancestral haplotype. Family F3 distinguished itself from the F1 and F2 in a single 5′-external polymorphic marker (Figure 4B). Because of the absence of upstream markers, we could not infer whether the affected chromosome detected in F3 resulted from mutation or recombination. In our study, we identified two derived chromosomes (F7-A carrier, and F2-F carrier) that differed within their families for the marker D16S514. This polymorphic variability may have resulted either from an acquired mutation, particularly if it only differed from the ancestral allele by one mutational step (single repeat—116 bp allele), or from recombination (120 bp in F7 and 124 bp in F2). Unfortunately, we could not discriminate which was the case here, as the derived haplotype could not be accurately inferred.

The two DHSMAP analysis (A and B) converged to the same time frame, 490 years ago, only diverging in their confidence intervals (CI). When the two descendent haplotypes, which differed only by a mutation or recombination event at D16S514, were added, the CI was narrower. However, as explained above, it was not possible to infer these chromosomes properly. Using a more conservative hypothesis, our CI was very broad, indicating future improvements to age estimation by including a higher number of samples from different c.1901C>T variant-carrying HDGC families.

This study estimated that the c.1901C>T variant has been segregating within the Portuguese population since the 16th century, close to the year 1528. This date was calculated by subtracting the estimated *CDH1* allele age (490 years) from the year 2018, when this analysis was performed (1528 = 2018−490 years) and assuming a 25-year generation time. These data may also indicate that more families are likely to be at risk and remain unidentified to date.

There is a limitation in this study regarding the number of carrier individuals used for the RNA analysis; however, as carriers herein analysed were from different families, we believe our data were still robust. Additionally, for the purpose of inferring haplotypes, the lack of samples from complete trios could interfere with the presented estimations. For this reason, families F4, F5, and F6 were excluded from the TMRCA analysis (Supplemental Figure S1).

One of the critical challenges in HDGC clinical management is to predict whether a carrier will develop the disease or not. Although proven pathogenic, this variant seems to present a low penetrance in HDGC families, as depicted by the DGC and LBC frequencies among carriers. From the 58 mutation carriers identified, only 27.6% (16/58) were clinically affected, and only one female was diagnosed with both LBC and DGC (Tables 1 and 2). Moreover, predicting whether premalignant lesions could lead to carcinomas or remain indolent is a challenge to solve in the future. Indeed, signet-ring cell carcinoma (SRCC) foci can remain indolent without proliferation, never leading to clinical disease, but some foci may also evolve, becoming very aggressive and often killing the patient. It has been reported in other studies that the number of cancer foci found in prophylactic gastrectomy specimens ranged from 1 to 487, and the size from < 0.1 mm to 16.0 mm, without a clear relation to patients' age or gender [40,41]. Therefore, some of the carriers herein presented may develop SRCC foci and/or clinically expressed cancer later in their lives. Additionally, unrecognized genetic modifiers, either protective or predisposing to disease development, may be conditioning the apparently low disease penetrance in the setting herein described.

5. Conclusions

The *CDH1* c.1901C>T variant, widely studied for its functional impact as the missense variant p.A634V, has been demonstrated to be a bona fide truncating variant (r.1900_1936del; p.Ala634ProfsTer7), supporting its classification as pathogenic according to ACMG/AMP *CDH1* guidelines.

The *CDH1* c.1901C>T variant was herein proven as the first *CDH1*-associated founder effect in Portugal, still restricted to a specific region of the northeastern part of the Porto district and aged approximately 490 years.

In our cohort of nine HDGC families carrying the *CDH1* c.1901C>T variant, the frequency of DGC and LBC was 18.9% and 19.4%, respectively. However, the prophylactic measures applied resulted in the identification of carcinoma foci in almost 81% of the carriers who underwent prophylactic gastrectomy and 50% of those submitted for prophylactic mastectomies, demonstrating the efficiency of these measures.

Despite the clearly incomplete penetrance of disease in these families, the c.1901C>T variant's early age of onset and aggressiveness highlights the importance and supports the use of intensive surveillance and prophylaxis in asymptomatic carriers of this variant.

The herein-established clinical actionability of the c.1901C>T variant, and the demonstration of a founder effect in this geographical region, empowered the monitoring of at-risk individuals and diagnosis of DGC and LBC during surveillance.

Publishing these data will increase awareness among clinicians on the need to adjust HDGC testing criteria in high-risk geographical areas such as Northern Portugal.

Supplementary Materials: The following are available online at https://www.mdpi.com/article/10.3390/cancers13174464/s1, Supplemental Figure S1: Partial pedigrees with corresponding haplotypes for all families studied; Supplemental Figure S2: Complete 2% agarose gels from colony PCR using c.1901C>T carriers RNA; Supplemental Figure S3: Sequencing results from LP and SP colonies; Supplemental Table S1: HDGC clinical criteria, updated in 2020; Supplemental Table S2: List of primers and PCR conditions for colony PCR strategy; Supplemental Table S3: List of primers and PCR conditions for polymorphic marker studies; Supplemental Table S4: Chromosomes' haplotypes from HDGC families used for TMRCA estimation.

Author Contributions: C.O. conceived and supervised the study; C.O., R.B.-M., R.L.S., and L.G. designed the study; R.B.-M., R.L.S. L.G., A.C.A., J.G.-P., A.A., S.S., H.P., and I.G. acquired and generated data; C.O., S.P.S., L.G., R.B.-M., R.L.S., and A.C.A. analysed and interpreted the data; S.S. designed and supervised the haplotype analysis; L.F., L.V., V.D., R.O., S.F., S.C.C., A.M., M.B., F.C., and S.C. acquired clinical data; L.G., R.B.-M., and R.L.S. drafted the manuscript; C.O. edited and reviewed the manuscript; and C.O. obtained funding. All authors critically revised the manuscript for important intellectual content. All authors have read and agreed to the published version of the manuscript.

Funding: This research and its authors were funded by FEDER—Fundo Europeu de Desenvolvimento Regional funds through the COMPETE 2020—Operational 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) and LEGOH (PTDC/BTM-TEC/6706/2020). This work was also financed by the projects NORTE-01-0145-FEDER-000003 (DOCnet)—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 Ethics Committee of *Centro Hospitalar Universitário São João* (CHUSJ), Porto (protocol code 307-17 approved on 25 January 2018).

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

Data Availability Statement: Clinical and genetic data from patients herein described are not publicly available due to ethical, legal and privacy issues.

Acknowledgments: The authors would like to thank the patients and caregivers for their consent to participate in this study. R.B.-M. acknowledges the support of the Institute of Biomedical Sciences Abel Salazar from University of Porto, specifically by the Doctoral Programme BiotechHealth and the FCT—Fundação para a Ciência e a Tecnologia for her PhD fellowship (ref. SFRH/BD/145132/2019). R.L.S. acknowledges the support of the Faculty of Medicine, University of Porto, specifically by the Doctoral Programme on Health Data Science and the FCT—Fundação para a Ciência e a Tecnologia, for her Ph.D. fellowship. J.G.-P. acknowledges the support of the Faculty of Medicine, University of Porto, specifically by the Doctoral Programme on Biomedicine, and the EU project Solve-RD (Grant Ref. 779257) for salary support. We acknowledge the European Reference Network on Genetic Tumour Risk Syndromes (ERN GENTURIS)—Project ID No. 739547. ERN GENTURIS is partly cofunded by the European Union within the framework of the Third Health Programme "ERN-2016—Framework Partnership Agreement 2017–2021. All figures present in the manuscript were created with BioRender.com.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. 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]
2. 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]
3. 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.; et al. Familial gastric cancer: Overview and guidelines for management. *J. Med. Genet.* **1999**, *36*, 873–880.
4. Oliveira, C.; Senz, J.; Kaurah, P.; Pinheiro, H.; Sanges, R.; Haegert, A.; Corso, G.; Schouten, J.; Fitzgerald, R.; Vogelsang, H.; et al. Germline *CDH1* deletions in hereditary diffuse gastric cancer families. *Hum. Mol. Genet.* **2009**, *18*, 1545–1555. [CrossRef]
5. 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]
6. Fitzgerald, R.C.; Hardwick, R.; Huntsman, D.; Carneiro, F.; Guilford, P.; Blair, V.; Chung, D.C.; Norton, J.; Ragunath, K.; Van Krieken, J.H.; et al. Hereditary diffuse gastric cancer: Updated consensus guidelines for clinical management and directions for future research. *J. Med. Genet.* **2010**, *47*, 436–444. [CrossRef]
7. 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.; 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] [PubMed]
8. 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]
9. Pharoah, P.D.; Guilford, P.; Caldas, C.; International Gastric Cancer Linkage Consortium. Incidence of gastric cancer and breast cancer in *CDH1* (E-cadherin) mutation carriers from hereditary diffuse gastric cancer families. *Gastroenterology* **2001**, *121*, 1348–1353. [CrossRef] [PubMed]
10. 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]
11. Xicola, R.M.; Li, S. Clinical features and cancer risk in families with pathogenic *CDH1* variants irrespective of clinical criteria. *J. Med. Genet.* **2019**, *56*, 838–843. [CrossRef] [PubMed]
12. Lee, K.; Krempely, K. Specifications of the ACMG/AMP variant curation guidelines for the analysis of germline *CDH1* sequence variants. *Hum. Mutat.* **2018**, *39*, 1553–1568. [CrossRef] [PubMed]
13. Melo, S.; Figueiredo, J. Predicting the Functional Impact of *CDH1* Missense Mutations in Hereditary Diffuse Gastric Cancer. *Int. J. Mol. Sci.* **2017**, *18*, 2687. [CrossRef]
14. Kaurah, P.; MacMillan, A.; Boyd, N.; Senz, J.; De Luca, A.; Chun, N.; Suriano, G.; Zaor, S.; Van Manen, L.; Gilpin, C.; et al. Founder and recurrent *CDH1* mutations in families with hereditary diffuse gastric cancer. *JAMA* **2007**, *297*, 2360–2372. [CrossRef]
15. Evans, J.A. Old meets new: Identifying founder mutations in genetic disease. *Can. Med. Assoc. J.* **2015**, *187*, 93–94. [CrossRef]
16. Kivisild, T. Founder Effect. In *Brenner's Encyclopedia of Genetics*, 2nd ed.; Academic Press: Massachusetts, MA, USA, 2013; pp. 100–101.
17. Keinan, A.; Clark, A.G. Recent explosive human population growth has resulted in an excess of rare genetic variants. *Science* **2012**, *336*, 740–743. [CrossRef]
18. Castro, C.; Antunes, L.; Lunet, N.; Bento, M.J. Cancer incidence predictions in the North of Portugal: Keeping population-based cancer registration up to date. *Eur. J. Cancer Prev.* **2016**, *25*, 472–480. [CrossRef]
19. Castro, C.; Peleteiro, B.; Bento, M.J.; Lunet, N. Trends in gastric and esophageal cancer incidence in northern Portugal (1994-2009) by subsite and histology, and predictions for 2015. *Tumori* **2017**, *103*, 155–163. [CrossRef]

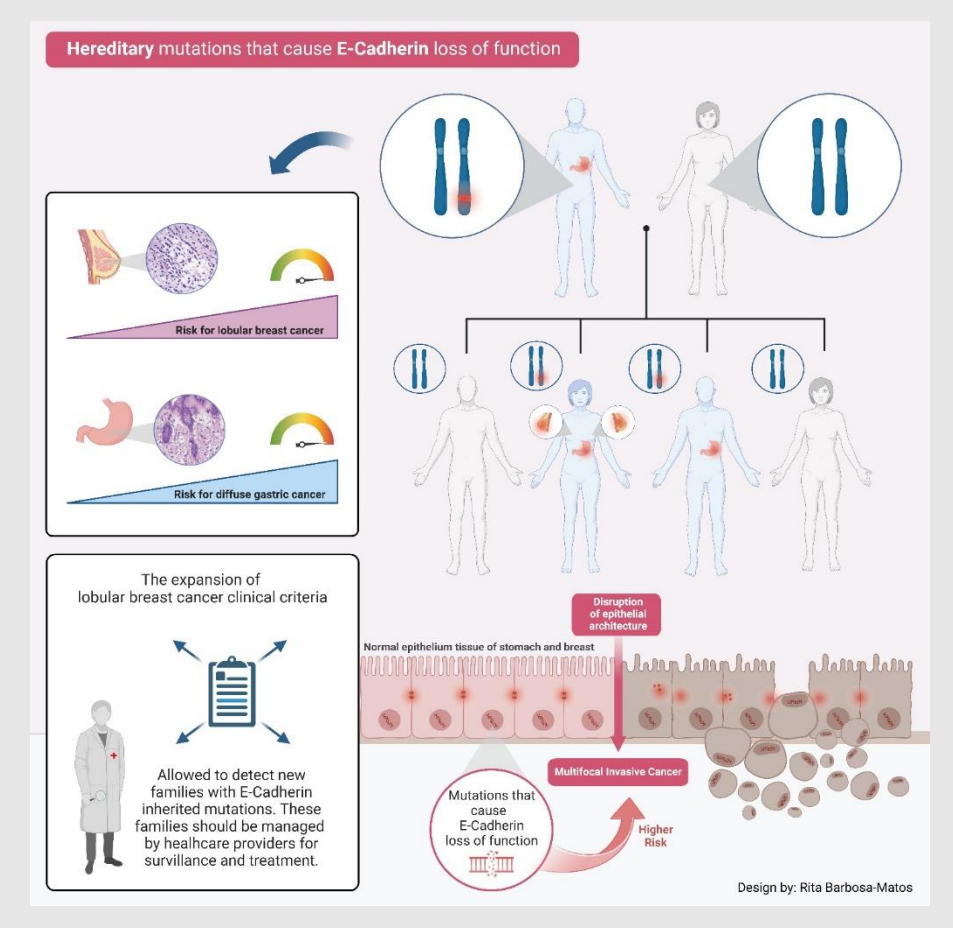
20. Ernst, M. *Systematics and the Origin of Species*; Columbia University Press: New York City, NY, USA, 1942; p. 237.
21. Vecsey-Semjen, B.; Becker, K.F.; Sinski, A.; Blennow, E.; Vietor, I.; Zatloukal, K.; Beug, H.; Wagner, E.; Huber, L.A. Novel colon cancer cell lines leading to better understanding of the diversity of respective primary cancers. *Oncogene* **2002**, *21*, 4646–4662. [CrossRef]
22. Corso, G.; Corso, F. Geographical Distribution of E-cadherin Germline Mutations in the Context of Diffuse Gastric Cancer: A Systematic Review. *Cancers* **2021**, *13*, 1269. [CrossRef]
23. McPeek, M.S.; Strahs, A. Assessment of linkage disequilibrium by the decay of haplotype sharing, with application to fine-scale genetic mapping. *Am. J. Hum. Genet.* **1999**, *65*, 858–875. [CrossRef]
24. Vaz Rodrigues, L.; Costa, F.; Marques, P.; Mendonça, C.; Rocha, J.; Seixas, S. Severe α-1 antitrypsin deficiency caused by Q0(Ourém) allele: Clinical features, haplotype characterization and history. *Clin. Genet.* **2012**, *81*, 462–469. [CrossRef]
25. Weber, J.L.; Wong, C. Mutation of human short tandem repeats. *Hum. Mol. Genet.* **1993**, *2*, 1123–1128. [CrossRef]
26. Nachman, M.W.; Crowell, S.L. Estimate of the mutation rate per nucleotide in humans. *Genetics* **2000**, *156*, 297–304. [CrossRef]
27. Chang, Y.F.; Imam, J.S.; Wilkinson, M.F. The nonsense-mediated decay RNA surveillance pathway. *Annu. Rev. Biochem.* **2007**, *76*, 51–74. [CrossRef]
28. Karam, R.; Carvalho, J.; Bruno, I.; Graziadio, C.; Senz, J.; Huntsman, D.; Carneiro, F.; Seruca, R.; Wilkinson, M.F.; Oliveira, C. The NMD mRNA surveillance pathway downregulates aberrant E-cadherin transcripts in gastric cancer cells and in *CDH1* mutation carriers. *Oncogene* **2008**, *27*, 4255–4260. [CrossRef] [PubMed]
29. Figueiredo, J.; Soderberg, O.; Simoes-Correia, J.; Grannas, K.; Suriano, G.; Seruca, R. The importance of E-cadherin binding partners to evaluate the pathogenicity of E-cadherin missense mutations associated to HDGC. *Eur. J. Hum. Genet.* **2013**, *21*, 301–309. [CrossRef]
30. Pinheiro, H.; Bordeira-Carrico, R.; Seixas, S.; Carvalho, J.; Senz, J.; Oliveira, P.; Inacio, P.; Gusmao, L.; Rocha, J.; Huntsman, D.; et al. Allele-specific *CDH1* downregulation and hereditary diffuse gastric cancer. *Hum. Mol. Genet.* **2010**, *19*, 943–952. [CrossRef]
31. Hwang, J.; Sato, H.; Tang, Y.; Matsuda, D.; Maquat, L.E. UPF1 association with the cap-binding protein, CBP80, promotes nonsense-mediated mRNA decay at two distinct steps. *Mol. Cell* **2010**, *39*, 396–409. [CrossRef] [PubMed]
32. Borràs, N.; Orriols, G.; Batlle, J.; Pérez-Rodríguez, A.; Fidalgo, T.; Martinho, P.; López-Fernández, M.F.; Rodríguez-Trillo, Á.; Lourés, E.; Parra, R.; et al. Unraveling the effect of silent, intronic and missense mutations on VWF splicing: Contribution of next generation sequencing in the study of mRNA. *Haematologica* **2019**, *104*, 587–598. [CrossRef]
33. Pereira, P.S.; Teixeira, A.; Pinho, S.; Ferreira, P.; Fernandes, J.; Oliveira, C.; Seruca, R.; Suriano, G.; Casares, F. E-cadherin missense mutations, associated with hereditary diffuse gastric cancer (HDGC) syndrome, display distinct invasive behaviors and genetic interactions with the Wnt and Notch pathways in Drosophila epithelia. *Hum. Mol. Genet.* **2006**, *15*, 1704–1712. [CrossRef]
34. Stella, A.; Lastella, P.; Loconte, D.C.; Bukvic, N. Accurate Classification of NF1 Gene Variants in 84 Italian Patients with Neurofibromatosis Type 1. *Genes* **2018**, *9*, 216. [CrossRef]
35. Gergics, P.; Smith, C.; Bando, H.; Jorge, A.A.L.; Rockstroh-Lippold, D.; Vishnopolska, S.A.; Castinetti, F.; Maksutova, M.; Carvalho, L.R.S.; Hoppmann, J.; et al. High-throughput splicing assays identify missense and silent splice-disruptive POU1F1 variants underlying pituitary hormone deficiency. *Am. J. Hum. Genet.* **2021**, *108*, 1526–1539. [CrossRef]
36. Suriano, G.; Oliveira, C.; Ferreira, P.; Machado, J.C.; Bordin, M.C.; De Wever, O.; Bruyneel, E.A.; Moguilevsky, N.; Grehan, N.; Porter, T.R.; et al. Identification of *CDH1* germline missense mutations associated with functional inactivation of the E-cadherin protein in young gastric cancer probands. *Hum. Mol. Genet.* **2003**, *12*, 575–582. [CrossRef] [PubMed]
37. More, H.; Humar, B.; Weber, W.; Ward, R.; Christian, A.; Lintott, C.; Graziano, F.; Ruzzo, A.M.; Acosta, E.; Boman, B.; et al. Identification of seven novel germline mutations in the human E-cadherin (*CDH1*) gene. *Hum. Mutat.* **2007**, *28*, 203. [CrossRef]
38. Oliveira, C.; Ferreira, P.; Nabais, S.; Campos, L.; Ferreira, A.; Cirnes, L.; Alves, C.C.; Veiga, I.; Fragoso, M.; Regateiro, F.; et al. E-Cadherin (*CDH1*) and p53 rather than SMAD4 and Caspase-10 germline mutations contribute to genetic predisposition in Portuguese gastric cancer patients. *Eur. J. Cancer* **2004**, *40*, 1897–1903. [CrossRef]
39. Shriver, M.D.; Jin, L.; Chakraborty, R.; Boerwinkle, E. VNTR allele frequency distributions under the stepwise mutation model: A computer simulation approach. *Genetics* **1993**, *134*, 983–993. [CrossRef] [PubMed]
40. Rocha, J.P.; Gullo, I.; Wen, X.; Devezas, V.; Baptista, M.; Oliveira, C.; Carneiro, F. Pathological features of total gastrectomy specimens from asymptomatic hereditary diffuse gastric cancer patients and implications for clinical management. *Histopathology* **2018**, *73*, 878–886. [CrossRef]
41. Barber, M.E.; Save, V.; Carneiro, F.; Dwerryhouse, S.; Lao-Sirieix, P.; Hardwick, R.H.; Caldas, C.; Fitzgerald, R.C. Histopathological and molecular analysis of gastrectomy specimens from hereditary diffuse gastric cancer patients has implications for endoscopic surveillance of individuals at risk. *J. Pathol.* **2008**, *216*, 286–294. [CrossRef]

Supplementary Material: Available through the link:

https://www.mdpi.com/article/10.3390/cancers13174464/s1

2.2. Paper 4, Chapter III

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



Author statement:
Rita Barbosa-Matos contributed to the investigation, methodology, figure design, review, and editing.

The main aim of this study is to determine associations between *CDH1* germline variants, and the cancer phenotypes reported in *CDH1* carriers, independently of fulfilling DGLBCS criteria. We prove that truncating pathogenic and likely-pathogenic *CDH1* mutations are positively associated with the development of DGC and LBC, but that there is no evidence for a positive association with these phenotypes and *CDH1* missense variants of unknown significance. Moreover, PV/LPVs occurred often in families enriched in LBC cases not fulfilling DGLBCS criteria, supporting the expansion clinical criteria to accommodate these cases. Our findings are significant for selecting *CDH1* carriers for management, prioritizing the target organs for surveillance.

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

José Garcia-Pelaez, Rita Barbosa-Matos, Silvana Lobo*, Alexandre Dias*, Luzia Garrido, Sérgio Castedo, Sónia Sousa, Hugo Pinheiro, Liliana Sousa, Rita Monteiro, Joaquin J Maqueda, Susana Fernandes, Fátima Carneiro, Nádia Pinto, Carolina Lemos, Carla Pinto, Manuel R Teixeira, Stefan Aretz, Svetlana Bajalica-Lagercrantz, Judith Balmaña, Ana Blatnik, Patrick R Benusiglio, Maud Blanluet, Vincent Bours, Hilde Brems, Joan Brunet, Daniele Calistri, Gabriel Capellá, Sergio Carrera, Chrystelle Colas, Karin Dahan, Robin de Putter, Camille Desseignés, Elena Domínguez-Garrido, Conceição Egas, D Gareth Evans, Damien Feret, Eleanor Fewings, Rebecca C Fitzgerald, Florence Coulet, María Garcia-Barcina, Maurizio Genuardi, Lisa Golmard, Karl Hackmann, Helen Hanson, Elke Holinski-Feder, Robert Hüneburg, Mateja Krajc, Kristina Lagerstedt-Robinson, Conxi Lázaro, Marjolijn J L Ligtenberg, Cristina Martínez-Bouzas, Sonia Merino, Geneviève Michils, Srdjan Novaković, Ana Patiño-García, Guglielmina Nadia Ranzani, Evelin Schröck, Inês Silva, Catarina Silveira, José L Soto, Isabel Spier, Verena Steinke-Lange, Gianluca Tedaldi, María-Isabel Tejada, Emma R Woodward, Marc Tischkowitz, Nicoline Hoogerbrugge, Carla Oliveira*

Summary

Background Truncating pathogenic or likely pathogenic variants of *CDH1* cause hereditary diffuse gastric cancer (HDGC), a tumour risk syndrome that predisposes carrier individuals to diffuse gastric and lobular breast cancer. Rare *CDH1* missense variants are often classified as variants of unknown significance. We conducted a genotype–phenotype analysis in families carrying rare *CDH1* variants, comparing cancer spectrum in carriers of pathogenic or likely pathogenic variants (PV/LPV; analysed jointly) or missense variants of unknown significance, assessing the frequency of families with lobular breast cancer among PV/LPV carrier families, and testing the performance of lobular breast cancer-expanded criteria for *CDH1* testing.

Methods This genotype-first study used retrospective diagnostic and clinical data from 854 carriers of 398 rare *CDH1* variants and 1021 relatives, irrespective of HDGC clinical criteria, from 29 institutions in ten member-countries of the European Reference Network on Tumour Risk Syndromes (ERN GENTURIS). Data were collected from Oct 1, 2018, to Sept 20, 2022. Variants were classified by molecular type and clinical actionability with the American College of Medical Genetics and Association for Molecular Pathology *CDH1* guidelines (version 2). Families were categorised by whether they fulfilled the 2015 and 2020 HDGC clinical criteria. Genotype–phenotype associations were analysed by Student's *t* test, Kruskal-Wallis, χ^2, and multivariable logistic regression models. Performance of HDGC clinical criteria sets were assessed with an equivalence test and Youden index, and the areas under the receiver operating characteristic curves were compared by *Z* test.

Findings From 1971 phenotypes (contributed by 854 probands and 1021 relatives aged 1–93 years), 460 had gastric and breast cancer histology available. *CDH1* truncating PV/LPVs occurred in 176 (21%) of 854 families and missense variants of unknown significance in 169 (20%) families. Multivariable logistic regression comparing phenotypes occurring in families carrying PV/LPVs or missense variants of unknown significance showed that lobular breast cancer had the greatest positive association with the presence of PV/LPVs (odds ratio 12·39 [95% CI 2·66–57·74], p=0·0014), followed by diffuse gastric cancer (8·00 [2·18–29·39], p=0·0017) and gastric cancer (7·81 [2·03–29·96], p=0·0027). 136 (77%) of 176 families carrying PV/LPVs fulfilled the 2015 HDGC criteria. Of the remaining 40 (23%) families, who did not fulfil the 2015 criteria, 11 fulfilled the 2020 HDGC criteria, and 18 had lobular breast cancer only or lobular breast cancer and gastric cancer, but did not meet the 2020 criteria. No specific *CDH1* variant was found to predispose individuals specifically to lobular breast cancer, although 12 (7%) of 176 PV/LPV carrier families had lobular breast cancer only. Addition of three new lobular breast cancer-centred criteria improved testing sensitivity while retaining high specificity. The probability of finding *CDH1* PV/LPVs in patients fulfilling the lobular breast cancer-expanded criteria, compared with the 2020 criteria, increased significantly (AUC 0·92 *vs* 0·88; *Z* score 3·54; p=0·0004).

Interpretation *CDH1* PV/LPVs were positively associated with HDGC-related phenotypes (lobular breast cancer, diffuse gastric cancer, and gastric cancer), and no evidence for a positive association with these phenotypes was found for *CDH1* missense variants of unknown significance. *CDH1* PV/LPVs occurred often in families with lobular breast cancer who did not fulfil the 2020 HDGC criteria, supporting the expansion of lobular breast cancer-centred criteria.

Lancet Oncol 2023; 24: 91–106

Published Online November 24, 2022 https://doi.org/10.1016/S1470-2045(22)00643-X

This online publication has been corrected. The corrected version first appeared at thelancet.com/oncology on January 3, 2022

[†]Contributed equally

Instituto de Investigação e Inovação em Saúde (i3S) (J Garcia-Pelaez MSc, R Barbosa-Matos MSc, S Lobo MSc, A Dias MSc, Prof S Castedo MD PhD, S Sousa BSc, H Pinheiro MD PhD, L Sousa MSc, C Lemos PhD, R Monteiro MSc, J J Maqueda MSc, Prof F Carneiro MD PhD, N Pinto PhD, Prof C Oliveira PhD), **Faculty of Medicine** (J Garcia-Pelaez, Prof S Castedo, S Fernandes PhD, Prof F Carneiro, Prof C Oliveira), **Institute of Biomedical Sciences Abel Salazar** (R Barbosa-Matos, S Lobo, C Lemos PhD, M R Teixeira MD PhD), **Center of Mathematics** (N Pinto), **and Institute for Molecular and Cell Biology** (C Lemos), **University of Porto, Porto, Portugal; Instituto de Patologia e Imunologia Molecular da Universidade do Porto, Porto, Portugal** (J Garcia-Pelaez, R Barbosa-Matos, S Lobo, A Dias, Prof S Castedo, S Sousa, H Pinheiro, L Sousa, R Monteiro, J J Maqueda, Prof F Carneiro, N Pinto, Prof C Oliveira); **Centro**

Hospitalar Universitário São João, Porto, Portugal (L Garrido RN, Prof S Castedo, Prof F Carneiro); Serviço de Medicina Interna, Centro Hospitalar Tâmega e Sousa, Penafiel, Portugal (H Pinheiro); Escola de Economia e Gestão, Universidade do Minho, Braga, Portugal (L Sousa); Bioinf2Bio, Porto, Portugal (J J Maqueda); Department of Laboratory Genetics, Portuguese Oncology Institute of Porto, Porto, Portugal (C Pinto MSc, M R Teixeira); Porto Comprehensive Cancer Center Raquel Seruca, Porto, Portugal (J Garcia-Pelaez, R Barbosa-Matos, S Lobo, A Dias, Prof S Castedo, S Sousa, H Pinheiro, L Sousa, C Lemos, Prof F Carneiro, N Pinto, C Pinto, M R Teixeira, Prof C Oliveira); Institute of Human Genetics, Medical Faculty, University of Bonn, Bonn, Germany (Prof S Aretz MD PhD, I Spier MD); Department of Internal Medicine I (R Hüneburg MD) and National Center for Hereditary Tumor Syndromes (Prof S Aretz, R Hüneburg, I Spier), University Hospital Bonn, Bonn, Germany; Department of Oncology-Pathology, Karolinska Institutet, Stockholm, Sweden (S Bajalica-Lagercrantz MD PhD, K Lagerstedt-Robinson PhD); Department of Clinical Genetics, Cancer Genetic Unit (S Bajalica-Lagercrantz, K Lagerstedt-Robinson) and Cancer Theme (S Bajalica-Lagercrantz), Karolinska University Hospital Solna, Stockholm, Sweden; Hospital Vall d'Hebron, Universitat Autonoma de Barcelona, Barcelona, Spain (J Balmaña MD PhD); Department of Molecular Diagnostics (Prof S Novaković PhD) and Department of Clinical Cancer Genetics (A Blatnik MD, M Krajc MD PhD), Institute of Oncology Ljubljana, Ljubljana, Slovenia; Medical Genetics Department, Pitié-Salpêtrière Hospital, AP-HP and Sorbonne University, Paris, France (P R Benusiglio MD, C Desseigné MSc, F Coulet PharmD PhD); Service de Génétique Oncologique, Institut Curie, Paris, France (L Golmard PharmD PhD, M Blanluet PharmD, C Colas MD PhD); Laboratory of

Funding European Reference Network on Genetic Tumour Risk Syndromes, European Regional Development Fund, Fundação para a Ciência e a Tecnologia (Portugal), Cancer Research UK, and European Union's Horizon 2020 research and innovation programme.

Copyright © 2022 Elsevier Ltd. All rights reserved.

Introduction

Hereditary diffuse gastric cancer (HDGC; OMIM number #137215) is an autosomal dominant tumour risk syndrome with increased predisposition to develop early onset diffuse gastric cancer and lobular breast cancer.[1–3] HDGC is mainly caused by germline single-nucleotide variants and copy-number variants in the E-cadherin (*CDH1*) gene (Ensembl gene ID ENSG00000039068; RefSeq ID NM_004360.4; Locus Reference Genomic ID LRG_301),[4,5] which are classified as pathogenic or likely pathogenic variants, according to the American College of Medical Genetics and Genomics and Association for Molecular Pathology *CDH1* variant curation guidelines.[6] The International Gastric Cancer Linkage Consortium (IGCLC) has published four sets of clinical criteria for *CDH1* testing to date (appendix p 1).[1,3,7] Although guidelines from 1999 supported *CDH1* genetic testing only in families with multiple cases of diffuse gastric cancer, the 2010 and 2015 guidelines widened testing to isolated patients with diffuse gastric cancer, aged 40 years or younger, and to individuals or families with both diffuse gastric cancer and lobular breast cancer before 50 years of age.[1,7] Phenotype-driven HDGC guidelines were superseded by genotype-driven guidelines in 2020, with clinical selection criteria being reformulated predominantly on the basis of actionable genetic test results (presence of a *CDH1* pathogenic or likely pathogenic variant).[2] The 2020 guideline proposed hereditary lobular breast cancer (HLBC) as a lobular breast cancer only predisposition syndrome, independent of HDGC (appendix p 1).[2,8]

In a large US cohort of probands (tested because of HDGC clinical ascertainment or multigene panel testing), individuals carrying *CDH1* pathogenic or likely pathogenic variants were calculated to have a lifetime risk of gastric cancer of 42% for men and 33% for women, whereas female breast cancer cumulative incidence was

Research in context

Evidence before this study

In 2018, we started a genotype-first study to define the cancer landscape in families carrying *CDH1* pathogenic or likely pathogenic variants and missense variants of unknown significance, assess the frequency of families with lobular breast cancer among families carrying pathogenic or likely pathogenic variants of *CDH1*, and test the performance of different sets of hereditary diffuse gastric cancer (HDGC) clinical criteria driving *CDH1* testing. For the latter, we considered different sets of HDGC criteria developed between 2015 and 2022 (2015, 2020, Yale—2022) for comparison with the lobular breast cancer-expanded criteria herein proposed. We searched PubMed for studies in English on "*CDH1* germline variants" and "diffuse gastric cancer" or "lobular breast cancer" or "LBC" or "hereditary diffuse gastric cancer" or "HDGC" from Jan 1, 2015 onwards (the year of publication of the HDGC policy review and the first rules for variant classification in the clinical context, by the American College of Medical Genetics and Genomics and Association for Molecular Pathology). References from relevant articles, HDGC guidelines papers before 2015, reviews, and previous meta-analyses were considered on a case-by-case basis to identify any additional or relevant study not retrieved by the PubMed search. In 2020, a new HDGC policy review highlighted the lack of genotype–phenotype correlations in *CDH1* variant carriers, which still vastly persists today, particularly regarding *CDH1*-related predisposition to lobular breast cancer, associations between missense *CDH1* variants of unknown significance and HDGC-related phenotypes, and *CDH1* variant-specific cancer predisposition. This knowledge gap provides the rationale for designing more accurate clinical criteria and for management of families with HDGC.

Added value of this study

To our knowledge, this is the first multicentre genotype–phenotype study and the largest dataset ever studied of individuals carrying rare *CDH1* variants and their relatives. It highlights specificities in *CDH1* variant-type related clinical phenotypes and supports the association of HDGC-related cancers with the presence of *CDH1* pathogenic or likely pathogenic variants, but did not find an association of HDGC-related cancers with the presence of *CDH1* missense variants of unknown significance. The study also demonstrates that families not fulfilling HDGC clinical criteria, but carrying a *CDH1* PV/LPVs, are mainly lobular breast cancer-enriched, supporting an expansion of current HDGC criteria to include additional lobular breast cancer-centred criteria.

Implications of all the available evidence

Given the lack of evidence supporting a positive association between the presence of *CDH1* germline missense variants and HDGC-related cancers, carriers of such variants are unlikely to be predisposed to HDGC. The recurrent identification of pathogenic or likely pathogenic variants in patients and families with lobular breast cancer, with or without gastric cancer, supports the expansion of clinical criteria for *CDH1* testing in this setting. The observations herein reported will contribute to clarify the context for clinical management of families carrying *CDH1* germline variants.

estimated at 55% at age 80 years.[9] These cumulative incidences are considerably lower for gastric cancer and higher for breast cancer compared with cohorts strictly fulfilling HDGC criteria and those from other world regions.[10–13] Given the high risk of cancer in carriers of *CDH1* pathogenic or likely pathogenic variants, current guidelines recommend intensive gastric and breast surveillance, prophylactic gastrectomy in men and women, and optional mastectomy in women, after a *CDH1* pathogenic or likely pathogenic variant is found in a family or individual with relevant history of cancer in these organs.[2]

Many *CDH1* pathogenic or likely pathogenic variants, found by multigene panel testing, occur in breast cancer patients not meeting the clinical criteria for HDGC.[12,14] Additionally, a 2022 cohort study of patients with cancer showed that *CDH1* germline pathogenic or likely pathogenic variants are enriched in patients with diffuse gastric cancer and lobular breast cancer.[15] The high number of patients with breast cancer carrying *CDH1* pathogenic or likely pathogenic variants in the USA led Lerner and colleagues[16] to propose the use of clinical criteria for hereditary breast and ovarian cancer, in addition to the modified and less restrictive 2020 criteria for HDGC (also known as Yale criteria), to select patients for *CDH1* testing. Although this approach increased the sensitivity of the criteria, no data were provided regarding their specificity or positive predictive value.[16]

CDH1 variants of unknown significance persist as a clinical challenge because of scarce genotype–phenotype data, and missense variants, the largest class of coding variants of unknown significance, have been claimed to preferentially predispose individuals to HLBC.[8] In parallel, endoscopic evidence of gastric cancer was found in more than 90% of families with HLBC, who carry *CDH1* pathogenic or likely pathogenic variants[17], raising questions about whether HLBC is an HDGC-independent entity. Although colorectal cancer has been considered as a possible *CDH1*-associated phenotype in some studies,[15,18] others claim it is rare among carriers of *CDH1* pathogenic or likely pathogenic variants.[9]

Genotype–phenotype association studies, particularly genotype-first studies, are expected to answer some of the abovementioned challenges. In this study, we aimed to conduct a genotype–phenotype analysis in carriers of rare *CDH1* variants and their relatives, to compare the cancer landscape in families carrying pathogenic or likely pathogenic variants (considered together for the purpose of our analyses) or missense variants of unknown significance, to assess the frequency of families with lobular breast cancer among families carrying pathogenic or likely pathogenic variants, and to test the performance of novel lobular breast cancer-expanded criteria for the selection of patients for *CDH1* testing. We used the collaborative environment of the European Reference Network on Genetic Tumour Risk Syndromes (ERN GENTURIS) to run a multicentre genotype-first study to identify associations between rare *CDH1* germline variants and cancer phenotypes, in the largest dataset of *CDH1* rare variant carriers and their relatives, the first and most detailed to date, to the best of our knowledge.

Methods

Participants and data collection

Data collection started on Oct 1, 2018, and ended on Sept 20, 2022. *CDH1* variant-related data were obtained from diagnostic laboratories across Europe (29 institutions in ten European countries; figure 1A; appendix pp 2–4) and were the result of genetic testing with either *CDH1*-targeted sequencing, or panel, whole-exome, or whole-genome sequencing. Only variants in the *CDH1* locus were collected for the purpose of the current study. The presence of a rare single-nucleotide variant or copy-number variant in *CDH1*, herein classified as any variant with frequency below 0·01 in gnomAD, in a proband irrespective of clinical diagnosis or HDGC suspicion, drove the collection of aggregated molecular and clinical information from carrier probands and their relatives (unconfirmed carriers). The term unconfirmed carrier refers to two different scenarios: affected relatives who were not submitted for carrier testing, particularly in the case of variants of unknown significance and likely benign or benign variants; and affected relatives from families carrying pathogenic or likely pathogenic variants who were generally submitted for carrier testing, but for whom the carrier status is generally unknown to this study. The cohort's general clinical features (appendix pp 2–3), information on participating institutions (appendix p 4), and molecular and clinical information indexed to probands from each family (appendix pp 5–46) were collected.

All patients signed an informed consent form for germline testing, either mentioning future research related to their susceptibility syndrome or allowing sample biobanking for future research. In most institutions, approval by ethics committees was waived, as this is a retrospective study; in the remaining institutions, local ethics approval was granted. The whole project received appraisal N19/CECRI/2022 from the Committee for Ethical and Responsible Conduct of Research of the study's leading institution, Instituto de Investigação e Inovação em Saúde, University of Porto, Portugal.

Clinical and variant data curation

Family history was revisited and compliance with the 2015 criteria,[1] 2020 criteria,[2] or Yale criteria[16] for HDGC (appendix pp 5–46) was registered. For the purpose of comparative analysis, a family with HDGC was defined as any family fulfilling the 2015 clinical criteria for HDGC. Phenotypes were collected from probands and relatives from families carrying *CDH1* germline variants.

Human Genetics, GIGA Institute, University of Liège, Liège, Belgium (V Bours MD PhD); **Center of Genetics, University Hospital, Liège, Belgium** (V Bours); **Department of Human Genetics, University of Leuven, Leuven, Belgium** (H Brems PhD, G Michils PhD); **Hereditary Cancer Programme, Catalan Institute of Oncology, Bellvitge Institute for Biomedical Research and Girona Biomedical Research Institute, Barcelona-Girona, Spain** (J Brunet MD PhD); **Laboratorio di Bioscienze, IRCCS Istituto Romagnolo per lo Studio dei Tumori (IRST) "Dino Amadori", Meldola, Italy** (D Calistri PhD, G Tedaldi PhD); **Hereditary Cancer Program, Catalan Institute of Oncology, Bellvitge Institute for Biomedical Research, Barcelona, Spain** (G Capellá MD PhD, C Lázaro PhD); **Centro de Investigación Biomédica en Red de Cáncer (CIBERONC), Madrid, Spain** (G Capellá, C Lázaro); **Oncology Service, Biocruces Bizkaia Health Research Institute, Cruces University Hospital, Cruces-Barakaldo, Bizkaia, Spain** (S Carrera MD); **Center of Human Genetics, IPG, Gosselies, Belgium** (K Dahan MD PhD, D Feret MSc); **Clinical Genetics Department, University Hospital of Ghent, Ghent, Belgium** (R de Putter MD); **Molecular Diagnostics Laboratory, Fundación Rioja Salud, Logroño, Spain** (E Domínguez-Garrido PhD); **CNC—Center for Neuroscience and Cell Biology, University of Coimbra, Coimbra, Portugal** (C Egas PhD); **Division of Evolution and Genomic Sciences, University of Manchester, Manchester, UK** (Prof D G Evans MD, E R Woodward MD PhD); **Manchester Centre for Genomic Medicine, Manchester, UK** (Prof D G Evans, E R Woodward); **Department of Medical Genetics, National Institute for Health Research Cambridge Biomedical Research Centre** (E Fewings PhD, Prof M Tischkowitz MD PhD), **and Early Cancer Institute** (Prof R C Fitzgerald MD), **University of Cambridge, Cambridge, UK; Genetics Unit, Biocruces Bizkaia Health**

Research Institute, Basurto University Hospital, Bilbao, Bizkaia, Spain (M Garcia-Barcina MD PhD, S Merino MD); Sezione di Medicina Genomica, Dipartimento di Scienze della Vita e Salute Pubblica, Università Cattolica del Sacro Cuore, Rome, Italy (Prof M Genuardi MD); UOC Genetica Medica, Dipartimento di Scienze di Laboratorio e Infettivologiche, Fondazione Policlinico Universitario A Gemelli IRCCS, Rome, Italy (Prof M Genuardi); Institute for Clinical Genetics, University Hospital Carl Gustav Carus and Faculty of Medicine, Technische Universität Dresden, Dresden, Germany (K Hackmann PhD, Prof E Schröck MD); National Center for Tumor Diseases, Dresden, Germany: German Cancer Research Center, Heidelberg, Germany (K Hackmann, Prof E Schröck); Helmholtz-Zentrum Dresden-Rossendorf, Dresden, Germany (K Hackmann, Prof E Schröck); German Cancer Consortium, Dresden, Germany (K Hackmann, Prof E Schröck); Max Planck Institute of Molecular Cell Biology and Genetics, Dresden, Germany (Prof E Schröck); SouthWest Thames Regional Genetics Service, St George's University Hospitals NHS Foundation Trust, London, UK (H Hanson MD); Medizinische Klinik und Poliklinik IV, Klinikum der Universität München, Munich, Germany (Prof E Holinski-Feder MD, V Steinke-Lange MD); Medizinisch Genetisches Zentrum, Munich, Germany (Prof E Holinski-Feder, V Steinke-Lange); Department of Human Genetics, Radboud University Medical Center, Nijmegen, Netherlands (Prof M J L Ligtenberg MD, Prof N Hoogerbrugge MD PhD); Department of Pathology (Prof M J L Ligtenberg) and Radboud Institute of Molecular Life Sciences (Prof M J L Ligtenberg, Prof N Hoogerbrugge), Radboud University Medical Center, Nijmegen, Netherlands; Genetics Service, Biocruces Bizkaia Health Research Institute, Cruces University Hospital, Cruces-Barakaldo, Bizkaia, Spain

Some individuals contributed more than one phenotype to the analysis if they had more than one phenotype. Information on age of cancer onset and gastric and breast cancer histology was collected (appendix pp 2–3).

Diffuse gastric cancer, lobular breast cancer, and gastric cancer were considered HDGC-related phenotypes. "Other" refers to all phenotypes that differed from HDGC-related phenotypes, breast cancer of unknown histotype (herein referred to as "non-specified breast cancer" or "breast cancer"), ductal breast cancer, ovarian cancer, and colorectal cancer (appendix p 3).

In the context of families, "gastric cancer only" refers to families enriched in gastric cancer, without breast cancer of any type, and could include cancers other than gastric or breast cancer; "breast cancer only" refers to families enriched in breast cancer, without gastric cancer of any type, but could include cancers other than gastric or breast cancer; "gastric cancer and breast cancer" refers to families with both gastric cancer and breast cancer, and could also include cancers other than gastric or breast cancer. "Other cancer" refers to families presenting exclusively cancers other than gastric or breast cancer.

Variant nomenclature was standardised according to the Human Genome Variation Society (HGVS) guidelines.[19] Variants were categorised into groups: coding truncating (including nonsense, frameshift, start loss, large deletion, large duplication, and splice-site variants); coding non-truncating (including missense, in-frame insertion, and in-frame exon skipping variants); non-coding (regulatory; including intronic large deletions, 5‾ untranslated region single-nucleotide variants, 3_ untranslated region single-nucleotide variants, and intronic single-nucleotide variants); and synonymous (referring to single-nucleotide variants unrelated to splicing), according to the Ensembl Variant Effect Predictor (VEP;[20] appendix p 47). Variants located until within two bp from the start or end of exons and within five bp from the start or end of introns were analysed with NetGene2 version 2.4 splice-site prediction software,[21] and classified as splice-site variants if a positive impact on splicing was predicted. Recurrence was considered if the same variant occurred in three or more families. With use of the American College of Medical Genetics and Genomics and Association for Molecular Pathology *CDH1* guidelines (version 2), all variants were classified for their actionability in HDGC as a pathogenic or likely pathogenic (herein analysed jointly and referred to in short as PV/LPV), a variant of unknown significance, or a likely benign or benign variant (herein analysed jointly and referred to in short as LBV/BV; figure 1A),[6] and were confirmed with ClinGen *CDH1* Variant Curation Expert Panel updates. In case a patient presented two *CDH1* germline variants, the most damaging variant was considered for genotype–phenotype analysis (appendix p 48). In the current study, only variants classified as either pathogenic or likely pathogenic were considered actionable, as their identification in probands triggers cascade testing in relatives and risk reduction measures in carrier probands and carrier relatives. Because the 2020 guidelines for HDGC recommend at least 2 years of gastric or breast surveillance, or both, for probands carrying a variant of unknown significance, but not for relatives, this type of variant was considered non-actionable in this study when assessing the performance of different HDGC clinical criteria.

The relative frequency of phenotypes, age of onset, and compliance with HDGC criteria were used in different comparisons involving families carrying different types of *CDH1* variants; families bearing recurrent missense variants of unknown significance or bearing recurrent truncating PV/LPVs; and families with gastric cancer only, breast cancer only, or gastric cancer and breast cancer.

The frequency of families carrying PV/LPVs among those complying with different HDGC criteria sets was used to evaluate the pick-up rate and performance of the 2015 and 2020 HDGC criteria, Yale criteria, and the lobular breast cancer-expanded criteria.

Statistical analysis

Continuous variables are presented as mean (SD). Phenotypes without an associated age of onset were excluded from statistical analyses. Multiple comparisons of age of disease onset in carrier probands and their relatives and different types of variants (PV/LPV, missense variant of unknown significance, or missense LBV/BV), were done using a Kruskal-Wallis test with Bonferroni correction. Categorical variables (phenotypes, compliance with HDGC criteria, and variant classification) are presented as count (%) and were compared with use of χ^2 tests. This approach was used to compare variant classes between families fulfilling or not fulfilling clinical criteria, and for the analysis of phenotypic distributions in families carrying different types of variants. Age of onset per phenotype was compared with use of an independent-samples Student's t test. Multivariable logistic regression models using the enter method, adjusting for phenotypes and age of onset to remove sampling bias in terms of age of onset, were used to identify independent predictors for distinct variant types (PV/LPVs, and missense variant of unknown significance). To evaluate the likelihood of each set of criteria producing false positive and false negative results, we calculated the sensitivity, specificity, positive predictive value, and negative predictive value, with corresponding 95% CIs, using R (version 4.1.3) with the library epiR (version 2.0.50),[23] whereas the Youden index (J) used to estimate the criteria's discriminating power, and the corresponding 95% CIs, were computed with the library ThresholdROC.[24] The term "lobular breast cancer-expanded criteria" represents a combination of all the 2020 HDGC criteria plus the new lobular-centred criteria herein proposed. To compare performance in terms of true positive or true negative detection for different sets of clinical criteria (2015 and 2020 HDGC criteria, Yale

criteria, and lobular breast cancer-expanded criteria), equivalence tests and corresponding 95% CIs were computed in R by directly implementing formulae. The equivalence limit (δ) was set to 0·12.[25] Areas under the receiver operating characteristic curves (AUCs) from different sets of clinical HDGC criteria (2015 and 2020 HDGC criteria, Yale criteria, and lobular breast cancer-expanded criteria) were compared against each other with *Z* tests, using the IBM SPSS Statistics for Windows (version 28.0.0). For a 95% CI, a *Z* score of 1·96 was used. Statistical significance was set at p<0·05 for all statistical tests.

Role of the funding source

The funders of the study had no role in study design, data collection, data analysis, data interpretation, or writing of the report.

Results

398 different rare *CDH1* germline variants were reported in 854 families (figure 1A; appendix pp 5–46). From 854 carrier probands and 1021 relatives, age of onset was available for 1495 patients' phenotypes and ranged from 1–93 years. Altogether, 1971 phenotypes were collected, of which 460 were gastric and breast cancers with available histology (figure 1A; appendix pp 2–3).

Among the 854 families, coding truncating variants occurred in 184 (22%), missense variants in 316 (37%), and other variant types in 354 families (41%; appendix p 47). In terms of variant clinical actionability,[6] 176 (21%) families carried PV/LPVs (100 different truncating variants); 169 (20%) families carried missense variants of unknown significance (109 different missense variants); 147 (17%) families carried missense LBV/BVs (29 different missense variants); 182 (21%) families carried variants of unknown significance other than missense (122 different variants); and 180 (21%) families carried LBV/BVs other than missense (38 different variants) (appendix pp 47, 49–51). Probands carrying missense variants of unknown significance or missense LBV/BVs, and their relatives, were all significantly older than probands carrying PV/LPVs and their relatives (p<0·05 in all comparisons), whereas probands carrying missense variants of unknown significance and missense LBV/BVs and their relatives had similar ages of onset (appendix p 52).

CDH1 PV/LPVs included both single-nucleotide variants and copy-number variants and were distributed across the whole *CDH1* locus, targeting all exons, as were all other variant types (appendix pp 49–51). Only 15 PV/LPVs (11 single-nucleotide variants and four copy-number variants) were recurrent, and were found in 81 (46%) of 176 families (appendix p 49). The most frequent PV/LPV single-nucleotide variant, which occurred in 13 families, was the c.1901C>T (p.Ala634ProfsTer7) splice-site variant, and the most frequent copy-number variant, occurring in ten families, was the exon 1–2 deletion NC_000016.10:g.(?_68737292)_(68738411_?)del (appendix p 49).

182 (21%) families fulfilled the 2015 clinical criteria for HDGC[1] and 672 (79%) did not. The cohort of families fulfilling the 2015 criteria was significantly enriched for PV/LPVs compared with the cohort of families not fulfilling the criteria (136 [75%] of 182 *vs* 40 [6%] of 672; p<0·0001; figure 1B; appendix p 53). By contrast, missense variants of unknown significance were enriched in the cohort of families not fulfilling the criteria (155 [23%] of 672; these 155 families are part of 159 carrying coding non-truncating variants in families not fulfilling criteria; figure 1B; appendix p 54) compared with the cohort of families that fulfilled the criteria (14 families carrying coding non-truncating variants [8%] of 182 families fulfilling criteria; p<0·0001; figure 1B; appendix p 54).

Early onset diffuse gastric cancer, lobular breast cancer, and gastric cancer, the classic HDGC-related phenotypes, were the most frequent among 176 PV/LPV carrier families, accounting for 459 (73%) of 631 phenotypes in probands and relatives (figure 2A; appendix p 55). HDGC-related phenotypes were repeatedly observed only among families bearing recurrent PV/LPVs, but not among families bearing recurrent missense variants of unknown significance (appendix p 56).

Among 176 families carrying PV/LPVs, 362 (92%) of 392 diffuse gastric cancer or gastric cancer cases occurred in families fulfilling the 2015 criteria for HDGC, whereas only 32 (48%) of 67 cases of lobular breast cancer occurred in families fulfilling the criteria (figure 2A; appendix p 55). Among families carrying missense variants of unknown significance, 32 (65%) of 49 diffuse gastric cancer or gastric cancer cases occurred in families fulfilling the HDGC criteria, and among families carrying missense LBV/BVs, this proportion was 18 (38%) of 48. Lobular breast cancers were extremely rare among families carrying missense variants of unknown significance or LBV/BVs (figure 2B; appendix pp 55, 57).

Among families carrying PV/LPVs, 50 (56%) of 90 non-specified breast cancer cases (mainly in relatives) occurred in families fulfilling the criteria (figure 2A; appendix pp 55, 58). By contrast, among families carrying missense variants of unknown significance, only one (<1%) of 219 cases of non-specified breast cancer occurred in families fulfilling the HDGC criteria, and among families carrying missense LBV/BVs, this proportion was three (2%) of 150 cases (figure 2B; appendix pp 55, 57–58).

Overall, in families fulfilling HDGC criteria, the frequency of either gastric cancers altogether or non-specified breast cancer in relatives was significantly higher in carriers of PV/LPVs than in carriers of missense variants of unknown significance or carriers of LBV/BVs (p<0·0001 in all comparisons; figure 2A, B; appendix pp 55, 57–58).

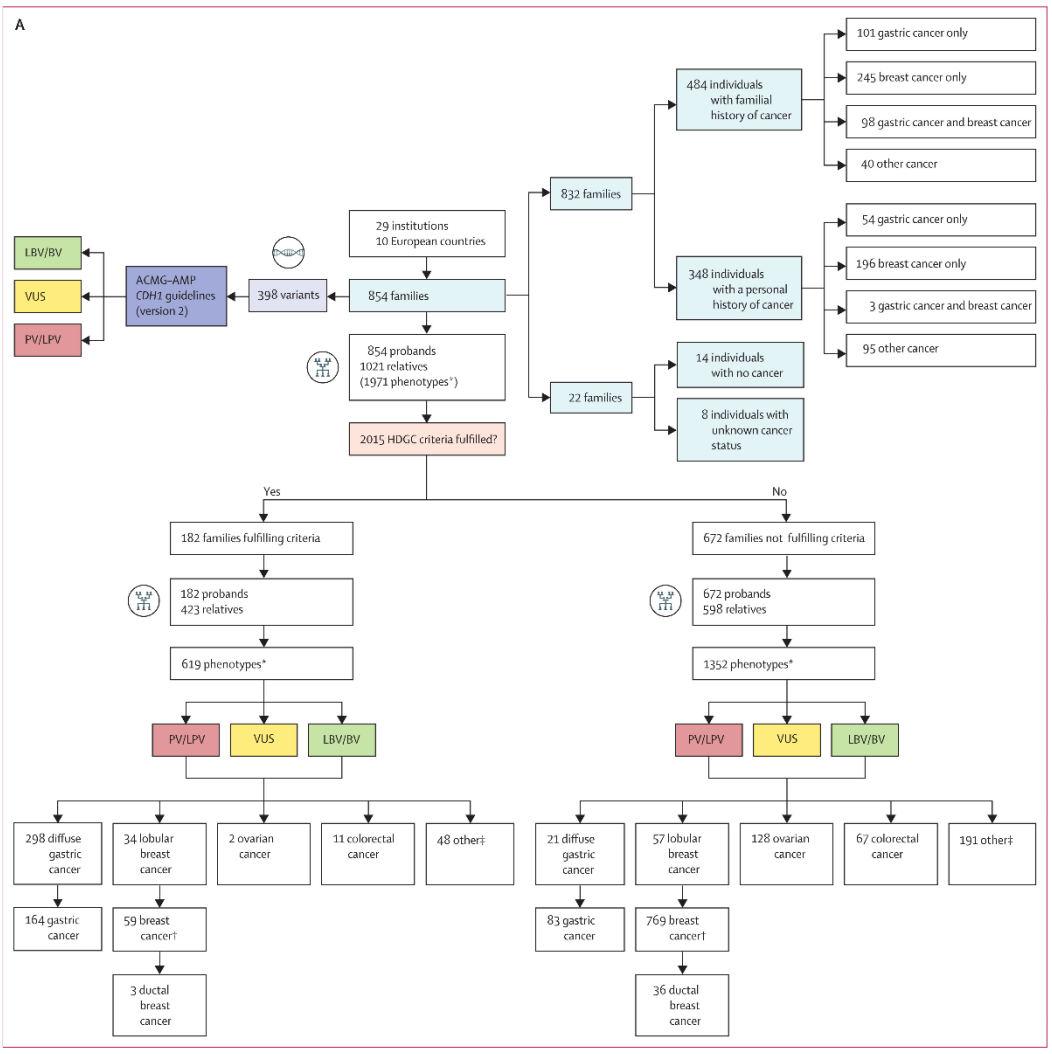
Multivariate logistic regression was used to evaluate the likelihood of occurrence of HDGC-related cancers in PV/LPV carrier families as compared with their occurrence in families carrying missense variants

(C Martínez-Bouzas PhD, M I Tejada PhD); **Unidad de Medicina Genómica y Pediatría, Clínica Universidad de Navarra, Programa de Tumores Sólidos, Centro de Investigación Médica Aplicada, Instituto de Investigación Sanitaria de Navarra, Pamplona, Navarra, Spain** (Prof A Patiño-García PhD); **Department of Biology and Biotechnology, University of Pavia, Pavia, Italy** (Prof G N Ranzani BSc PhD); **GenoMed—Diagnósticos de Medicina Molecular, Lisbon, Portugal** (I Silva MSc, C Silveira MSc); **Molecular Genetics Laboratory, Elche University Hospital, Elche, Spain** (J L Soto BSc PhD); **European Reference Network on Genetic Tumour Risk Syndromes (ERN GENTURIS), Porto, Portugal** (Prof S Castedo, Prof F Carneiro, M R Teixeira, Prof C Oliveira); **ERN GENTURIS, Bonn, Germany** (Prof S Aretz, R Hüneburg, I Spier); **ERN GENTURIS, Stockholm, Sweden** (S Bajalica-Lagercrantz, K Lagerstedt-Robinson); **ERN GENTURIS, Barcelona, Spain** (J Balmaña, J Brunet, G Capellá, C Lázaro); **ERN GENTURIS, Ljubljana, Slovenia** (A Blatnik, M Krajc); **ERN GENTURIS, Liège, Belgium** (V Bours); **ERN GENTURIS, Paris, France** (C Colas); **ERN GENTURIS, Ghent, Belgium** (R de Putter); **ERN GENTURIS, Rome, Italy** (Prof M Genuardi); **ERN GENTURIS, Munich, Germany** (Prof E Holinski-Feder, V Steinke-Lange); **ERN GENTURIS, Nijmegen, Netherlands** (Prof M J L Ligtenberg, Prof N Hoogerbrugge); **ERN GENTURIS, Dresden, Germany** (Prof E Schröck)

Correspondence to: Prof Carla Oliveira, Instituto de Investigação e Inovação em Saúde (i3S), University of Porto, Porto 4200-135, Portugal **carlaol@i3s.up.pt**

See Online for appendix

of unknown significance (figure 2C). This analysis highlighted positive associations between the classic HDGC-related cancers (lobular breast cancer, diffuse gastric cancer, and gastric cancer) and the presence of PV/LPVs, in contrast to missense variants of unknown significance. Lobular breast cancer (a phenotype present in 47 (27%) of 176 PV/LPV carrier families and in four (2%) of 169 families carrying missense variants of unknown significance) had the greatest positive association (odds ratio [OR] 12·39 [95% CI 2·66–57·74], p=0·0014), followed by diffuse gastric cancer (8·00 [2·18–29·39], p=0·0017), and gastric cancer (7·81 [2·03–29·96], p=0·0027; figure 2C; appendix p 59). Cancers, such as non-specified breast cancer, ovarian



(Figure 1 continues on next page)

cancer, colorectal cancer, and all other cancers together, were not positively associated with the presence of a PV/LPV (figure 2C). The likelihood of ovarian cancer occurring in families carrying missense variants of unknown significance was significantly higher than in families carrying PV/LPVs (OR 0·12 [95% CI 0·02–0·71], p=0·019; figure 2C; appendix p 59). In the multivariate analysis, age of cancer onset did not differ with regard to cancer type or between families carrying PV/LPVs and missense variants of unknown significance (figure 2C; appendix p 59).

We next analysed the evolution of HDGC clinical criteria, considering the above genotype–phenotype associations. From 176 PV/LPV carrier families, 40 (23%) did not meet the 2015 criteria. Of these families, 11 (6%) fulfilled the 2020 HDGC criteria that considers eligible for *CDH1* testing isolated cases of diffuse gastric cancer in people younger than 50 years and those families with at least two cases of lobular breast cancer in family members younger than 50 years (appendix p 60). From the remaining 29 PV/LPV families who did not meet the 2020 HDGC criteria, only four families did not show breast cancer involvement (figure 3A; appendix p 61). 18 (62%) of 29 were families with lobular breast cancer with or without gastric cancer (figure 3B). Detailed analysis of family cancer history revealed that these 18 families presented at least one confirmed case of lobular breast cancer, either isolated before age 55 years (four of 18), in families with additional non-specified breast cancers and no gastric involvement (seven out of 18), or in families with history of non-specified breast cancer and gastric cancer (seven out of 18). This analysis resulted in a proposal of three new lobular breast cancer-centred criteria to be added to the 2020 HDGC criteria, and herein called lobular breast cancer-expanded criteria (figure 3B).

For 11 of 29 families who did not meet the 2020 HDGC criteria, representing 6% of 176 *CDH1* PV/LPV carrier families, no possible criteria could be envisioned due to very scarce or missing clinical information (figure 3B; appendix p 61).

We next stratified the full PV/LPV carrier cohort into family subgroups of gastric cancer only, gastric cancer and breast cancer, and breast cancer only, and plotted the relative phenotype frequency and distribution in each family subgroup. This approach allowed evaluation of the contribution of the lobular breast cancer-centred criteria in the different subgroups (figure 4; appendix pp 62–68).[1,2]

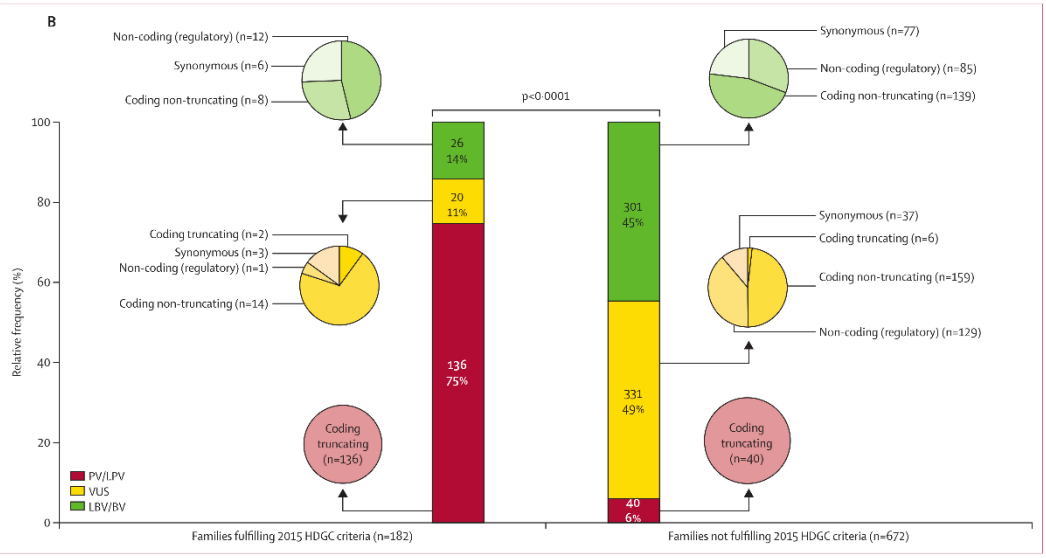
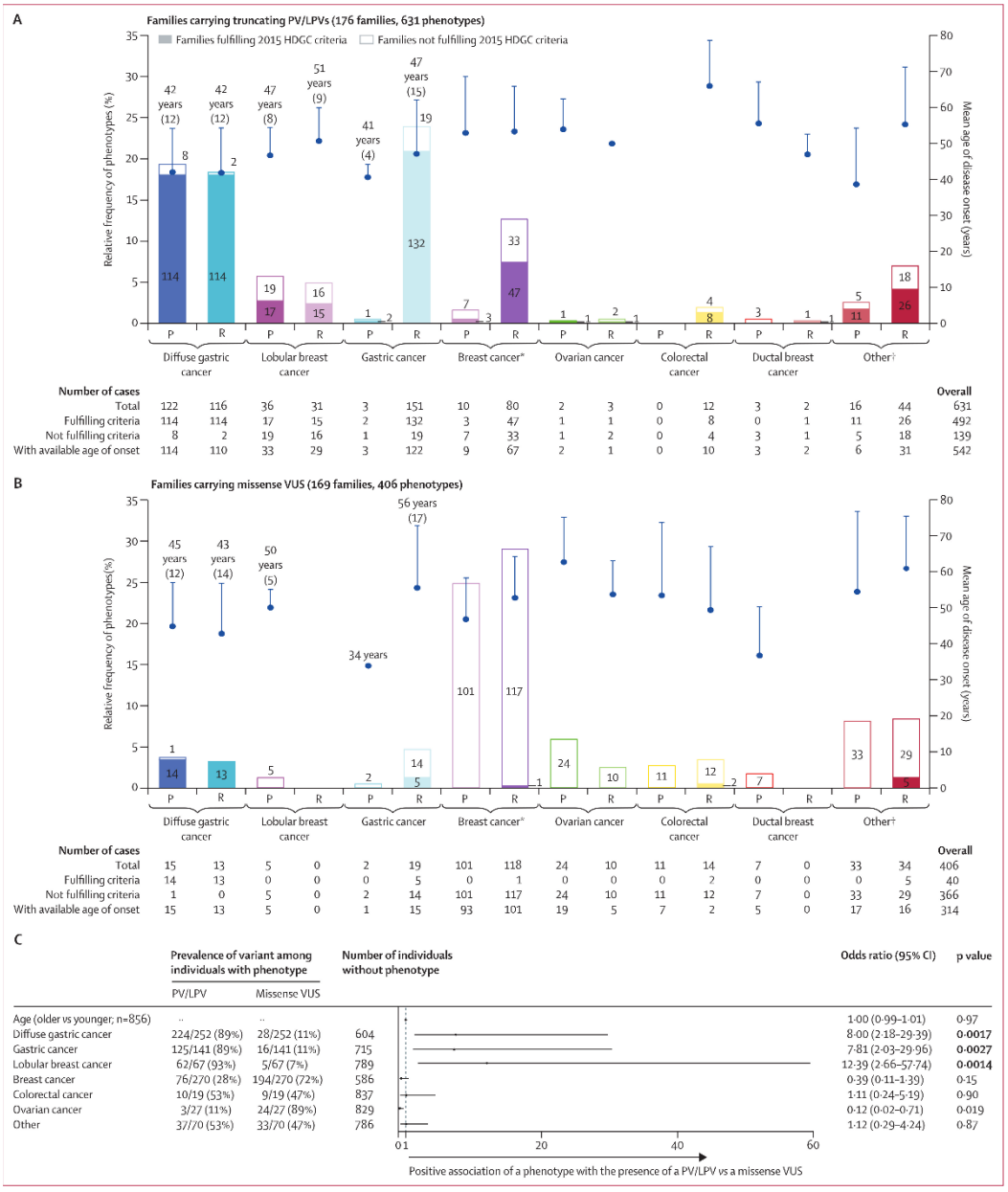


Figure 1: **Flowchart of the genotype–phenotype study and molecular and clinical features of the cohort of families with rare germline *CDH1* variants**
(A) Flowchart of the genotype–phenotype study. (B) Relative frequency of variant groups and clinical classification in families fulfilling or not fulfilling the 2015 HDGC criteria. A χ^2 test was used to compare the frequency of variant clinical classifications between families fulfilling and not fulfilling the 2015 HDGC criteria (appendix p 53). 14 of 14 coding non-truncating variants of unknown significance were missense variants occurring in families fulfilling the 2015 HDGC criteria. 155 of 159 coding non-truncating variants of unknown significance were missense variants occurring in families not fulfilling the 2015 HDGC criteria. Details on variant type and correspondence with clinical criteria and clinical classification are depicted in appendix (pp 47, 54). ACMG–AMP=American College of Medical Genetics and Genomics and Association for Molecular Pathology. HDGC=hereditary diffuse gastric cancer. LBV/BV=likely benign or benign variant. PV/LPV=pathogenic or likely pathogenic variant. VUS=variant of unknown significance. *Probands could contribute more than one phenotype. †Breast cancer of unknown histotype. ‡Other phenotypes not including gastric cancer, breast cancer, colorectal cancer, or ovarian cancer (detailed in appendix p 3).


A
Families carrying truncating PV/LPVs (176 families, 631 phenotypes)
Families fulfilling 2015 HDGC criteria
Families not fulfilling 2015 HDGC criteria
Relative frequency of phenotypes (%)
Mean age of disease onset (years)
42 years (12)
42 years (12)
47 years (8)
51 years (9)
41 years (4)
47 years (15)
Diffuse gastric cancer
Lobular breast cancer
Gastric cancer
Breast cancer*
Ovarian cancer
Colorectal cancer
Ductal breast cancer
Other†
Number of cases
Overall
Total 122 116 36 31 3 151 10 80 2 3 0 12 3 2 16 44 631
Fulfilling criteria 114 114 17 15 2 132 3 47 1 1 0 8 0 1 11 26 492
Not fulfilling criteria 8 2 19 16 1 19 7 33 1 2 0 4 3 1 5 18 139
With available age of onset 114 110 33 29 3 122 9 67 2 1 0 10 3 2 6 31 542
B
Families carrying missense VUS (169 families, 406 phenotypes)
Relative frequency of phenotypes(%)
Mean age of disease onset (years)
45 years (12)
43 years (14)
50 years (5)
34 years
56 years (17)
Diffuse gastric cancer
Lobular breast cancer
Gastric cancer
Breast cancer*
Ovarian cancer
Colorectal cancer
Ductal breast cancer
Other†
Number of cases
Overall
Total 15 13 5 0 2 19 101 118 24 10 11 14 7 0 33 34 406
Fulfilling criteria 14 13 0 0 0 5 0 1 0 0 0 2 0 0 0 5 40
Not fulfilling criteria 1 0 5 0 2 14 101 117 24 10 11 12 7 0 33 29 366
With available age of onset 15 13 5 0 1 15 93 101 19 5 7 2 5 0 17 16 314
C
Prevalence of variant among individuals with phenotype
PV/LPV
Missense VUS
Number of individuals without phenotype
Odds ratio (95% CI)
p value
Age (older vs younger; n=856) .. .. 1·00 (0·99–1·01) 0·97
Diffuse gastric cancer 224/252 (89%) 28/252 (11%) 604 8·00 (2·18–29·39) 0·0017
Gastric cancer 125/141 (89%) 16/141 (11%) 715 7·81 (2·03–29·96) 0·0027
Lobular breast cancer 62/67 (93%) 5/67 (7%) 789 12·39 (2·66–57·74) 0·0014
Breast cancer 76/270 (28%) 194/270 (72%) 586 0·39 (0·11–1·39) 0·15
Colorectal cancer 10/19 (53%) 9/19 (47%) 837 1·11 (0·24–5·19) 0·90
Ovarian cancer 3/27 (11%) 24/27 (89%) 829 0·12 (0·02–0·71) 0·019
Other 37/70 (53%) 33/70 (47%) 786 1·12 (0·29–4·24) 0·87
0 1 20 40 60
Positive association of a phenotype with the presence of a PV/LPV vs a missense VUS

174 of 176 families fit into the following subgroups: 91 families with gastric cancer only (representing 52% of all families carrying PV/LPVs) and 255 phenotypes; 67 families with gastric cancer and breast cancer (representing 38% of all families carrying PV/LPVs) and 331 phenotypes, and; 16 families with breast cancer only (representing 9% of all families carrying PV/LPVs) and 43 phenotypes (figure 4; appendix pp 62–68). In families with gastric cancer only, 230 (90%)of 255 phenotypes are diffuse gastric cancer and gastric cancer. Of the 91 of families with gastric cancer only, 89 (98%) fulfilled the 2015 or 2020 HDGC criteria (figure 4A; appendix pp 62–68).[2] In families with gastric cancer and breast cancer, 162 (49%) of 331 phenotypes were diffuse gastric cancer and gastric cancer, 48 (15%) were lobular breast cancer, and 78 (24%) were non-specified breast cancer. 56 (84%) of 67 families fulfilled either the 2015 or 2020 HDGC criteria, and lobular breast cancer-centred criteria contributed additional seven (10%) of 67 families (figure 4B; appendix pp 62–68). From the 16 families with breast cancer only (figure 4C), 12 (7% of 176 families carrying PV/LPVs) presented lobular breast cancer only. In these 16 families, 19 (44%) of 43 phenotypes were lobular breast cancer, and 12 (28%) unspecified breast cancer. Two (13%) of 16 families fulfilled either the 2015 or 2020 HDGC criteria, and lobular breast cancer-centred criteria contributed additional 11 (69%) of 16 families (figure 4; appendix pp 62–68).

We next analysed whether specific PV/LPVs were more frequent or exclusive of subgroups of gastric cancer only, gastric cancer and breast cancer, or breast cancer only families (appendix pp 64–69). A single variant (c.1565+2dup) was recurrent and consistently associated with gastric cancer only families, and no variant was particularly associated with gastric cancer and breast cancer families (appendix pp 64–69). Variants occurring exclusively in families with breast cancer only occurred in a single family each (appendix pp 64–69).

Figure 2: **Distribution of phenotypes and cancer average age of onset, in probands and relatives from germline *CDH1* variant carrier families, and multivariable logistic regression models for the association of cancer phenotypes with PV/LPV and missense-variant of unknown significance**
Relative frequency of phenotypes, phenotype distribution, number of cases, and average age of onset in probands and relatives in families carrying PVs or LPVs molecularly classified as truncating (A) and in families carrying variants of unknown significance molecularly classified as missense (B) (data further detailed in appendix p 55). The left axis shows the relative frequency (%) of each phenotype out of the total number of phenotypes. Mean age of onset (SD) is shown for HDGC-associated phenotypes. Detailed data on the number of cases and average age of onset for all phenotypes are provided in the appendix (p 55). (C) Forest plot showing the results of a multivariable logistic regression model for the association of each phenotype with the presence of a PV/LPV *vs* a missense VUS in 856 patients with available age of disease onset and clinical phenotype. Probands could contribute more than one phenotype. HDGC=hereditary diffuse gastric cancer. PV/LPV=pathogenic or likely pathogenic variant. P=probands. R=relatives. VUS=variants of unknown significance. *Breast cancer of unknown histotype. †Other phenotypes not including gastric cancer, breast cancer, colorectal cancer, or ovarian cancer (detailed in appendix p 3).

We next plotted a summary of the 854 families, bearing variants classified according to the American College of Medical Genetics and Genomics and Association for Molecular Pathology *CDH1* variant classification (version 2), against the 2020 HDGC or the lobular breast cancer-expanded criteria to graphically represent improvement of clinical criteria in identifying PV/LPV carrier families (figure 5A). Further, the performance of different sets of HDGC criteria (2015, 2020, Yale, lobular breast cancer-expanded) was tested (figure 5B). For these tests, we considered PV/LPVs as actionable, and variants of unknown significance or LBV/BVs as non-actionable (appendix p 70).

The lobular breast cancer-expanded set of criteria showed better performance than the 2020 HDGC criteria (AUC 0·92 *vs* 0·88; *Z* score 3·54; p=0·0004), as well as better specificity (0·90 [95% CI 0·88–0·93] *vs* 0·34 [0·30–0·37]), and superiority in true negative detection (p<0·0001) compared with the Yale criteria[16] (specificity 0·34 [95% CI 0·30–0·37]; figure 5B, C; appendix pp 70–72), and equivalent true negative detection to that of the 2015 and 2020 HDGC criteria (p=0·059 and p=0·21, respectively; figure 5C). The lobular breast cancer-expanded criteria also showed better sensitivity (0·94 [95% CI 0·89–0·97]) than the 2015 HDGC criteria (0·77 [0·70-0·83] and the 2020 HDGC criteria (0·84 [0·77–0·89]), with superior true positive detection (p<0·0001 and p=0·0025, respectively), while maintaining high specificity (figure 5C; appendix pp 70–72).

Discussion

Our study shows that *CDH1* PV/LPVs are positively associated with HDGC-related phenotypes (lobular breast cancer, diffuse gastric cancer, and gastric cancer), and found no evidence of a positive association of these phenotypes with *CDH1* missense variants of unknown significance. As *CDH1* PV/LPVs occurred often in families with lobular breast cancer who did not meet the 2020 HDGC criteria, this study supports the expansion of HDGC criteria towards inclusion of additional lobular breast cancer-centred criteria.

The identification of *CDH1* pathogenic or likely pathogenic variants in patients with a relevant cancer history results in the application of clinical recommendations, including prophylactic removal of target organs in at-risk asymptomatic carriers. Therefore, it is crucial to know which *CDH1* variants are actionable and cause disease, whether variant classification is accurate, which organs are prone to cancer development in carriers, and if there are other histopathological cancer features specific to a particular *CDH1* variant, besides diffuse gastric cancer and lobular breast cancer. Answering these questions will improve clinical guidelines and ascertainment for genetic testing. Given that HDGC is a very rare tumour risk syndrome, ERN GENTURIS led a European multicentre study using a genotype-first

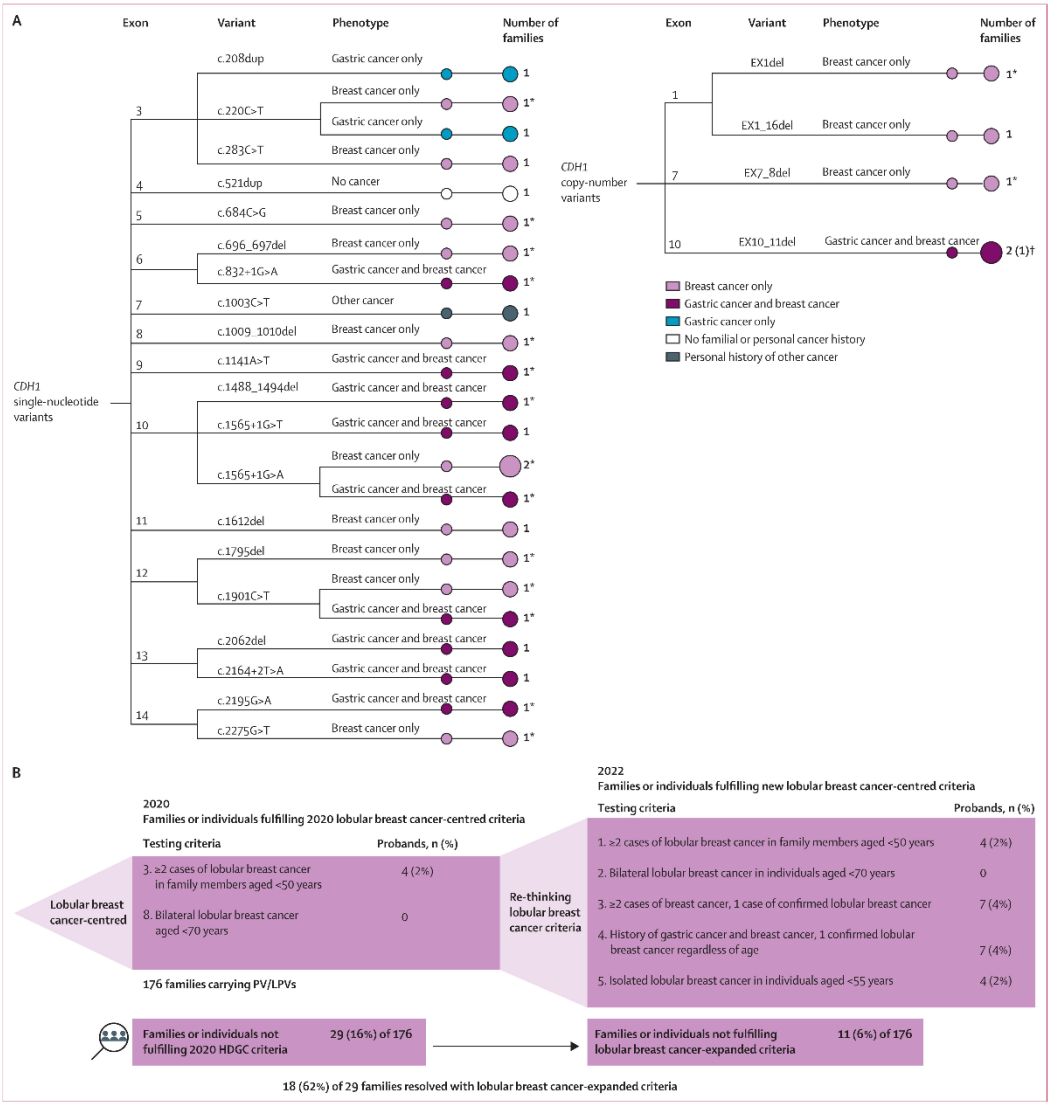
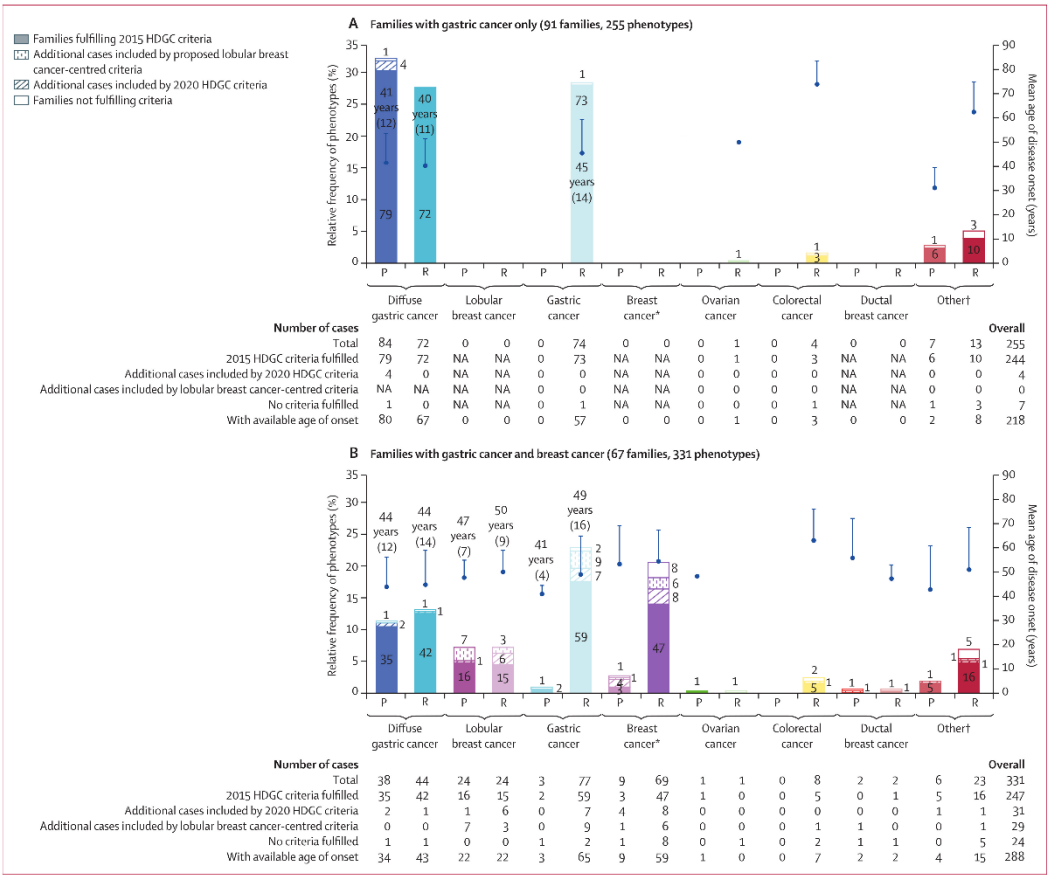


Figure 3: ***CDH1*** **PV/LPV carrier families and lobular breast cancer-related criteria**
(A) *CDH1* single-nucleotide and copy-number variants occurring in PV/LPV carrier families not fulfilling the 2020 HDGC criteria. PV/LPVs found in the cohort are shown by *CDH1* exon location. The size of each circle is proportional to the number of families found to carry a particular variant. (B) Proposed lobular breast cancer-expanded criteria and proportion of the 176 PV/LPV carrier families explained by these criteria. Lobular breast cancer-centred criteria 3, 4, and 5 are proposed as an addition to the 2020 HDGC criteria. dup=duplication. EX=exon. del=deletion. HDGC=hereditary diffuse gastric cancer. PV/LPV=pathogenic or likely pathogenic variant. *Families resolved by lobular breast cancer-centred criteria. †One of two families is resolved by the lobular breast cancer-centred criteria.

approach to address these questions, using the largest series to date (to our knowledge) of clinically characterised rare *CDH1* variant carriers and their relatives, who were tested irrespective of fulfilling clinical criteria.

Our data support germline truncating pathogenic or likely pathogenic variants as being major risk factors for HDGC-related cancers, and do not provide evidence of such an association for missense variants of unknown significance. These data confirm previous findings,[2,6,9] while challenging the role of most *CDH1* missense variants of unknown significance as being causative for HDGC.[8,27,28] Indeed, the limited number of missense variants, that were initially deposited in VEP and ClinVar as pathogenic or likely pathogenic variants, have later been shown to affect splicing, generating premature truncation.[11,29] We should not, however, discard the possibility that some purely missense variants can cause HDGC. Large databases and increasingly robust clinical and functional studies are needed to show whether some missense variants of unknown significance could cause HDGC.[6]

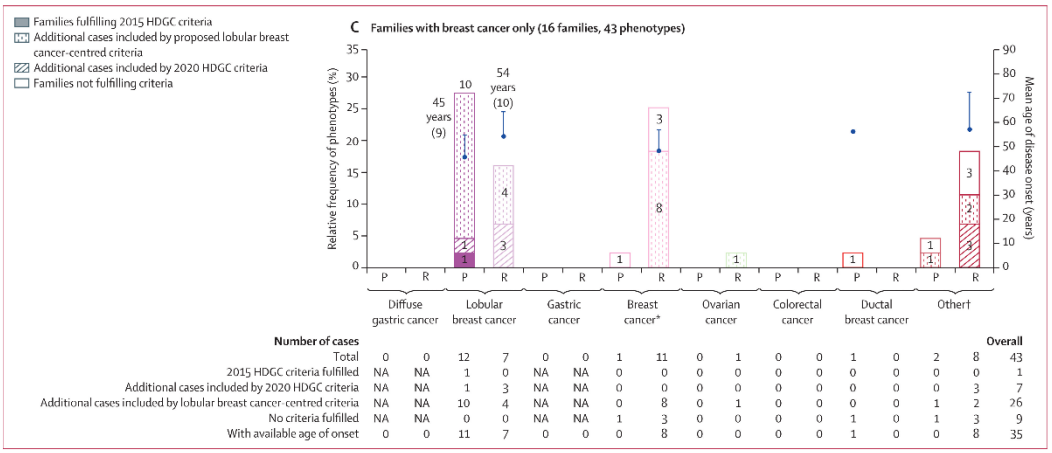
Our results consolidate the spectrum of cancers associated with *CDH1* PV/LPVs, which is enriched for early onset, histologically confirmed, diffuse gastric cancer and lobular breast cancer in probands, and gastric cancer in relatives, and occurs mainly in families fulfilling the 2020 HDGC clinical criteria.[2] Furthermore, to the best of our knowledge, we showed for the first time that lobular breast cancer had the greatest positive association with *CDH1* PV/LPVs, followed by



Number of cases	Diffuse gastric cancer		Lobular breast cancer		Gastric cancer		Breast cancer*		Ovarian cancer		Colorectal cancer		Ductal breast cancer		Other†		Overall
Total	84	72	0	0	0	74	0	0	0	1	0	4	0	0	7	13	255
2015 HDGC criteria fulfilled	79	72	NA	NA	0	73	NA	NA	0	1	0	3	NA	NA	6	10	244
Additional cases included by 2020 HDGC criteria	4	0	NA	NA	0	0	NA	NA	0	0	0	0	NA	NA	0	0	4
Additional cases included by lobular breast cancer-centred criteria	NA	NA	NA	NA	0	0	NA	NA	0	0	0	0	NA	NA	0	0	0
No criteria fulfilled	1	0	NA	NA	0	1	NA	NA	0	0	0	1	NA	NA	1	3	7
With available age of onset	80	67	0	0	0	57	0	0	0	1	0	3	0	0	2	8	218

Number of cases	Diffuse gastric cancer		Lobular breast cancer		Gastric cancer		Breast cancer*		Ovarian cancer		Colorectal cancer		Ductal breast cancer		Other†		Overall
Total	38	44	24	24	3	77	9	69	1	1	0	8	2	2	6	23	331
2015 HDGC criteria fulfilled	35	42	16	15	2	59	3	47	1	0	0	5	0	1	5	16	247
Additional cases included by 2020 HDGC criteria	2	1	1	6	0	7	4	8	0	0	0	0	0	0	1	1	31
Additional cases included by lobular breast cancer-centred criteria	0	0	7	3	0	9	1	6	0	0	0	1	1	0	0	1	29
No criteria fulfilled	1	1	0	0	1	2	1	8	0	1	0	2	1	1	0	5	24
With available age of onset	34	43	22	22	3	65	9	59	1	0	0	7	2	2	4	15	288

(Figure 4 continues on next page)



Number of cases																	Overall
Total	0	0	12	7	0	0	1	11	0	1	0	0	1	0	2	8	43
2015 HDGC criteria fulfilled	NA	NA	1	0	NA	NA	0	0	0	0	0	0	0	0	0	0	1
Additional cases included by 2020 HDGC criteria	NA	NA	1	3	NA	NA	0	0	0	0	0	0	0	0	0	3	7
Additional cases included by lobular breast cancer-centred criteria	NA	NA	10	4	NA	NA	0	8	0	1	0	0	0	0	1	2	26
No criteria fulfilled	NA	NA	0	0	NA	NA	1	3	0	0	0	0	1	0	1	3	9
With available age of onset	0	0	11	7	0	0	0	8	0	0	0	0	1	0	0	8	35

Figure 4: **Distribution of phenotypes in probands and relatives from families carrying germline *CDH1* PV/LPVs and having gastric cancer only, gastric cancer and breast cancer, and breast cancer only, with reference to HDGC clinical criteria**

Relative frequency of phenotypes, phenotype distribution, number of cases, and average age of onset of probands and relatives in families carrying PV/LPVs and having gastric cancer only, gastric cancer and breast cancer, or breast cancer only (data further detailed in appendix pp 62–63). The left axis shows the relative frequency (%) of each phenotype out of the total number of phenotypes. Mean age of onset (SD) is shown for HDGC-associated phenotypes. Detailed data on number of cases and average age of onset for other phenotypes are provided in appendix (pp 62–63). Probands could contribute more than one phenotype. HDGC=hereditary diffuse gastric cancer. PV/LPV=pathogenic or likely pathogenic variant. P=probands. R=relatives. VUS=variants of unknown significance. *Breast cancer of unknown histotype. †Other phenotypes not including gastric cancer, breast cancer, colorectal cancer, or ovarian cancer (detailed in appendix p 3).

diffuse gastric cancer and gastric cancer. These findings contribute to our understanding of the relationship between *CDH1* PV/LPVs and diffuse gastric cancer and lobular breast cancer in probands, and gastric cancer in relatives.[9,12,10] Other cancer types—including ductal breast cancer, ovarian cancer, and colorectal cancer—were extremely rare in PV/LPV carrier families, showing no evidence of association with the presence of PV/LPVs. This finding supports the exclusion of these cancer types from the spectrum of cancers associated with *CDH1* PV/LPVs.[11] Altogether, these genotype–phenotype data support clinical recommendations for intensive surveillance and prophylactic surgery of stomach and breast only in carriers of *CDH1* pathogenic or likely pathogenic variants.[7] The apparent association between the occurrence of ovarian cancer in families carrying missense variants of unknown significance merits further research, which is beyond the scope of the current study.

This European cohort had a predominance of families with gastric cancer only and gastric cancer and breast cancer over families with breast cancer only, which represented 9% of all *CDH1* PV/LPV carrier families. By contrast, similar cohorts in the USA reported breast cancer only (presumably HLBC) in more than 35% of families carrying *CDH1* pathogenic or likely pathogenic variants.[9,12] Our observation that a large fraction (38%) of PV/LPV carrier families have both gastric cancer and breast cancer are further supported by additional data from the USA, which showed occult diffuse gastric cancer in risk-reduction gastrectomy specimens from 93·8% of individuals with *CDH1* pathogenic or likely pathogenic variants in presumed HLBC families.[17]

27% of *CDH1* PV/LPV carrier families, herein described, had at least one histologically confirmed case of lobular breast cancer, with or without gastric cancer (sometimes unconfirmed diffuse gastric cancer), highlighting the need (as reinforced by others)[32] to widen the 2020 HDGC criteria.[7] To maximise the diagnosis of *CDH1* PV/LPV carriers, we propose three novel lobular breast cancer-centred criteria to drive *CDH1* testing in patients with histologically confirmed lobular breast cancer with or without a history of gastric cancer, representing most clinical contexts among carriers of *CDH1* PV/LPVs who do not meet the 2020 HDGC criteria.[7] This widening needs to assure that the criteria remain sensitive enough to identify as many pathogenic or likely pathogenic variant carriers as possible, with genetic testing being done preferentially in individuals who are most likely to carry such variants. Lerner and colleagues[16] followed a similar rationale and proposed testing for *CDH1* variants in everyone fulfilling the modified 2020 HDGC criteria and the Hereditary Breast and Ovarian Cancer criteria (Yale criteria), reporting an increased sensitivity and simplified system. Although a high sensitivity was attained, it was accompanied by a substantial decrease in specificity and Youden index (appendix p 70). Using the Yale criteria to select patients for *CDH1* testing will lead to

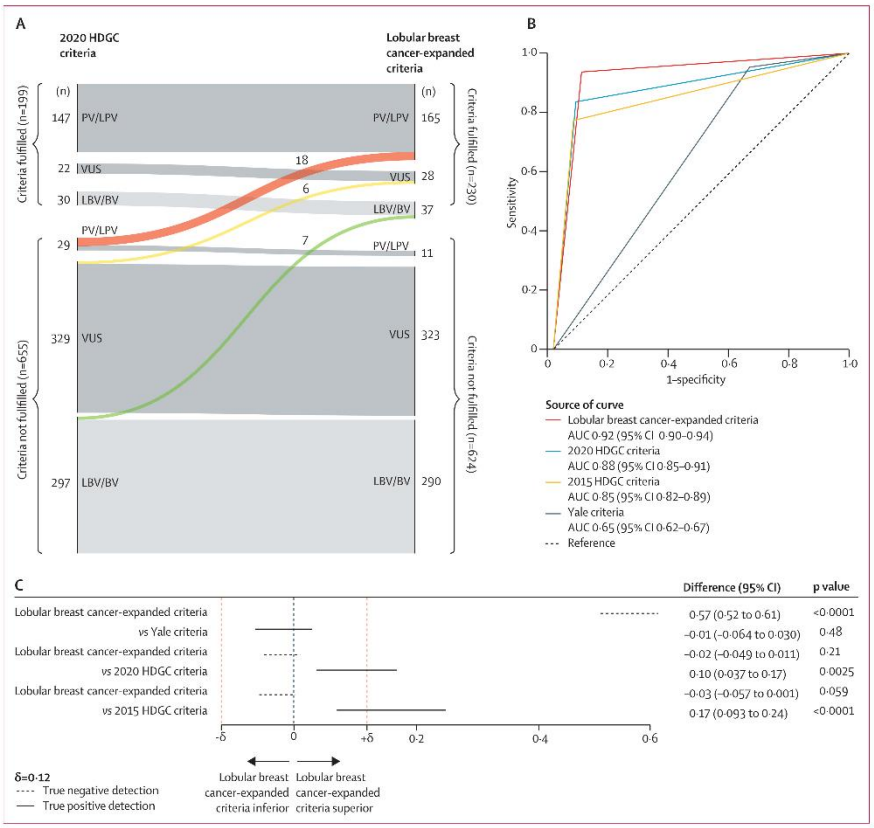


Figure 5: ***CDH1*** **variant classification and distribution according to compliance with 2020 HDGC clinical criteria, and performance evaluation of HDGC clinical criteria sets**

(A) Alluvial plot (created with the visualisation platform developed by Mauri and colleagues)[45] displaying the number of families carrying *CDH1* variants and fulfilling 2020 HDGC criteria or lobular breast cancer-expanded criteria. Variants are categorised according to the American College of Medical Genetics and Genomics and Association for Molecular Pathology (version 2). (B) Receiver operating characteristic analysis comparing the AUCs for four different clinical criteria sets (lobular breast cancer-expanded *vs* 2020 HDGC criteria Z=3·54, AUC difference 0·042 [95% CI 0·02–0·06], p=0·0004). (C) Forest plot representing the lobular breast cancer-expanded clinical criteria equivalence test. The equivalence limit (δ) was set to 0·12. HDGC=hereditary diffuse gastric cancer. LBV/BV=likely benign or benign variant. PV/LPV=pathogenic or likely pathogenic variant. VUS=variants of unknown significance.

an excess of patients with breast cancer being tested with a very low pickup rate, representing a waste of resources. Contrarily, the lobular breast cancer-expanded criteria proposed in this Article showed superior performance to that of previous criteria, and are expected to identify a larger number of carriers of *CDH1* pathogenic or likely pathogenic variants with high specificity and sensitivity.

Despite the established *CDH1*-related causality in lobular breast cancer, HLBC as an entity remains controversial since its definition in 2020.[2,8,12,17] Our genotype–phenotype analysis does not provide evidence that specific *CDH1* PV/LPVs predispose individuals differentially to HDGC (families with gastric cancer only or gastric cancer and breast cancer) or to HLBC (families with lobular breast cancer only), as suggested previously, particularly in the HLBC context.[8,17,27,28] These data are supported by Gamble and colleagues, who studied 31 families tested because of suspicion of HLBC, and carrying 19 different *CDH1* pathogenic or likely pathogenic variants.[17] These authors conclude that a proportion of families with HLBC who initially did not report history of gastric cancer exhibited high rates of

occult signet ring cell gastric cancer in a prospective study. Additionally, our literature review showed that most *CDH1* PV/LPVs reported in families with breast cancer only have been reported in families with HDGC elsewhere (appendix pp 73–77).[17] Alternatively, HLBC and HDGC might differentially occur in pathogenic or likely pathogenic variant carrier families, if they originate from particular geographical regions and due to exposure to particular environmental factors or the presence of modifier genetic events,[33–35] which might favour clinical expression of diffuse gastric cancer only or lobular breast cancer only, or both (appendix pp 73–77). Additionally, these modifier genes or environmental factors might prevent the progression of intramucosal gastric signet ring cell carcinoma,[35] leading to familial presentations of lobular breast cancer only, explaining HLBC in some instances. In such cases, *CDH1*-associated HLBC might exist as a HDGC-independent clinical entity, according to the 2020 guidelines paper.[2] In any case, as available data indicate that *CDH1* pathogenic or likely pathogenic variants predispose individuals to both diffuse gastric cancer and lobular breast cancer, clinical management should address both organs.

An additional factor that might constitute a confounder is variant classification. Six *CDH1* missense variants earlier described as related to HLBC[8,17,27,28] were revisited and reclassified in this study as LBV/BVs or variants of unknown significance, in accordance with American College of Medical Genetics and Genomics and Association for Molecular Pathology *CDH1* guidelines (appendix pp 73–77). This example highlights the need of accurate variant classification for correct evaluation of disease-causal factors, and adequate disease management. Clinical classification is challenging, mainly for newly discovered variants with missing clinical data.[6,32] This is one of the reasons that led us to share as many clinical details of the families analysed as possible (appendix pp 5–46).

A limitation of this study is the fact that Sweden, the Netherlands, and the UK contributed 83 (10%) of 854 families tested specifically for *CDH1* due to clinical suspicion of HDGC, as opposed to all other countries that contributed data irrespective of HDGC clinical suspicion, as requested in the frame of the study design. Additionally, as a multicentre study, we recognise a possible information bias inherent in the use of retrospective data from multiple institutions, with restricted and non-harmonised phenotypic descriptions. Furthermore, assumption of *CDH1*-positive carrier status in clinically affected relatives, and lack of confirmed histology in many cases of gastric cancer and breast cancer among relatives, are limitations. We also recognise the possibility of overfitting in the present study, and, to our knowledge, a cohort with the same level of detail is currently not available for cross-validation purposes. This is the most detailed description of clinical data from rare *CDH1* variant carriers published to date, to the best of our knowledge, which we believe minimises the study's limitations.

In summary, this is the first multicentre genotype-first study to identify associations between rare *CDH1* germline variants and cancer phenotypes, using the largest dataset of rare *CDH1* variant carriers and their relatives, to the best of our knowledge, completed under the umbrella of ERN GENTURIS. The study highlights associations between a specific category of rare *CDH1* germline variants and HDGC specific phenotypes, and proposes novel clinical criteria to trigger *CDH1* germline testing. Altogether, these findings will contribute to clarify the context for clinical management in families carrying *CDH1* germline variants.

Contributors

JG-P and CO conceptualised the study and did the formal analysis. JG-P, RB-M, SL, AD, and CO contributed to investigation and methodology. RB-M, JG-P, SL, AD, and CO contributed to figure design. JG-P and SL contributed to classification of variants using American College of Medical Genetics and Genomics and Association for Molecular Pathology *CDH1* guidelines. AD, NP, and CLe contributed to the statistical analysis. LS contributed to data management. RM contributed to the Variant Effect Predictor analysis. JJ-M contributed to the *CDH1* gene models. LGa, SCas, SS, HP, SF, FC, CP, MRT, SA, SB-I, JBa, AB, PRB, MB, VB, HB, JBr, DC, GC, SCar, CC, KD, RdP, CD, ED-G, CE, DGE, DF, EF, RCF, CF, MG-B, MG, LGo, KH, HH, EH-F, RH, MK, KL-R, CLa, MJLL, CM-B, SM, GM, SN, AP-G, GNR, ES, ISi, CS, JLS, ISp, VS-I, GT, M-IT, ERW, MT, NH, and CO contributed to data collection. JG-P and CO wrote the original draft of the manuscript. All authors contributed to reviewing and editing the manuscript. CO contributed to supervision, project administration, and funding acquisition. JG-P, RB-M, SL, AD, LGa, and CO directly accessed and verified the underlying raw data in all research articles needed for this submission. All authors had access to the data in the study, critically revised the manuscript for important intellectual content, and have agreed to the submitted version of the results in accordance with the underlying data.

Declaration of interests

DGE declares fees from Astrazeneca and Recursion. ERW declares grants from International Alliance for Cancer Early Detection, for which they are codirector of the research domain. GNR declares receipt of funding for study materials, medical writing, and article processing charges from Italian Ministry of Education (GNR). MJLL declares consulting fees (via the Radboud University Medical Center) from Merck Sharp & Dohme (MSD), AstraZeneca, Lilly, Janssen-Cilag, Illumina, GlaxoSmithKline. PRB declares fees from AstraZeneca, MSD, and Bristol Myers Squibb; and is a scientific committee member for the Geneticancer patients association. JBa declares fees from AstraZeneca, Lilly, and Pfizer. SA is a member of APC subVCEP of the InSiGHT/ClinGen Hereditary Colorectal Cancer/Polyposis Variant Curation Expert Panel; is an unpaid member of the German Gene Diagnostics Commission and speaker of the Centre for Hereditary Tumour Syndromes of the University of Bonn. RH declares grants from SLA Pharma and Janssen Pharmaceuticals; consulting fees from Janssen and One Two Therapeutics; equipment from Fujifilm; is the head of German Consortium for Familial Gastrointestinal Cancer; and is an unpaid advisory board member of the Lynch Syndrome advocacy Group and the Familial Polyposis Group. ES declares grants from NCT/DKTK Master. ES declares honoraria for presentations from AstraZeneca, Georg Thieme Verlag KG, and payment for expert testimony from Illumina; is a member of the board of directors of Deutsche Gesellschaft für HumanGenetik; an advisor for Dresden-concept Genome Center; and is board of directors president (paid) for LNS laboratoire National de Santé. RdP declares support for presentations (via his institution) from MSD and AstraZeneca. GC declares to receive funding for study materials, medical writing, article processing charges from the Spanish Ministry of Science and Innovation, the Instituto de Salud Carlos III CIBERONC, and the Government of Catalonia.; consulting fees from VCN Biosciences Synthetic Biologics; is the chair of the Council of the International

Society of Hereditary Gastrointestinal Tumours and the FUREGA Fundació Recerca en Gastroenterologia; and stock in Synthetic Biologics. CLa declares consulting fees and honoraria from AstraZeneca and MSD, and is a paid advisory board member for Illumina.

Data sharing

Data, aggregated and pseudo-anonymised at the source, on *CDH1* germline variant molecular and clinical classification, disease spectrum, family history, and age of disease onset, will be made available through a bulk submission to the ClinVar database and in the appendix (pp 5–46) of this Article. All participating institutions have confirmed their agreement in making the information available and public.

Acknowledgments

This research was supported by ERN GENTURIS (project ID 739547); ERN GENTURIS is partly cofunded by the EU within the framework of the Third Health Programme ERN-2016—Framework Partnership Agreement 2017–2021. This work has received funding from the European Union's Horizon 2020 research and innovation programme (grant number 779257, Solve-RD: H2020-SC1–2017-Single-Stage-RTD) via a PhD studentship awarded to JG-P. The work has also been supported by the European Regional Development Fund through the COMPETE 2020-Operacional Programme for Competitiveness and Internationalisation, Portugal 2020, Programa Operacional Regional do Norte under the project "The Porto Comprehensive Cancer Center" with the reference NORTE-01-0145-FEDER-072678—Consórcio PORTO.CCC—Porto.Comprehensive Cancer Center; the Portuguese national funds through Fundação para a Ciência e Tecnologia, Ministério da Ciência, Tecnologia e Inovação (project references PTDC/BTM-TEC/30164/2017 [3DChrome] and PTDC/BTM-TEC/6706/2020 [LEGOH]); the University of Porto through the Doctoral Programme in Biomedicine (Faculty of Medicine; awarded to JG-P), Doctoral Programme in Molecular and Cellular Biotechnology Applied to Health Sciences (ICBAS; awarded to RB-M and SL), Doctoral Programme In Molecular And Cell Biology (ICBAS; awarded to AD); the Manchester National Institute for Health and Care Research (NIHR) Biomedical Research Centre (IS-BRC-1215-20007), providing support to DGE and ERW; the NIHR Cambridge Biomedical Research Centre (BRC-1215-20014), providing support to MT; Medical Research Council (RG84369) and Cancer Research UK (C14478/A12088; RG80420/A21047/A29225), providing support to RF; the Carlos III National Health Institute and Ministerio de Ciencia e Innovación, funded by the European Regional Development Fund (PI19/00553; PID2019-111254RB-I00) and CIBERONC (CB16/12/00234), providing support to CLa, GC, and JBr; the Government of Catalonia (Pla estratègic de recerca i innovació en salut [PERIS_MedPerCan and URDCat projects], 2017SGR1282 and 2017SGR496); and the Fundação para a Ciência e Tecnologia (UIDB/04539/2020, UIDP/04539/2020, LA/P/0058/2020, and POCI-01-0145-FEDER-022184) providing support to CE. We are grateful to the ERN GENTURIS National Coordinators for promoting data collection at the country level for the current project; Rui Magalhães (ICBAS, University of Porto, and Centro Hospitalar Universitário do Porto) for consultancy in statistics; and Víctor Treviño (Escuela de Medicina y Ciencias de la Salud del Tecnológico de Monterrey, Monterrey, Mexico) for the R script for gene models.

References

1 van der Post RS, Vogelaar IP, Carneiro F, et al. Hereditary diffuse gastric cancer: updated clinical guidelines with an emphasis on germline *CDH1* mutation carriers. *J Med Genet* 2015; **52:** 361–74.

2 Blair VR, McLeod M, Carneiro F, et al. Hereditary diffuse gastric cancer: updated clinical practice guidelines. *Lancet Oncol* 2020; **21:** e386–97.

3 Caldas C, Carneiro F, Lynch HT, et al. Familial gastric cancer: overview and guidelines for management. *J Med Genet* 1999; **36:** 873–80.

4 Guilford P, Hopkins J, Harraway J, et al. E-cadherin germline mutations in familial gastric cancer. *Nature* 1998; **392:** 402–05.

5 Oliveira C, Senz J, Kaurah P, et al. Germline *CDH1* deletions in hereditary diffuse gastric cancer families. *Hum Mol Genet* 2009; **18:** 1545–55.

6 Lee K, Krempely K, Roberts ME, et al. Specifications of the ACMG/AMP variant curation guidelines for the analysis of germline *CDH1* sequence variants. *Hum Mutat* 2018; **39:** 1553–68.

7 Fitzgerald RC, Hardwick R, Huntsman D, et al. Hereditary diffuse gastric cancer: updated consensus guidelines for clinical management and directions for future research. *J Med Genet* 2010; **47:** 436–44.

8 Corso G, Figueiredo J, La Vecchia C, et al. Hereditary lobular breast cancer with an emphasis on E-cadherin genetic defect. *J Med Genet* 2018; **55:** 431–41.

9 Roberts ME, Ranola JMO, Marshall ML, et al. Comparison of *CDH1* penetrance estimates in clinically ascertained families *vs* families ascertained for multiple gastric cancers. *JAMA Oncol* 2019; **5:** 1325–31.

10 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:** 1348–53.

11 Kaurah P, MacMillan A, Boyd N, et al. Founder and recurrent *CDH1* mutations in families with hereditary diffuse gastric cancer. *JAMA* 2007; **297:** 2360–72.

12 Xicola RM, Li S, Rodriguez N, et al. Clinical features and cancer risk in families with pathogenic *CDH1* variants irrespective of clinical criteria. *J Med Genet* 2019; **56:** 838–43.

13 Hansford S, Kaurah P, Li-Chang H, et al. Hereditary diffuse gastric cancer syndrome: *CDH1* mutations and beyond. *JAMA Oncol* 2015; **1:** 23–32.

14 Lowstuter K, Espenschied CR, Sturgeon D, et al. Unexpected *CDH1* mutations identified on multigene panels pose clinical management challenges. *JCO Precis Oncol* 2017; **1:** 1–12.

15 Adib E, El Zarif T, Nassar AH, et al. *CDH1* germline variants are enriched in patients with colorectal cancer, gastric cancer, and breast cancer. *Br J Cancer* 2022; **126:** 797–803.

16 Lerner BA, Xicola RM, Rodriguez NJ, Karam R, Llor X. Simplified and more sensitive criteria for identifying individuals with pathogenic *CDH1* variants. *J Med Genet* 2022; published online Jan 25. https://doi.org/10.1136/jmedgenet-2021-108169.

17 Gamble LA, Rossi A, Fasaye GA, et al. Association between hereditary lobular breast cancer due to *CDH1* variants and gastric cancer risk. *JAMA Surg* 2022; **157:** 18–22.

18 Oliveira C, Bordin MC, Grehan N, et al. Screening E-cadherin in gastric cancer families reveals germline mutations only in hereditary diffuse gastric cancer kindred. *Hum Mutat* 2002; **19:** 510–17.

19 den Dunnen JT, Dalgleish R, Maglott DR, et al. HGVS recommendations for the description of sequence variants: 2016 update. *Hum Mutat* 2016; **37:** 564–69.

20 McLaren W, Gil L, Hunt SE, et al. The ensembl variant effect predictor. *Genome Biol* 2016; **17:** 122.

21 Hebsgaard SM, Korning PG, Tolstrup N, Engelbrecht J, Rouzé P, Brunak S. Splice site prediction in *Arabidopsis thaliana* pre-mRNA by combining local and global sequence information. *Nucleic Acids Res* 1996; **24:** 3439–52.

23 Stevenson M, Sergeant E. EpiR: tools for the analysis of epidemiological data. https://cran.r-project.org/web/packages/epiR/index.html (accessed Aug 25, 2022).

24 Perez-Jaume S, Skaltsa K, Pallarès N, Carrasco JL. ThresholdROC: optimum threshold estimation tools for continuous diagnostic tests in R. *J Stat Softw* 2017; **82:** 1–21.

25 Walker E, Nowacki AS. Understanding equivalence and noninferiority testing. *J Gen Intern Med* 2011; **26:** 192–96.

26 Mauri M, Elli T, Caviglia G, Uboldi G, Azzi M. RAWGraphs: a visualisation platform to create open outputs. 12th Biannual Conference on Italian SIGCHI Chapter; Sept 18–20, 2017 (abstr 28).

27 Girardi A, Magnoni F, Vicini E, et al. *CDH1* germline mutations in families with hereditary lobular breast cancer. *Fam Cancer* 2022; **31:** 274–78.

28 Corso G, Intra M, Trentin C, Veronesi P, Galimberti V. *CDH1* germline mutations and hereditary lobular breast cancer. *Fam Cancer* 2016; **15:** 215–19.

29 Barbosa-Matos R, Leal Silva R, Garrido L, et al. The *CDH1* c.1901C>T variant: a founder variant in the Portuguese population with severe impact in mRNA splicing. *Cancers (Basel)* 2021; **13:** 4464.

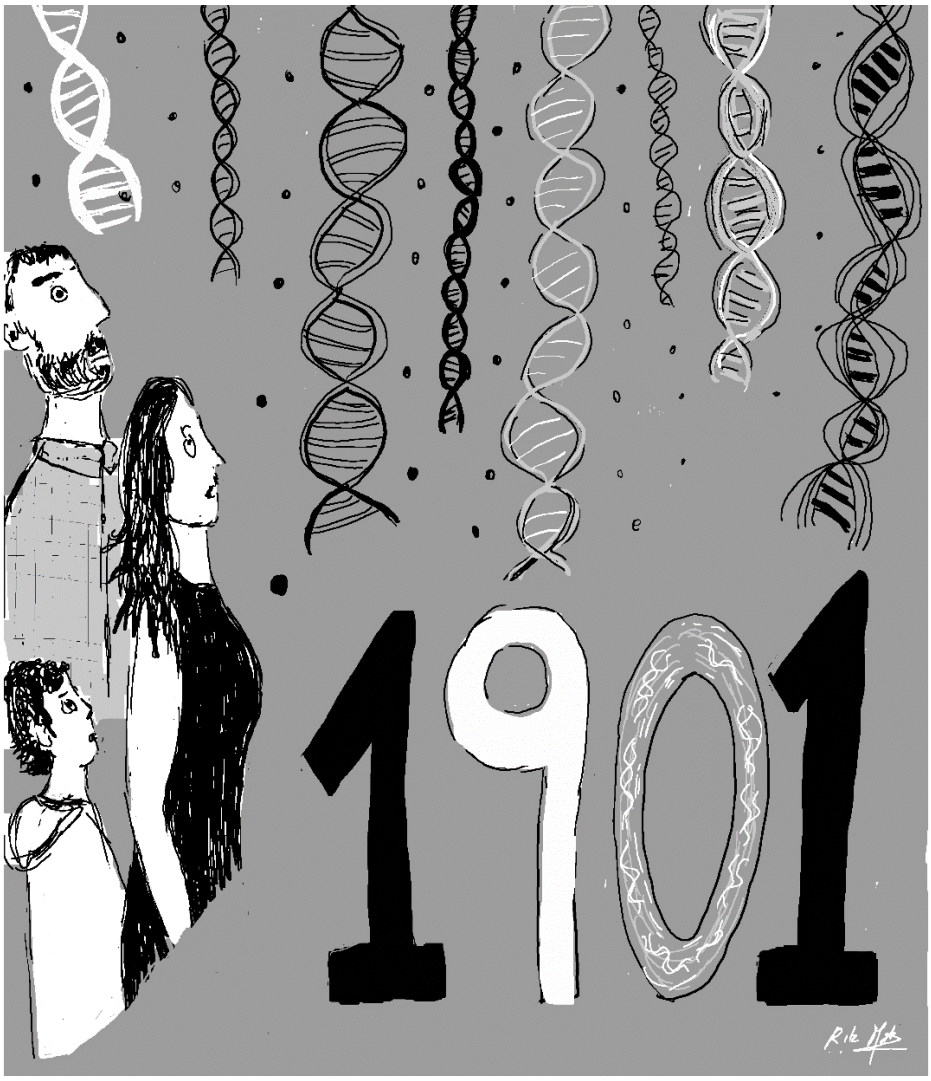
For **ClinVar** see https://www.ncbi.nlm.nih.gov/clinvar/

30 Lo W, Zhu B, Sabesan A, et al. Associations of *CDH1* germline variant location and cancer phenotype in families with hereditary diffuse gastric cancer (HDGC). *J Med Genet* 2019; **56:** 370–79.
31 Massari G, Magnoni F, Favia G, et al. Frequency of *CDH1* germline mutations in non-gastric cancers. *Cancers (Basel)* 2021; **13:** 2321.
32 Katona BW, Clark DF, Domchek SM. *CDH1* on multigene panel testing: look before you leap. *J Natl Cancer Inst* 2020; **112:** 330–34.
33 Zhang H, Ahearn TU, Lecarpentier J, et al. Genome-wide association study identifies 32 novel breast cancer susceptibility loci from overall and subtype-specific analyses. *Nat Genet* 2020; **52:** 572–81.
34 Tanikawa C, Kamatani Y, Toyoshima O, et al. Genome-wide association study identifies gastric cancer susceptibility loci at 12q24.11-12 and 20q11.21. *Cancer Sci* 2018; **109:** 4015–24.
35 Ellison-Loschmann L, Sporle A, Corbin M, et al. Risk of stomach cancer in Aotearoa/New Zealand: a Māori population based case-control study. *PLoS One* 2017; **12:** e0181581.

Supplementary Material: Available through the link:

https://www.thelancet.com/journals/lanonc/article/PIIS1470-2045(22)00643-X/fulltext#supplementaryMaterial

3. Lifetime risk in *CDH1* carriers and diagnostic signatures: insights from the Portuguese founder Population



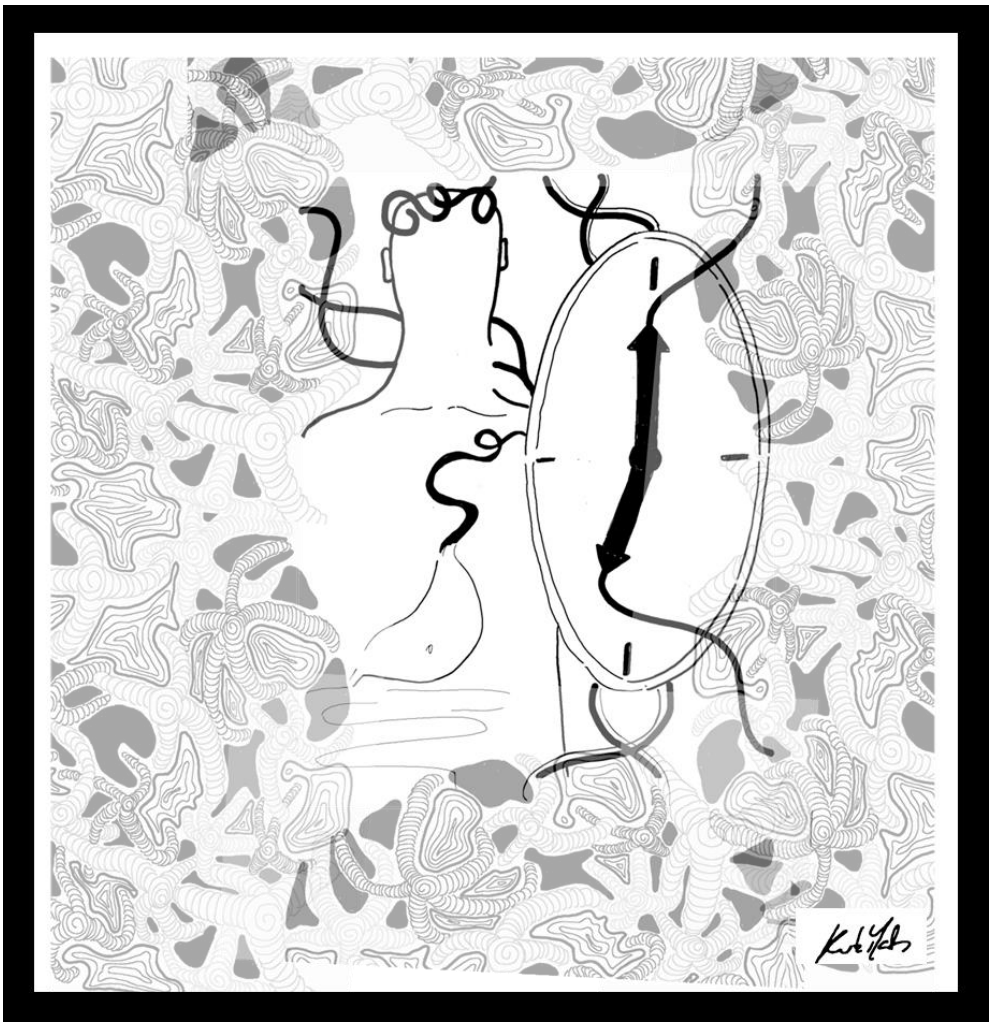
Paper 5: Lifetime cancer risk associated with the most recurrent *CDH1* germline variant in DGLBCS: A retrospective study of a founder effect.

Paper 6: Combined whole-genome and transcriptome profiling of Early Onset Cancer Patients and Asymptomatic Carriers of a *CDH1* founder variant

3.1. Paper 5, Chapter III

Lifetime cancer risk associated with the most recurrent *CDH1* germline variant in DGLBCS: A retrospective study of a founder effect.

Barbosa-Matos R, Fonseca J*, Garrido L*, Seixas S, Oliveira R, Freixo J, Quental R, Louro P, Fernandes S, Grangeia A, Baptista M, Carneiro F, Pinheiro H, Castedo S, Gullo I, Drouet Y, Oliveira C.



Author statement:

Rita Barbosa-Matos contributed to the conceptualization, investigation, data collection, methodology, figure design and writing (original draft)

Abstract

Families impacted by Diffuse Gastric Cancer and Lobular Breast Cancer Syndrome (DGLBCS) exhibit distinct disease penetrance and expressivity. Current penetrance estimates for *CDH1* carriers, reflecting the "one-size-fits-all" approach to surveillance and risk-reducing options for carriers, underscore the need for more tailored risk estimation.

Objective: To estimate *CDH1-related* cancer risk in a patient cohort linked by a founder variant, living in the same geographical area for centuries and homogeneously tested and monitored at a single institution.

Design and Participants: In this retrospective study, we compiled comprehensive pedigree and clinical data from 285 individuals across nine DGLBCS families residing in a region distinguished by a well-established *CDH1* founder effect. The *CDH1* c.1901 C>T pathogenic variant, the most recurrent *CDH1* variant in Europe, was identified in 95 individuals from 171 genotyped individuals.

Main Outcome and Measures: The cumulative risk of Diffuse Gastric Cancer (DGC) and Lobular Breast Cancer (LBC) were assessed relative to the incidence data of the European Cancer Information System (ECIS) exclusively to the founder effect region and estimated using the genotype restricted likelihood (GRL) method. In addition, an exhaustive description of the clinical data and patient management profile was performed.

Results: 285 individuals were included, 33.3% (95/285) were *CDH1* carriers, 9.1% (16/285) were diagnosed with GC, 147 were females and 8.1% (12/147) females presented BC. The cumulative risk of DGC was estimated at 5.8% at 60 years for both genders, and no other DGCs were identified beyond this age. For LBC, the cumulative risk estimate was 19.1% at 60 years. Risk ratios compared to the general population were higher in early-age DGC- and LBC-groups and decreased with age.

Conclusion and Relevance: This is the largest familial study of DGLBCS sharing the most recurrent *CDH1* variant in Europe, revealing the lowest cumulative cancer risk for *CDH1* mutation carrier published to date. Variant-specific risk estimates, based on detailed family data, may assist patient decision-making regarding surveillance and prophylactic measures.

INTRODUCTION

Diffuse Gastric Cancer and Lobular Breast Cancer Syndrome (DGLBCS) is a rare autosomal-dominant syndrome associated with pathogenic (P) or likely-pathogenic (LP) variant in CDH1 and CTNNA1 genes. It leads to early onset Diffuse Gastric Cancer (DGC) that accounts for nearly 30% of all gastric cancers (GC)(Ansari et al., 2018) and Lobular Breast Cancer (LBC) which constitute about 10% of breast cancer (BC) cases(Christgen et al., 2016; Dossus & Benusiglio, 2015; Phipps et al., 2010). Both tumours exhibit a loss of intercellular adhesion, characterized by a diffuse invasive growth pattern. According to the 2020 guidelines from the International Gastric Cancer Linkage Consortium (IGCLC), clinical diagnosis considers family history of gastric and breast cancer at specific ages of disease onset, particularly the occurrence of DGC with or without LBC. Furthermore, a positive molecular diagnosis through genetic testing is crucial to meet the criteria for DGLBCS.

Individuals with P/LP CDH1 variants exhibit variable cancer risks across several studies. The cumulative lifetime risk for GC ranges from 37.2%–70.0% in males, and 24.7%–83.0% in females. Prophylactic total gastrectomy (PTG) is recommended in individuals between 20–70 years of age fulfilling clinical criteria independently of their age at the time of genetic testing. However, if the family history of GC is weak, the emphasis on PTG is reduced. Additionally, females also have an increased risk of developing BC between 39–55%, and annual surveillance or bilateral risk-reducing mastectomy are recommended (Hansford et al., 2015; Kaurah et al., 2007; Pharoah et al., 2001; Roberts et al., 2019; Xicola et al., 2019).

Optimal timing for risk reducing surgeries is unknown and long-term complications might arise, especially in the context of PTG associated with a decreased quality-of-life immediately after clinical intervention (Gallanis et al., 2024; Muir et al., 2016). For CDH1 carriers who do not immediately pursue a PTG, regular endoscopies should be performed and upon a positive biopsy, total gastrectomy is endorsed.

DGLBCS families are dispersed across the globe and the hundreds of clinically actionable CDH1 variants identified so far display incomplete penetrance, distinct ages of onset and expressivity. Hence, it is recognized that disease risk varies between families, and a comprehensive family history, spanning several generations, is essential for accurate lifetime risk estimations.

Recent studies indicating lower cancer risks often involve CDH1 carriers not meeting DGLBCS criteria and exhibit different cancer histology apart from DGC or LBC. Notably, current risk estimates derive from general incidence rates of GC and BC, without consideration for the specific histological subtypes associated with this syndrome. Few penetrance studies in DGLBCS family data using bias-correction statistical method have been published (Hansford et al., 2015; Kaurah et al., 2007; Pharoah et al., 2001; Roberts et al., 2019; Xicola et al., 2019), whereas it is recognized that estimation from retrospective family data requires specific bias correction to avoid overestimation (Carayol & Bonaïti-Pellié, 2004).

To date, a generalized risk is considered for decision-making surrounding the timing of surgery of an asymptomatic CDH1 carrier. This underscores the need for refined risk assessments that account for the unique histological characteristics and hereditary background defining DGLBCS families.

Our aim was to estimate unbiased lifetime risk for CDH1 c.1901 C>T germline variant carriers, a well-established founder effect in the Northern region of Portugal (Barbosa-Matos et al., 2021). For that, we used genetic, clinical and follow up extended data from nine families comprising 285 individuals living in the same geographical area and homogeneously tested and monitored at a single institution. Additionally, we report the results on surveillance and disease management of these families over a 20-year period.

MATERIALS AND METHODS

Study population

This retrospective cohort study comprised 195 individuals tested for *CDH1* c.1901 C>T mutation from nine families (initially from 11 probands) recruited to the Centro Hospital de São João at Porto (CHUSJ), Portugal between 2000 and 2023. Ethical and legal aspects of the study were approved by the CHUSJ ethics committee (ethics number CES 469/22) and a signed informed consent was obtained by all living participants or under clinical investigation law (21/2014) fulfilling DGLBCS genetic testing criteria.

All individuals underwent a complete evaluation of at least three-generation family history and pertinent clinical findings. Anonymous pedigrees were created for all families (n=400 probands and relatives), and information regarding sex, year

of birth, *CDH1* mutation status when available, age at last follow-up, or death and cancer history (tumour site, age at diagnosis) was available for 285 individuals (**Figure 1A**). This data was obtained in genetic consultations based on medical history obtained in interviews of the probands and their relatives, supplemented, when possible, by medical or pathology reports. Secondly, individuals carrying the *CDH1* c.1901 C>T variant, opting for CHUSJ follow-up, received comprehensive care from a specialized team in gastroenterology, gastric surgery, pathology, cancer genetics, psychosocial support, and nutrition. For these individuals (n=64 carriers) clinical data was also collected on gastric and breast screening, surveillance, and whether any gastrectomy or mastectomy had been performed. Information was collected under the guidance of clinicians ensuring full adherence to ethical and data protection standards.

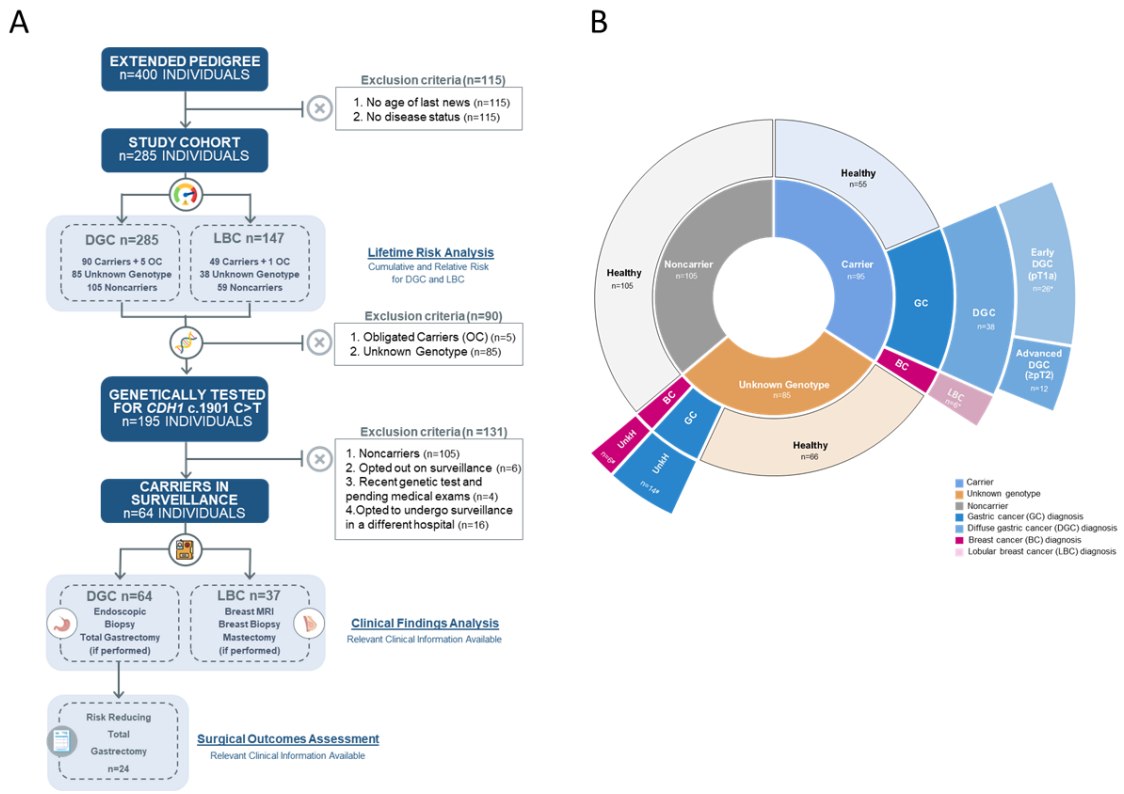


Figure 1. Detailed description of study cohort and phenotypic distribution of the *CDH1* c.1901 C>T families. (A) Diagram depiction of cases selection and exclusion criteria applied in each study analysis. (B) Sunburst diagram depicting carrier status and disease onset in the 285 individuals' cohort. Carrier status was classified as carrier, noncarrier, or unknown genotype. Tumour histology of patients with gastric cancer (GC) and breast cancer (BC) was also attributed when

proven. UnkH – Unknown histological classification of tumour (GC and BC). *Four female variant carriers had both LBC diagnosis and early DGC findings in surveillance. #One woman with unknown genotype presented both GC and BC.

Statistical analysis

Due to family ascertainment bias (commonly detecting positive carriers in families with strong cancer history), a correction is necessary when estimating cancer risks from retrospective family data. The genotype-restricted likelihood (GRL) method was previously applied in penetrance studies of other hereditary syndromes involving CTNNA1, MSH2, MLH1, and MSH6 genes (Bonadona et al., 2011; Coudert et al., 2022). It uses all available information, accounting individuals with unknown genotype status, and conditioning on the phenotypes of all individuals and the genotype of the proband (index case) since this information contributes to the ascertainment of the family (Carayol & Bonaïti-Pellié, 2004).

A non-parametric function with three fixed age-nodes, and monotone cubic spline interpolation (NP GRL) was used to model the shape of the penetrance function with the GRL for DGC and LBC. For each phenotype, age nodes were delineated by the age distribution of the cohort, encompassing carriers and individuals with unknown genotypes. Each node corresponded to the first quartile (Q1), median, and third quartile (Q3), leading to age nodes at 32, 43 and 59 years for DGC and 36, 45 and 60 years for LBC. Due to the limited number of cases/families, it was not possible to compute confidence intervals (CIs) using the bootstrap approach, thus a 'leave-one-out' strategy was employed. This involved replicating all analyses while systematically excluding one family at a time providing insights into the potential impact of individual families on the overall findings.

For cumulative and relative risk calculations, incidence data for GC and BC was available from the Northern Portuguese population between 2000 and 2010 resorting to data from European Cancer Information System (ECIS) provided by the European Commission (https://ecis.jrc.ec.europa.eu/index.php). We considered as cumulative incidence 30% of gastric cancer incidence for DGC, and 10% of breast cancer incidence for LBC. (Supplementary Table 1). These parameters were fixed to represent age-specific cancer incidence for the noncarriers. Additionally, CDH1 c.1901 C>T variant was absent from population databases and its frequency in the Northern region of Portugal was assumed to be 1/2000 (rare variant).

All descriptive statistics are presented as medians (interquartile ranges) or frequencies with corresponding percentages. Individuals with incomplete information were excluded using specific criteria from relevant analysis (Figure 1A). Continuous variables underwent a normality assessment using the Shapiro-Wilk test, and non-parametric distributions were examined through the independent Mann-Whitney U test. For discrete variables. For comparing discrete variables, $\chi 2$ or Fisher's exact tests were used when adequate.

Kaplan-Meier (KM) curves were calculated to depict the incidence of DGC and LBC in our cohort and for comparison with the GRL estimates. A p-value of <0.05 represented statistical significance, 95% confidence intervals were estimated using the infinitesimal jackknife variance estimate considering the non-independence of relatives, using the robust and cluster options of the survfit(.) function of the survival R package.

All data manipulation and statistical analyses were performed in R (version 4.1). Estimation with the GRL method was performed using the generisk package available on github (https://github.com/youenndrouet/generisk).

Outcomes

Our primary goal was to estimate the lifetime risk for DGC and LBC development in families with DGCLBC sharing a founder effect. Given the context of following these families since 2000, we had the necessary information to perform a prospective study regarding the follow-up of *CDH1* c.1901 C>T positive carriers for both target organs (stomach and breast). Therefore, as another primary outcome, we assessed whether number of fragments analyzed in endoscopic biopsy, time from genetic testing to organ removal, number of tissue blocks analyzed from gastrectomy and other clinical findings registered in endoscopic biopsies or total gastrectomy pathological reports reflect any association with the identification of Early or Advanced DGC lesions within our cohort. Secondary outcomes were short-and long-term consequences of stomach risk-reducing surgery.

RESULTS

Lifetime risk Analysis

From 2000 to 2023, 195 individuals underwent genetic testing at CHUSJ to detect the *CDH1* c.1901 C>T germline variant. Originally, eleven families were presumed to be identified, yet three of them (F2, F8, and F9) were merged due to a non-consanguineous marriage and a shared progenitor (**Supplemental Figure S1A**). As a result, we had access to nine families sharing the same *CDH1* germline alteration and created an extended pedigree of 400 individuals from the same region. One family (F10) recently identified, did not meet the criteria for DGLBCS. The proband, diagnosed with oligodendroglioma at 38 years old, requested genetic testing owing to the familial gastric cancer history and residence in the region affected by the founder effect. The characteristics of the families are summarized in **Supplemental Table 2**.

In total, 285 individuals (148 females and 137 males) were included for DGC penetrance estimates (**Figure 1A**). Our cohort included 95 carriers (including obligated carriers), 105 noncarriers and 85 individuals with unknown genotype. Females constituted 53.7% (51/95) of carriers, 44.7% (38/85) of those with an unknown genotype, and 56.2% (59/105) of noncarriers. Based on histology reports, 10.5% (10/95) of carriers were clinically affected by Advanced DGC and two cases (2.1%) of Advanced GC of unknown histology were reported in obligated carriers (**Figure 1B**). Overall, 14 cases (16.5%) classified as Advanced GC of unknown histology were identified in individuals with an unknown genotype. Median age of Advanced GC diagnosis was 35 years of age for carriers (IQR 29–43) and 40 years for unknown genotype (IQR 35–51). No GC cases were identified in the noncarriers group. (**Figure 1B; Figure 2A; Supplemental Table 3**)

From the 147 females considered for LBC penetrance studies, 50/147 (34.0%) were carriers and 6/50 (12.0%) had invasive LBC (**Figure 1B**). In females with an unknown genotype, we identified 6 BC cases (15.8%). Median age of BC diagnosis was 49 years of age for carriers (IQR 42–59) and 46 years for unknown genotype (IQR 45–50). Also, no BC cases were identified in the noncarriers group. (**Figure 1B; Figure 2B; Supplemental Table 3**)

In both lifetime cancer risk estimates, GC or BC cases with unknown histology were considered affected within the DGCLBC disease spectrum as either DGC or LBC. This classification was applied considering our population, since these

individuals were consistently first or second-degree relatives of confirmed carriers (data not shown), and their age of disease onset overlapped with those of cases with confirmed histology (**Figure 2**; **Supplemental Figure S1B**). Furthermore, all noncarriers were cancer-free supporting that unknown genotype individuals affected by GC or BC were likely carriers of the *CDH1* variant (**Figure 1B; Figure 2**).

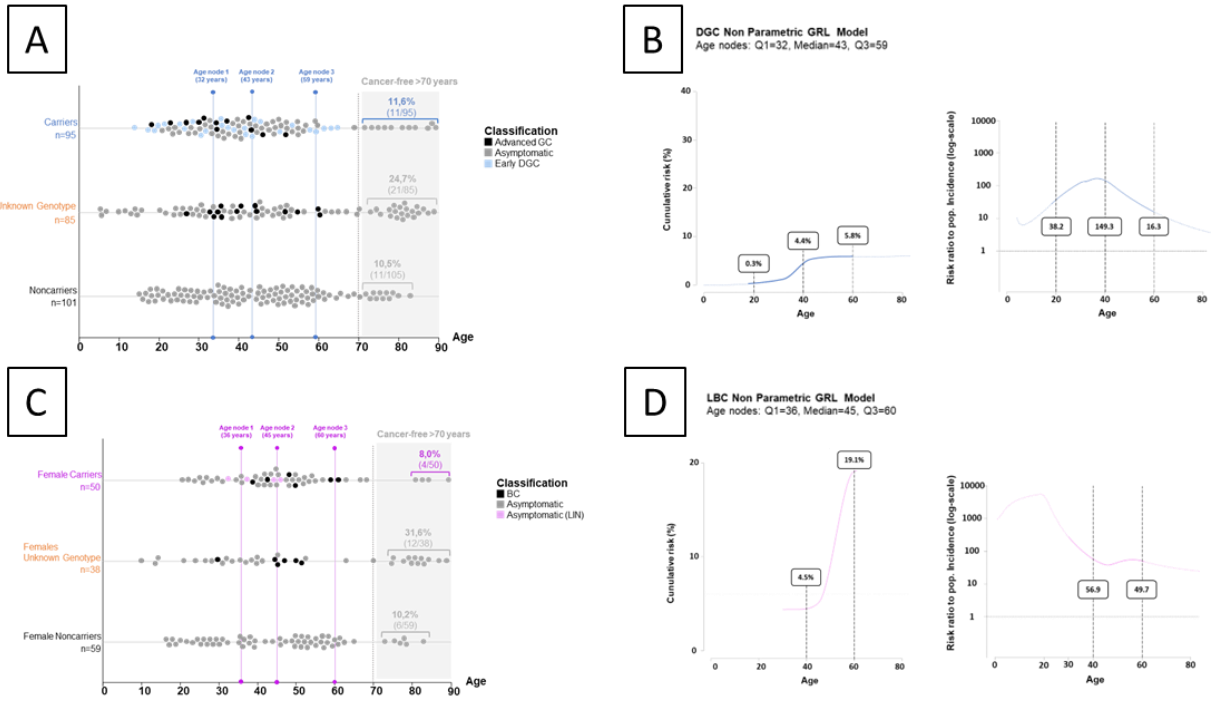


Figure 2. DGC and LBC lifetime risk (A) Distribution of cases used in DGC penetrance and identification of age nodes used in the NP-GRL model; (B) DGC cumulative and relatives risks by age estimated with the genotype restricted likelihood (GRL) method using a non-parametric function (NP GRL); (C) Distribution of cases used in LBC penetrance and identification of age nodes used in the NP-GRL model; (D) LBC cumulative and relatives risks by age estimated with the genotype restricted likelihood (GRL) method using a non-parametric function (NP GRL)

The cumulative risk of DGC development was calculated jointly for both males and females since there were not enough cases to retrieve adequate estimates by sex. Based on our estimates using the GRL method, DGC risk is 4.4% and 6.0% by the ages of 40 and 80, respectively (**Figure 2; Supplemental Table 4 and Supplemental Figure S3**). For the same ages, using carriers only, the

estimates provided by the Kaplan-Meier (KM) method were 10.3 (95%CI, 3.2%–16.9%) and 19.1% (95% CI, 8.0%–28.8%), respectively (**Supplemental Figure S1B**).

A supplemental analysis using GRL and KM methods considering individuals with Early DGC (identified in Figure 2A) as cancer-affected was performed. The cumulative risk of combined Early and Advanced GC using the GRL method at 40 and 80 years were 9.3% and 33.5%, respectively (**Supplemental Figure 2)**. Within our DGC risk estimates cohort, 41/285 (15.1%) of individuals were older than 70 years and remained asymptomatic for GC being 11/285 (3.6%) carriers of the *CDH1* c.1901 C>T variant (**Figure 2A**). The relative risk for DGC of *CDH1* carriers compared with the general population estimated with the GRL method decreased with age, ranging from 1236 at 20 years to 24.1 at 80 years of age **(Figure 2C, Supplemental Table 4)**.

For LBC estimates, the GRL method provided cumulative risks of 4.5% and 19.7% by the ages of 40 and 80, respectively. The relative risk for LBC for female carriers compared to the general population exhibits a changing trend nearly until 60 years old. From this point onwards, it decreases, reaching 26.1 at 80 years of age. Similarly, we calculated the cumulative risk with the KM method (**Supplemental Figure 1B, Supplemental Table 4).** A total of 22 out of the 147 (15.0%) females considered for the analysis were older than 70 years and remained free of BC symptoms with 2.7% (4/147) being positive carriers.

The remaining age-specific cumulative and relative risks for DGC and LBC calculated by both GRL and KM methods are provided in **Supplemental Table 4**. Using the leave-one-out strategy we could observe similar risk estimates when families were removed one by one, except when excluding large families with more than 45 individuals such as F2, F3 and F5 revealing their weight on the risk estimations (**Supplemental Figure 3)**.

Clinical Findings Analysis

Concerning the 195 individuals tested at CHUSJ, 90 of them (46.2%) received a positive genetic test for the *CDH1* c.1901 C>T variant but not all of them decided to be followed at this institution for disease management. In fact, 16/90 (17.8%) decided to undergo surveillance in a different hospital and 6/90 (6.7%) chose to forego surveillance, due to personal reasons or because they were asymptomatic older individuals. Additionally, 4/90 (4.4%) recently received the genetic test result and no surveillance exams were performed yet (**Figure 1A**). 74.4% of individuals

(64/90) were monitored for DGC, with 41.1% (37/90) among them being females also under surveillance for LBC.

Of the 64 carriers followed at CHUSJ, 8 (12.5%) individuals were clinically diagnosed with DGC before genetic testing. From the 56 (87.5%) asymptomatic under surveillance carriers, two cases (3.6%) were diagnosed with Advanced DGC and five individuals (8.9%) had positive signet ring cell carcinoma findings (pT1a lesions) at endoscopic surveillance. All cases diagnosed with Early DGC by endoscopic biopsy tested negative for lymph nodes (pT1aN0M0) and opted for TG with curative intent (**Figure 3A**). The number of biopsy fragments examined was similar between patients with positive and negative endoscopic findings.

Among the 49 individuals without previous positive endoscopic findings (normal biopsy result in at least one documented endoscopic exam), 24 (48.9%) decided to perform prophylactic total gastrectomy (PTG). Of these, 83.3% (20/24) received diagnosis of Early DGC and 3.3% (8/24) presented in situ lesions (pTis) after undergoing risk-reducing surgery (**Figure 3A; Table 1**). At a median follow-up of 4 years (IQR 1-7), the remaining 25/49 (51%) carriers remained asymptomatic after deciding to delay risk-reducing surgery. Their median age of latest endoscopic biopsy is 41 years (IQR 31-54).

Among the 10 patients that received a diagnosis of Advanced DGC, 6/10 (60%) presented distant metastasis and 70% (7/10) had a tumour grading classification of IV. From the 6 DGC patients that underwent curative gastrectomy, 3 survived. (**Figure 3A; Table 1**). The median age of death caused by DGC was 36 years (IQR 20-43)

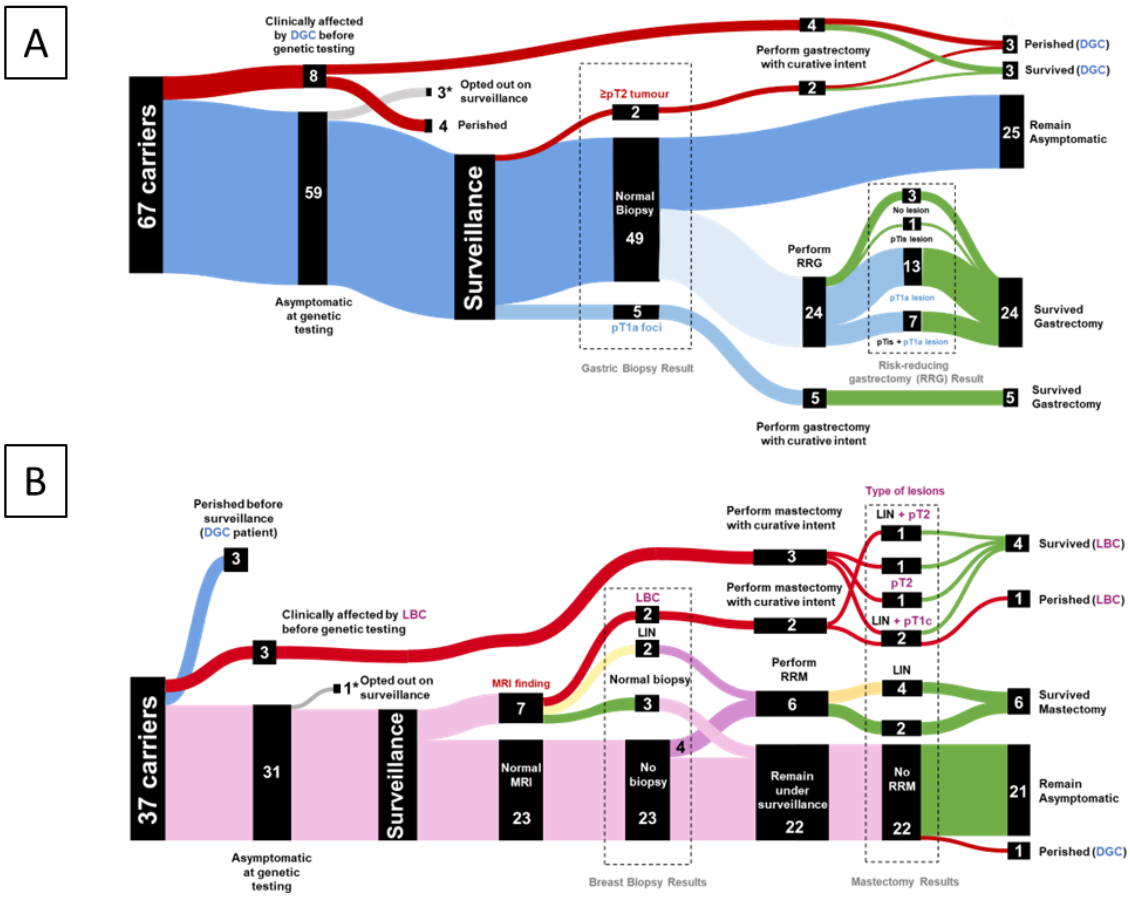


Figure 3. Profile of clinical management for DGC and LBC *CDH1* c.1901 C>T carriers followed at CHUSJ (A) Alluvial diagram depicting the distribution of cases under follow-up at CHUSJ for DGC disease management and surveillance (B) Alluvial diagram depicting the distribution of cases under follow-up at CHUSJ for LBC disease management and surveillance.

Moreover, we could perform a retrospective study based on the pathological reports data from patients under management at CHUSJ. Available information was summarized in **Table 1 and Table 2** for DGC and LBC, respectively. We found that 57.8% (37/64) of carriers had a positive history of HP infection during the follow-up period **(Table 1)**. The most reported histological finding in endoscopic biopsies was gastritis affecting 44/64 (68.8%) of carriers, especially frequent in 24 (82.8%) of the 29 asymptomatic individuals that either remain under surveillance without positive findings (n=25) or had no lesions after PTG (n=4). The presence of intestinal metaplasia (n=5) or foveolar hyperplasia (n=5) were also mainly reported in the asymptomatic carriers **(Table 1).** No significant association was identified between histological findings in endoscopic biopsies, the number of fragments analyzed, HP status and the diagnosis of Early DGC. The presence of gastric ulcer

was significantly higher in patients with Advanced DGC than in those with normal endoscopic biopsies ($P < 0.05$) reported in **Table 1**.

Table 1. Relevant clinical findings in DGC endoscopic screening, surveillance, and total gastrectomy of *CDH1* c.1901 C>T asymptomatic and affected carriers.

Cases followed at a single-center	Total (N=64)	Advanced DGC (N=10)	Early DGC (pT1a) (N=25)	Asymptomatic (N=29)	*p-value*
Female (n) (%)	36 (56.2)	5 (50.0)	15 (60.0)	16 (55.2)	0.840
Age at Molecular Study (Years)					
Median [Q1,Q3]	37.0 [26.5, 52.0]	34.0 [31.5, 41.0]	34.0 [25.0, 44.0]	39.0 [29.0, 52.0]	0.521
Pathway of Care (n) (%)					
Performed Curative Gastrectomy	8 (12.5)	3 (30.0)	5 (20.0)	0 (0)	
Performed Risk Reducing Gastrectomy	24 (37.5)	0 (0)	20 (80.0)	4 (13.8)	<0.001
Under Surveillance	25 (39.1)	0 (0)	0 (0)	25 (86.2)	
Perished by Disease	7 (10.9)	7 (70.0)	0 (0)	0 (0)	
Age at Latest Endoscopic Biopsy (Years)					
Median [Q1,Q3]	39.0 [31.0, 52.0]	35.0 [31.5, 41.0]	38.0 [27.0, 45.0]	41.0 [31.0,55.0]	0.274
***Helicobacter Pylori* History (n) (%)**					
Yes	38 (59.4)	4 (40.0)	17 (68.0)	17 (58.6)	0.919
Cases with missing data (n) (%)	7 (10.9)	4 (50.0)	1 (4.0)	2 (6.9)	
Neoplasic Lesion in Endoscopic Biopsy (n) (%)					
≥pT2	10 (15.6)	10 (100)	0 (0)	0 (0)	<0.001
pT1a	5 (7.8)	0 (0)	5 (20.0)	0 (0)	
No Findings	49 (76.6)	0 (0)	20 (80.0)	29 (100)	
Gastritis in Endoscopic Biopsy (n) (%)					
Yes	44 (68.8)	3 (30.0)	17 (68.0)	24 (82.8)	0.542
Cases with missing data (n) (%)	13 (20.3)	6 (60.0)	5 (20.0)	2 (6.9)	
Other Histological Findings in Endoscopic Biopsy (n) (%)					
Gastric Ulcer	5 (7.8)	4(40.0)	0 (0)	1 (3.4)	<0.001
Intestinal Metaplasia	8 (12.5)	1 (10.0)	2 (8.0)	5 (17.2)	0.632
Foveolar Hyperplasia	6 (9.4)	0 (0)	1 (4.0)	5 (17.2)	0.268
No Findings	34 (53.1)	0 (0)	17 (68.0)	17 (58.6)	0.001
Cases with missing data (n) (%)	12 (18.8)	5 (50.0)	5 (20.0)	2 (6.9)	
Cases diagnosed with DGC	**Total (N=35)**	**Advanced DGC (N=10)**	**Early DGC (pT1a) (N=25)**	**Asymptomatic (N=0)**	***p-value***
Timing of GC Diagnosis (n) (%)					
Before Genetic Testing	8 (22.8)	8 (80.0)	0 (0)	0 (0)	<0.001
Endoscopic Surveillance	7 (20.0)	2 (20.0)	5 (20.0)	0 (0)	
Gastrectomy	20 (57.1)	0 (0)	20 (80.0)	0 (0)	
Age of Onset at GC Diagnosis (Years)					
Median [Q1,Q3]	35.0 [28.5, 44.0]	35.0 [30.8, 41.0]	39.0 [27.0, 45.0]	-	0.365
Tumour Staging at GC diagnosis (n) (%)					
<pT2	25 (71.4)	0 (0)	25 (100.0)	0(0)	<0.001
≥pT2	10 (28.6)	10(100.0)	0 (0)	0(0)	
Tumour Grading at GC diagnosis (n) (%)					
IV	7 (20.0)	7 (70.0)	0(0)	0(0)	
IIB	1 (2.9)	1 (10.0)	0(0)	0(0)	<0.001
IB	1 (2.9)	1 (10.0)	0(0)	0(0)	
IA	24 (68.6)	-	25 (100.0)	0(0)	
Cases with missing data (n) (%)	1 (2.9)	1 (10.0)	0(0)	0(0)	
GC with Lymph Node Metastasis (n) (%)					
Yes	3 (4.7)	3 (30.0)	0 (0)	0(0)	0.004
Cases with missing data (n) (%)	4(6.3)	4 (40.0)	0(0)		
GC Distant Metastasis (n) (%)					
Yes	6 (17.1)	6 (60.0)	0 (0)	-	<0.001
Cases with missing data (n) (%)	1 (2.9)	1			
Age of Death (Years)					
Median [Q1,Q3]	36.0 [27.5, 37.5]	36.0 [27.5, 37.5]	-	-	NA
Cases submitted to gastrectomy	**Total (N=35)**	**Advanced DGC (N=6)**	**Early DGC (pT1a) (N=25)**	**Asymptomatic (N=4)**	***p-value***
Years From Genetic Testing to Gastrectomy					
Median [Q1,Q3]	1.0 [0, 2.0]	0 [0, 0.5]	1.0 [0, 2.0]	1.5 [1.0, 2.0]	0.049
Age at Gastrectomy (Years)					
Median [Q1,Q3]	39.5 [29.2, 46.5]	36.5 [33.5, 41.8]	39.5 [29.0, 46.5]	49.5 [39.0, 59.0]	0.421
Premalignant Lesions (pTis) in Gastrectomy (n) (%)	8 (22.9)	0 (0)	7 (28.0)	1 (25.0)	0.449
Neoplasic Lesions in Gastrectomy (n) (%)					
≥pT2	6 (17.1)	6 (60.0)	0 (0)	0 (0)	<0.001
pT1a	25 (71.4)	0 (0)	25 (100.0)	0 (0)	
No lesions	4 (11.4)	0 (0)	0 (0)	4 (100.0)	
Gastritis in Gastrectomy (n) (%)					
Yes	21 (60.0)	3 (50.0)	16 (64.0)	2 (50.0)	0.764
Cases with missing data (n) (%)	9 (25.7)	3 (50.0)	5 (20.0)	1 (25.0)	
Histological Findings in Gastrectomy (n) (%)					
Gastric Ulcer	1 (2.9)	1 (16.7)	0 (0)	0 (0)	0.764
Foveolar Hyperplasia	6 (17.1)	0 (0)	4 (16.0)	2 (50.0)	0.386
Intestinal Metaplasia	7 (20.0)	0 (0)	5 (20.0)	2 (50.0)	0.386
Tufting	4 (11.4)	0 (0)	3 (12.0)	1 (25.0)	0.562
No Findings	13 (37.1)	1 (16.7)	12 (48.0)	0 (0)	0.068
Cases with missing data (n) (%)	12 (34.3)	5 (83.3)	6 (24.0)	1 (20.7)	

The median age at which TGs were performed was 40 years (IQR 14-64), excluding the case with less than 1 year after surgery. Again, the most reported histological finding was gastritis present in 79.2% (19/24) of individuals submitted to PTG including 3 of the 4 carriers with no lesions identified in the full stomach **(Table 1)**.

Next, we analyze the rates of short-and long-term events associated with PTG for the 23 patients that had at least one and a half years of follow-up (**Supplemental Table 5**). Of the 23 cases, only one female individual (4.3%) had postoperative complications within the first month of recovery. A second patient required procedural intervention due to two hernias more than 12 months after gastrectomy. The most frequently described symptom within the first year affecting 30.4% (7/23) of cases was nausea, followed by asthenia in 17.4% (4/23) (**Supplemental Table 5**).

After more than 12 months post-surgery, the most recurrent symptoms reported were dumping syndrome and muscle cramps, both occurring in 17.4% (4/23) of patients with a median follow-up time of 7 years (IQR 5.5-8.0). (**Supplemental Table 5**). For 21 individuals, data on Body Max Index (BMI) alterations before and after PTG were available (**Supplemental Figure 4**). At least 57.1% (12/21) had excess of weight at the preoperative stage (BMI over 25), this changed to 23.8% (5/21) at least 1.5 years after. In total, 15/21 (71.4%) were reported with a normal BMI (18.5-24.9) at the latest nutritional consultation. All patients were prescribed with commercially available multivitamin supplements and the micronutrient deficiencies reported were vitamin D (10/23; 43.5%) iron (9/23;39.1%), vitamin B12 (7/23;30.4%), magnesium (7/23; 30.4%) and a low level of total proteins was reported in one patient (1/23; 4.3%) (**Supplemental Figure 5**).

Regarding LBC surveillance and management, from the 37 female carriers followed at CHUSJ, only 33 (91.9%) were managed for LBC development since 3 perished from Advanced DGC and 1 opted out on surveillance (**Figure 3B**). Four (12.1%) of the 33 were clinically diagnosed with LBC before genetic testing, performed curative mastectomy and survived. An abnormal Magnetic Resonance Imaging (MRI) finding lead to a breast biopsy in seven (23.3%) of the 30 asymptomatic female carriers resulting in the diagnosis of two (6.6%) LBC and two (6.6%) Lobular Intraepithelial Neoplasia (LIN) cases. The remaining three carriers with normal biopsy decided to remain under surveillance. Overall, 28 (84.8%) of the

33 individuals are asymptomatic for BC with a median follow-up time of 5.5 years (IQR 1.0-8.0). 78.6% (22/28) of the asymptomatic carriers decided to remain under surveillance instead of prophylactic mastectomy (although one patient died from DGC during follow-up) with a median age of 45 years (IQR 30-57). The remaining six (21.4%) carriers performed risk-reducing mastectomies (RRM) at a median age of 43 years (IQR 32-63), with four (14.2%) cases in the absence of previous abnormal findings or lesions. No LBC cases were identified at RRM, but 4 (66.6%) of the 6 carriers presented LIN (two of them already identified by previous biopsy). In total, five (15.2%) of the initially 33 women under management were diagnosed with LBC at a median age of 49 years (IQR 42-61), with one death at 61 years of age with distant metastasis **(Table 2; Figure 3B).** Besides DGC and LBC, one male carrier was diagnosed with prostate cancer at 57 years before genetic testing, a female carrier was diagnosed with a neuroendocrine tumour during TG after Early DGC diagnosis during in endoscopic surveillance and an Oligodendroglioma stage II cancer affected the proband of Family 10.

Table 2. Relevant clinical findings in LBC screening, surveillance, and total mastectomy of *CDH1* c.1901 C>T carriers

Female cases followed at a single-center	Overall (N=33)	Lobular Breast Cancer (N=5)	Asymptomatic (N=28)
Age at Molecular Study (Years)			
Median [Q1,Q3]	42.0 [15.0, 32.0]	50.0 [49.0, 59.0]	40.5 [24.8, 52.0]
Pathway of Care (n) (%)			
Perished by Disease	1 (3.0)	1 (20.0)	0 (0)
Performed Curative Mastectomy	5 (15.1)	5 (100.0)	0 (0)
Performed Prophylactic Mastectomy	6 (21.2)	0 (0)	6 (21.4)
Under Surveillance	22 (66.7)	0 (0)	22 (78.6)
Abnormal finding on MRI (n) (%)	7 (21.2)	2 (40.0)	5 (17.9)
Performed Biopsy (n) (%)	9 (27.3)	4 (80.0)	5 (17.9)
Cases with missing data (n) (%)	1 (3.0)	1 (20.0)	0 (0)
Age at Latest Breast Biopsy (Years)			
Median [Q1,Q3]	43.0 [34.0, 49.0]	49.0 [47.2, 51.5]	34.0 [31.0, 43.0]
Pre-neoplasic Lesions (LIN) Breast Biopsy (n) (%)	2 (6.1)	0 (0)	2 (7.1)
Neoplastic Lesions at Breast Biopsy (n) (%)			
pT2	2 (6.1)	2 (40.0)	0 (0)
pT1c	1 (3.0)	1 (20.0)	0 (0)
pT1	1 (3.0)	1 (20.0)	0 (0)
Females carriers submitted to mastectomy	**Overall (N=11)**	**Lobular Breast Cancer (N=5)**	**Asymptomatic (N=6)**
Age at Mastectomy (Years)			
Median [Q1,Q3]	44.0 [42.5, 54.0]	54.0 [47.5, 59.5]	43.0 [39.5, 45.5]
Mastectomy Unilateral or Bilateral (n) (%)			
Bilateral	10 (90.1)	4 (80.0)	6 (100.0)
Unilateral*	1 (9.1)*	1 (20.0)*	0 (0)
Pre-neoplasic Lesions (LIN) in Mastectomy (n) (%)	7 (63.6)	3 (60.0)	4 (14.3)
Neoplastic Lesions in Mastectomy (n) (%)	5 (45.5)	5 (100.0)	0 (0)
Females carriers diagnosed with LBC	**Overall (N=5)**	**Lobular Breast Cancer (N=5)**	**Asymptomatic (N=0)**
Timing of Breast Cancer Detection (n) (%)			
Before Genetic Testing	3 (60.0)	3 (60.0)	0 (0)
During surveillance	2 (40.0)	2 (40.0)	0 (0)
Breast Cancer Age Onset (Years)			
Mean (SD)	51.8 (8.0)	51.8 (8.0)	0 (0)
Median [Q1,Q3]	49.0 [48.0, 59.0]	49.0 [48.0, 59.0]	0 (0)
Tumour Stage (n) (%)			
T2	2 (40.0)	2 (40.0)	0 (0)
T1c	2 (40.0)	2 (40.0)	0 (0)
Cases with missing data (n) (%)	1 (20.0)	1 (20.0)	0 (0)
Lymph Node Metastasis (n) (%)			
N2	1 (20.0)	1 (20.0)	0 (0)
N1	1 (20.0)	1 (20.0)	0 (0)
N0	2 (40.0)	2 (40.0)	0 (0)
Cases with missing data (n) (%)	1 (20.0)	1 (20.0)	0 (0)
Distant Metastasis (n) (%)			
M1	1 (20.0)	1 (20.0)	0 (0)
M0	3 (60.0)	3 (60.0)	0 (0)
Cases with missing data (n) (%)	1 (20.0)	1 (20.0)	0 (0)
Laterality of LBC (n) (%)			
Unilateral (Left)	2 (40.0)	2 (40.0)	0 (0)
Unilateral (Right)	2 (40.0)	2 (40.0)	0 (0)
Unknown	1 (20.0)	1 (20.0)	0 (0)
Age of Death (Years)			
Median [Min, Max]	61.0 [61.0, 61.0]	61.0 [61.0, 61.0]	-

* Case with unilateral mastectomy with LBC diagnosed on the right breast and performed breast reduction on the left

DISCUSSION

Penetrance studies of *CDH1* variants show significant variability depending on the on the examined cohort. The identification of founder populations offers a unique opportunity to explore disease development in conditions with minimal impact of genetic heterogeneity and environment factors. Furthermore, the identification of founder germline variants can impact disease management on a nationwide perspective, with considerable impact on economic systems due to genetic testing exams and particular pathways of care for individuals at risk. In the Northern region of Portugal, several families have been identified carrying the same *CDH1* genetic variant. Over the last 20 years, large kindreds with a shared core uniform haplotype encompassing the *CDH1* c.1901 C>T variant have been followed at CHUSJ. This creates a unique opportunity to study lifetime DGC and LBC risks for the most recurrent *CDH1* variant described in Europe. As the probands were tested and followed-up with their relatives at a single institution, we were able to obtain detailed clinical data that was close to exhaustive. This allowed us to take advantage of the GRL method requiring extended family data to obtain unbiased penetrance estimates taking into account all the available information.

To our knowledge, this is the largest single-center DGLBCS penetrance study involving families carrying a specific *CDH1* variant. Prior to this, only Kaurah et al., had conducted penetrance estimates on a founder variant (*CDH1* c.2398delC) from the Newfoundland region with a cumulative risk by 75 years for BC of 52%, and for advanced GC of 40% for males and 63% for women.

We developed a comprehensive database with clinical data from 285 individuals for specific lifetime risk assessments in DGC and LBC, that can be filtered according to other types of intended analysis (Figure 1A). The disease phenotypes and age of onset distributions reveals that GC occurrence is the most common observed cancer being diagnosed in 9.1% all families.

Cumulative and relative risk estimations for GC and BC for *CDH1* carriers have been calculated in several publications, and along the years, they are becoming less biased through family ascertaining and limited cases on cohort composition.

Description of cumulative risk has been steadily decreasing since the first *CDH1* penetrance study by Pharoah et al reporting 67% of GC risk at age 80 for

males and 83% for women, to values varying between 42.0-37.2% in men and 33.0-24.7% in women at 80 years of age, described by Xicola et al and Roberts et al. Cumulative risk of LBC also showed some variation starting at 39.0% in the first study and now varying between 55.0-42.9% for the same age.

Carriers' management follow specific guidelines adopted worldwide, with risk reducing surgeries interrupting the natural disease development. The consequences of individuals being closely followed and benefiting from prophylactic interventions are similarly reflected in all cohorts used for *CDH1* cancer risk estimates.

In this work, the cumulative risk for DGC at 80 years old is predicted to be 6.0% for both males and females while for LBC risk is 14.7% in women. These are the lowest cancer estimates ever described, especially for a cohort enriched in families fulfilling DGLBCS criteria, reinforcing the need of a complete family history with access to cancer-free carriers. The depth of pedigree knowledge is crucial for informed choices regarding preventive measures, genetic testing, and personalized healthcare strategies.

This study also suggests that carriers of *CDH1* c.1901 C>T variant have particular lower risks of disease development than what has been reported in other studies for *CDH1* variants(São José et al., 2023). It is important to acknowledge that comparisons between risk estimations are limited since each study has their own cohort, calculation methods and constraints.

Individuals with Early DGC (identified with pT1a signet ring cell carcinoma lesions detected by an endoscopic biopsy or gastrectomy) were not considered affected in the main analysis (**Figure 2A and Figure 2B**). However, pT1a lesions are characterized as early-stage cancers and their detection during surveillance leads to total gastrectomy despite their suspected indolent behavior(Hayakawa et al., 2015; Humar et al., 2007; Mimata et al., 2011; Nikaido et al., 2021). Therefore, we explored the potential impact on penetrance estimates when these individuals are also considered as affected. As expected, the cumulative risk highly increased, reaching 33.5% by 80 years of age (**Supplemental Figure S2,**). This finding questions the clinical significance attributed to the risk of pT1a lesions progressing to advanced disease. Notably, among 24 PTGs, 83.3% (20/24) resulted in an Early Diffuse Gastric Cancer (DGC) diagnosis despite previous negative endoscopic biopsy results. For the 11 individuals (11.6% of the 95 carriers) who did not undergo PTG older than 70 years and remain cancer-free at a median age of 81 years (IQR

71-90), we hypothesize that they present indolent Early DGC lesions or throughout their lives pT1a dynamically formed and degenerated. This is further supported by the observation that in the unknown genotype group, at a higher risk for pT1a progression, there is not a significant increase in number of GC cases observed. In the carriers' group, 12.6% (12/95) experienced GC, with 26.3% opting for total gastrectomy and interrupting disease progression, compared to 16.4% GC cases diagnosed in the unknown genotype group without any disease management, leaving them completely exposed to the risk of disease.

Furthermore, advocating that PTG should be considered on an individual basis, and considering family history, gains support in our cohort. Notably, our carriers and those with an unknown genotype do not present GC cases beyond the age of 60. This challenges the conventional perspective of reporting lifetime risk, often referencing an 80-year age range, incongruent with the typical onset behavior in these families. Additionally, it questions the utility of performing PTG in three carriers without previous positive endoscopic findings at ages of 61 and 64 years old (**Figure 2A**).

To align with more realistic patterns, we propose presenting cancer risk estimates between 20 and 60 years of age, a critical interval where most cases manifest, and statistical models can provide more accurate predictions (**Figure 2B&D**).

This extensive retrospective study assessed clinical information collected over 20 years. The inclusion of data for such an extended period is an important advantage to evaluate the effectiveness of the management strategies applied and redefine them to improve carriers' quality of life.

Despite that not all individuals used for lifetime risk calculations were under follow-up at our medical center, we collected clinical documents or health records on 70.5% (67/95) of the identified carriers considered in this study. A total of 35 gastrectomies and 11 mastectomies were performed, reflecting compliance with the current IGCLC recommendations for stomach removal independently of endoscopic surveillance results while a more deliberate decision-making process is employed for breast.

In our cohort, endoscopic sensitivity was very low at detecting early DGC lesions in asymptomatic individuals, despite 83.3% (20/24) were diagnosed with pT1a lesions when the entire gastric mucosa was pathologically evaluated after

PTG. The high rate of false positives agrees with other studies(Buckley et al., 2023; Castro et al., 2020; Friedman et al., 2021; Vos et al., 2020). Subtle gastric mucosal abnormalities were reported, with gastritis and HP infection affecting more than half of our carriers. However, we did not find any association between mucosal abnormalities and the presence early lesions. Moreover, there are no reports of pale lesions of mucosa being targets of biopsy in this cohort.
In recent and independent prospective longitudinal studies(Asif et al., 2023; Friedman et al., 2021; Lee et al., 2023) of endoscopic surveillance using multiple biopsies, random biopsies were reported as effective in detecting advanced lesions. Throughout our follow-up period, only one carrier developed advanced DGC concomitant with the presence of a gastric ulcer, a progress successfully identified through endoscopy. Interestingly, in accordance with Asif et al., gastric ulcers were the only mucosal abnormality associated with advanced GC lesions ($p<0.05$). If performed in a reference center, endoscopy may be a reasonable alternative to PTG but further studies with greater statistical power and longer follow-up periods for asymptomatic patients postponing surgery are needed.

We were able to infer that in the absence of cancer or premalignant lesions, clinical reports of TG were more detailed regarding other gastric mucosa abnormal findings. This highlights the need for a gastrectomy report template with clear and consistent description of key findings, ideally adopted by all DGLBCS reference centers.

Given that PTG is still considered the only effective risk reducing measure, we were interested in exploring its short- and long-term impact in our cohort. There were no cases of mortality or major morbidity, and a low rate of serious complications were reported being the most common outcomes nausea, asthenia, dumping syndrome, muscle cramps and vitamins deficiency.

LIMITATIONS

Despite being one of the largest cohorts reported so far for a *CDH1* founder population, the main limitations of this study are related with the sample size (inherent to the rarity of this *CDH1* variant) and its retrospective design, limited access to older medical or histological reports, which would confirm all cancer diagnoses and its respective histology.

For the lifetime risk calculations since our study was based on inferred incidences of DGC and LBC in the general population of Northern Portugal, efforts should be made to collect real epidemiological data from national registries regarding these histological types of GC. The same rational is applicable to the exact allele frequency of the *CDH1* c.1901 C>T variant.

Conclusion

The development of clinical guidelines for patient management involves the integration of lifetime risk and relies on the available options for prevention, surveillance, and risk reduction. Our extensive retrospective study defines the lowest risk for DGC and LBC ever described, specific to carriers of a founder-effect variant in *CDH1*. Moreover, our study is outstanding not only due to the large number of carriers involved, but also covering all levels of the pathway of care established for patients and families with DGLBCS. Our findings support that surveillance can be an acceptable alternative for PTG especially in the context of a low penetrant *CDH1* variant. This is a first step in the context of personalized management of healthy carriers and can contribute to improved outcomes and quality of care of patients and their families.

REFERENCES

Ansari, S., Gantuya, B., Tuan, V., & Yamaoka, Y. (2018). Diffuse Gastric Cancer: A Summary of Analogous Contributing Factors for Its Molecular Pathogenicity. *International Journal of Molecular Sciences*, *19*(8), 2424. https://doi.org/10.3390/ijms19082424

Asif, B., Sarvestani, A. L., Gamble, L. A., Samaranayake, S. G., Famiglietti, A. L., Fasaye, G.-A., Quezado, M., Miettinen, M., Korman, L., Koh, C., Heller, T., & Davis, J. L. (2023). Cancer surveillance as an alternative to prophylactic total gastrectomy in hereditary diffuse gastric cancer: a prospective cohort study. *The Lancet Oncology*, *24*(4), 383–391. https://doi.org/10.1016/S1470-2045(23)00057-8

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., … Oliveira, C. (2021). The CDH1 c.1901C>T Variant: A Founder Variant in the Portuguese Population with Severe Impact in mRNA Splicing. *Cancers*, *13*(17), 4464. https://doi.org/10.3390/cancers13174464

Bonadona, V., Bonaïti, B., Olschwang, S., Grandjouan, S., Huiart, L., Longy, M., Guimbaud, R., Buecher, B., Bignon, Y.-J., Caron, O., Colas, C., Noguès, C., Lejeune-Dumoulin, S., Olivier-Faivre, L., Polycarpe-Osaer, F., Nguyen, T. D., Desseigne, F., Saurin, J.-C., Berthet, P., … French Cancer Genetics Network, for

the. (2011). Cancer Risks Associated With Germline Mutations in MLH1, MSH2, and MSH6 Genes in Lynch Syndrome. *JAMA*, *305*(22), 2304–2310. https://doi.org/10.1001/jama.2011.743

Buckley, K. H., Tondon, R., Guilford, P., & Katona, B. W. (2023). Non-targeted biopsies in hereditary diffuse gastric cancer: Necessary, but not enough. *Translational Cancer Research*, *12*(7), 1883–1886. https://doi.org/10.21037/tcr-23-541

Carayol, J., & Bonaïti-Pellié, C. (2004). Estimating penetrance from family data using a retrospective likelihood when ascertainment depends on genotype and age of onset. *Genetic Epidemiology*, *27*(2), 109–117. https://doi.org/10.1002/gepi.20007

Castro, R., Lobo, J., Pita, I., Videira, F., Pedro-Afonso, L., Dinis-Ribeiro, M., & Brandao, C. (2020). Random biopsies in patients harboring a CDH1 mutation: time to change the approach? *Revista Española de Enfermedades Digestivas*, *112*. https://doi.org/10.17235/reed.2020.6720/2019

Christgen, M., Steinemann, D., Kühnle, E., Länger, F., Gluz, O., Harbeck, N., & Kreipe, H. (2016). Lobular breast cancer: Clinical, molecular and morphological characteristics. *Pathology - Research and Practice*, *212*(7), 583–597. https://doi.org/10.1016/j.prp.2016.05.002

Coudert, M., Drouet, Y., Delhomelle, H., Svrcek, M., Benusiglio, P. R., Coulet, F., Clark, D. F., Katona, B. W., van Hest, L. P., van der Kolk, L. E., Cats, A., van Dieren, J. M., Nehoray, B., Slavin, T., Spier, I., Hüneburg, R., Lobo, S., Oliveira, C., Boussemart, L., … Colas, C. (2022). First estimates of diffuse gastric cancer risks for carriers of *CTNNA1* germline pathogenic variants. *Journal of Medical Genetics*, *59*(12), 1189–1195. https://doi.org/10.1136/jmg-2022-108740

Dossus, L., & Benusiglio, P. R. (2015). Lobular breast cancer: incidence and genetic and non-genetic risk factors. *Breast Cancer Research*, *17*(1), 37. https://doi.org/10.1186/s13058-015-0546-7

Friedman, M., Adar, T., Patel, D., Lauwers, G. Y., Yoon, S. S., Mullen, J. T., & Chung, D. C. (2021). Surveillance Endoscopy in the Management of Hereditary Diffuse Gastric Cancer Syndrome. *Clinical Gastroenterology and Hepatology*, *19*(1), 189–191. https://doi.org/10.1016/j.cgh.2019.10.033

Gallanis, A. F., Gamble, L. A., Samaranayake, S. G., Lopez, R., Rhodes, A., Rajasimhan, S., Fasaye, G.-A., Juma, O., Connolly, M., Joyce, S., Berger, A., Heller, T., Blakely, A. M., Hernandez, J. M., & Davis, J. L. (2024). Costs of Cancer Prevention: Physical and Psychosocial Sequelae of Risk-Reducing Total Gastrectomy. *Journal of Clinical Oncology*, *42*(4), 421–430. https://doi.org/10.1200/JCO.23.01238

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., … Huntsman, D. G. (2015). Hereditary Diffuse Gastric Cancer Syndrome. *JAMA Oncology*, *1*(1), 23. https://doi.org/10.1001/jamaoncol.2014.168

Hayakawa, Y., Ariyama, H., Stancikova, J., Sakitani, K., Asfaha, S., Renz, B. W., Dubeykovskaya, Z. A., Shibata, W., Wang, H., Westphalen, C. B., Chen, X., Takemoto, Y., Kim, W., Khurana, S. S., Tailor, Y., Nagar, K., Tomita, H., Hara, A.,

Sepulveda, A. R., … Wang, T. C. (2015). Mist1 Expressing Gastric Stem Cells Maintain the Normal and Neoplastic Gastric Epithelium and Are Supported by a Perivascular Stem Cell Niche. *Cancer Cell*, *28*(6), 800–814. https://doi.org/10.1016/j.ccell.2015.10.003

Humar, B., Fukuzawa, R., Blair, V., Dunbier, A., More, H., Charlton, A., Yang, H. K., Kim, W. H., Reeve, A. E., Martin, I., & Guilford, P. (2007). Destabilized Adhesion in the Gastric Proliferative Zone and c-Src Kinase Activation Mark the Development of Early Diffuse Gastric Cancer. *Cancer Research*, *67*(6), 2480–2489. https://doi.org/10.1158/0008-5472.CAN-06-3021

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., … Huntsman, D. (2007). Founder and Recurrent CDH1 Mutations in Families With Hereditary Diffuse Gastric Cancer. *JAMA*, *297*(21), 2360. https://doi.org/10.1001/jama.297.21.2360

Lee, C. Y. C., Olivier, A., Honing, J., Lydon, A.-M., Richardson, S., O'Donovan, M., Tischkowitz, M., Fitzgerald, R. C., & di Pietro, M. (2023). Endoscopic surveillance with systematic random biopsy for the early diagnosis of hereditary diffuse gastric cancer: a prospective 16-year longitudinal cohort study. *The Lancet Oncology*, *24*(1), 107–116. https://doi.org/10.1016/S1470-2045(22)00700-8

Mimata, A., Fukamachi, H., Eishi, Y., & Yuasa, Y. (2011). Loss of E-cadherin in mouse gastric epithelial cells induces signet ring-like cells, a possible precursor lesion of diffuse gastric cancer. *Cancer Science*, *102*(5), 942–950. https://doi.org/10.1111/j.1349-7006.2011.01890.x

Muir, J., Aronson, M., Esplen, M.-J., Pollett, A., & Swallow, C. J. (2016). Prophylactic Total Gastrectomy: a Prospective Cohort Study of Long-Term Impact on Quality of Life. *Journal of Gastrointestinal Surgery*, *20*(12), 1950–1958. https://doi.org/10.1007/s11605-016-3287-8

Nikaido, M., Kakiuchi, N., Miyamoto, S., Hirano, T., Takeuchi, Y., Funakoshi, T., Yokoyama, A., Ogasawara, T., Yamamoto, Y., Yamada, A., Setoyama, T., Shimizu, T., Kato, Y., Uose, S., Sakurai, T., Minamiguchi, S., Obama, K., Sakai, Y., Muto, M., … Seno, H. (2021). Indolent feature of Helicobacter pylori-uninfected intramucosal signet ring cell carcinomas with CDH1 mutations. *Gastric Cancer*, *24*(5), 1102–1114. https://doi.org/10.1007/s10120-021-01191-8

Pharoah, P. D. P., 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*(6), 1348–1353. https://doi.org/10.1053/gast.2001.29611

Phipps, A. I., Li, C. I., Kerlikowske, K., Barlow, W. E., & Buist, D. S. M. (2010). Risk Factors for Ductal, Lobular, and Mixed Ductal-Lobular Breast Cancer in a Screening Population. *Cancer Epidemiology, Biomarkers & Prevention*, *19*(6), 1643–1654. https://doi.org/10.1158/1055-9965.EPI-10-0188

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

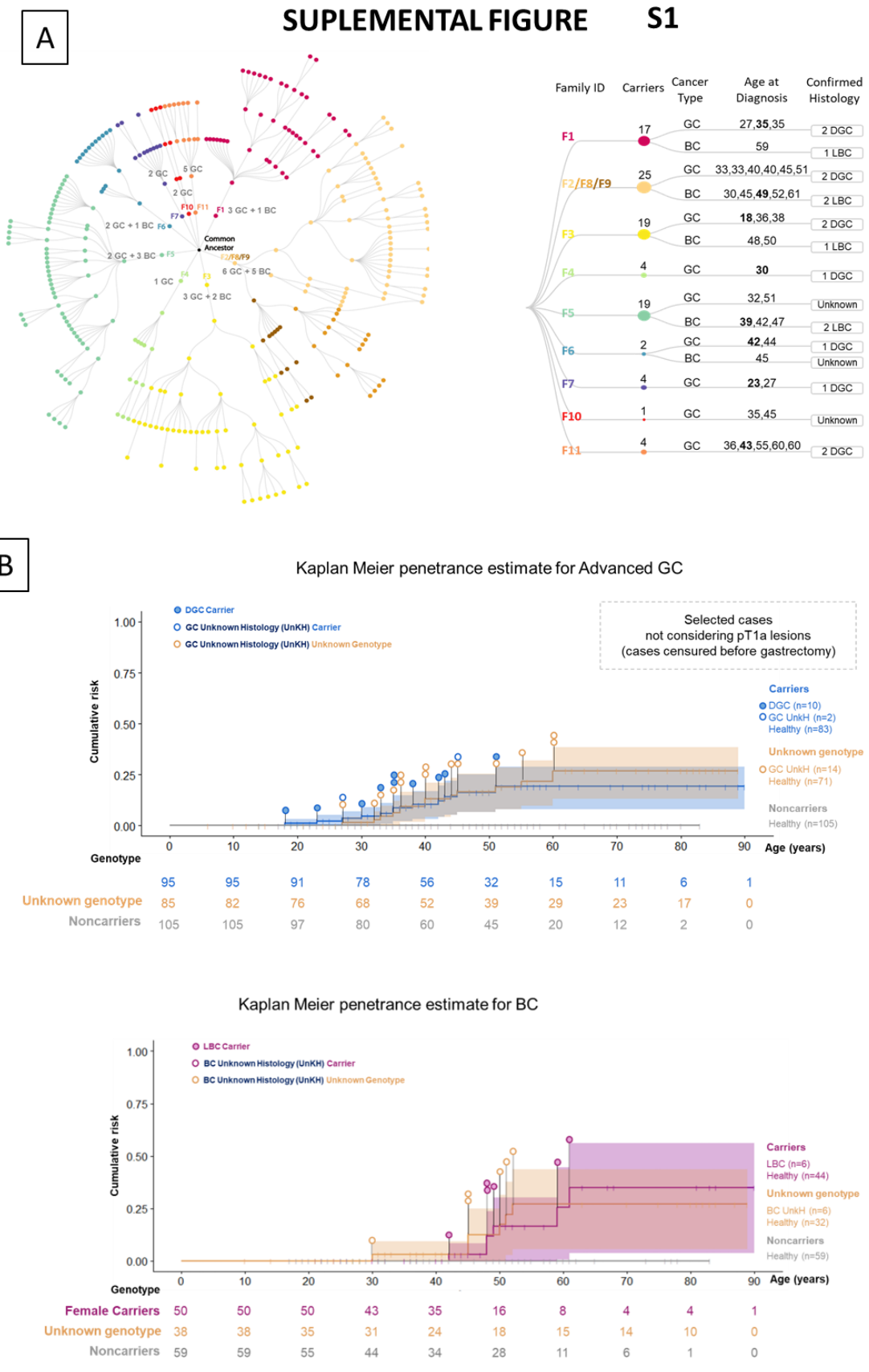
Multiple Gastric Cancers. *JAMA Oncology*, *5*(9), 1325. https://doi.org/10.1001/jamaoncol.2019.1208

São José, C., Garcia-Pelaez, J., Ferreira, M., Arrieta, O., André, A., Martins, N., Solís, S., Martínez-Benítez, B., Ordóñez-Sánchez, M. L., Rodríguez-Torres, M., Sommer, A. K., te Paske, I. B. A. W., Caldas, C., Tischkowitz, M., Tusié, M. T., Aretz, S., Capella, G., Castedo, S., de Voer, R. M., … Oliveira, C. (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*, *26*(5), 653–666. https://doi.org/10.1007/s10120-023-01395-0

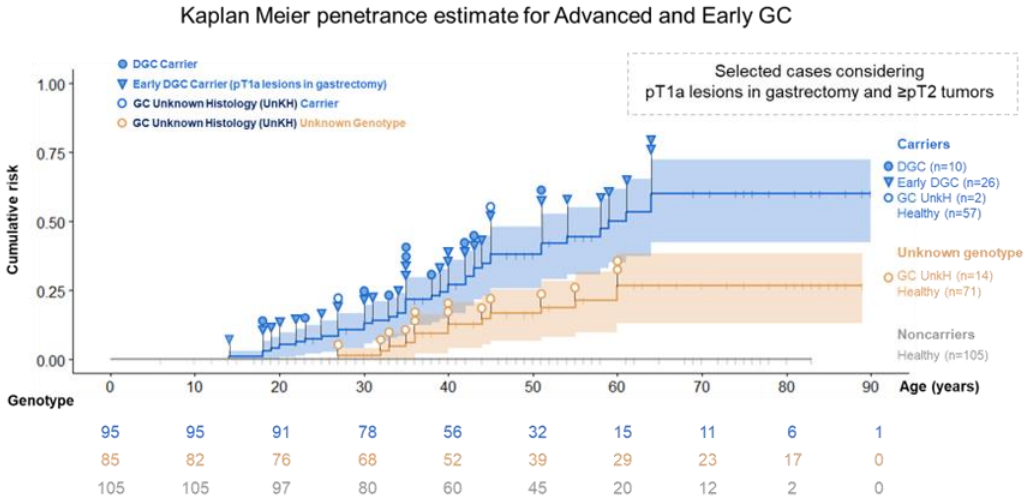
Vos, E. L., Salo-Mullen, E. E., Tang, L. H., Schattner, M., Yoon, S. S., Gerdes, H., Markowitz, A. J., Mandelker, D., Janjigian, Y., Offitt, K., Coit, D. G., Stadler, Z. K., & Strong, V. E. (2020). Indications for Total Gastrectomy in *CDH1* Mutation Carriers and Outcomes of Risk-Reducing Minimally Invasive and Open Gastrectomies. *JAMA Surgery*, *155*(11), 1050. https://doi.org/10.1001/jamasurg.2020.3356

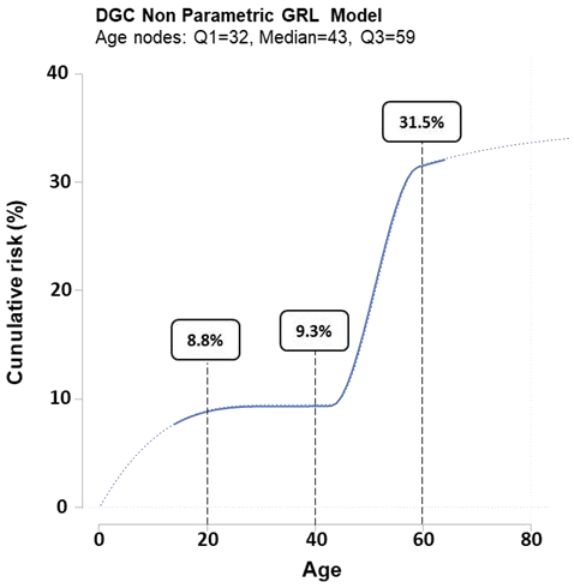
Xicola, R. M., Li, S., Rodriguez, N., Reinecke, P., Karam, R., Speare, V., Black, M. H., LaDuca, H., & Llor, X. (2019). Clinical features and cancer risk in families with pathogenic *CDH1* variants irrespective of clinical criteria. *Journal of Medical Genetics*, *56*(12), 838–843. https://doi.org/10.1136/jmedgenet-2019-105991

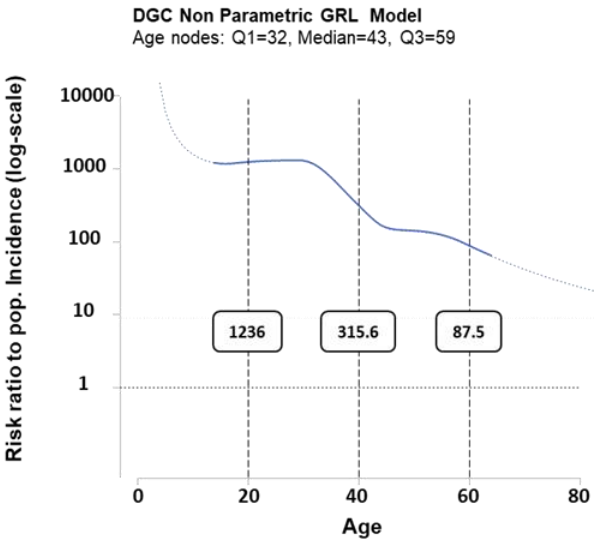
Supplementary Material:



SUPLEMENTAL FIGURE S2

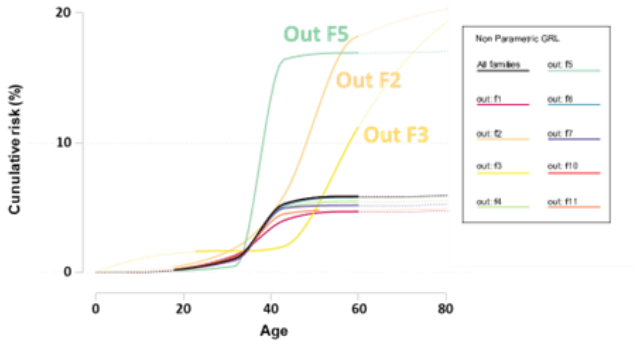


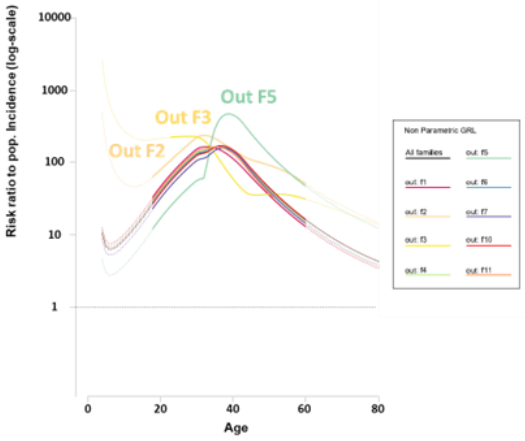




SUPLEMENTAL FIGURE S3

DGC Non Parametric GRL Model
Age nodes: Q1=32, Median=43, Q3=59



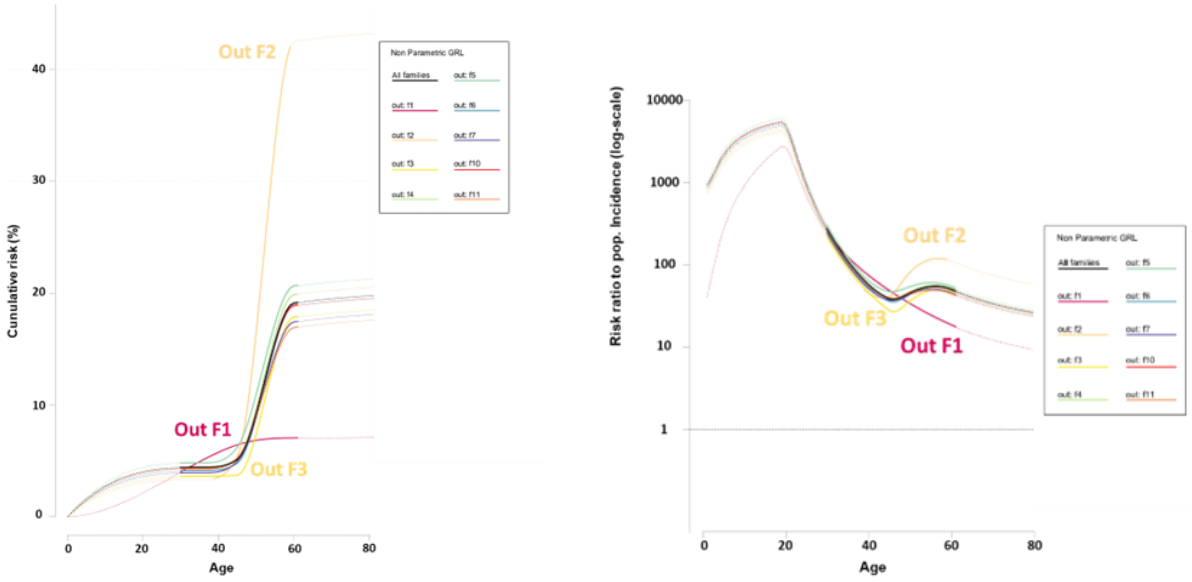


DGC Leave one-out Analysis:

Leave one out	Families considered	Cumulative Risk at 20 years	Cumulative Risk at 40 years	Cumulative Risk at 60 years	Nr of carriers	Nr. of affected carriers	Nr. of aymptomatic carriers	Nr. of affected unkown genotype cases	Nr. of asymptomatic unkown genotype	Total nr of affected cases considered	Total nr of asymptomatic cases considered	Total	GC cases %
W/out F1	F2, F3,F4,F5,F6,F7,F10,F11	0,3	3,4	4,6	78	10	68	13	67	23	135	158	15%
W/out F2	F1,F3,F4,F5,F6,F7,F10,F11	0,6	4,4	18,2	70	9	61	11	43	20	104	124	16%
W/outF3	F1,F2,F4,F5,F6,F7,F10,F11	1,5	1,8	11,1	76	10	66	13	63	23	129	152	15%
W/out F4	F1,F2,F3,F5,F6,F7,F10,F11	0,3	4,1	5,4	91	11	80	14	71	25	151	176	14%
W/out F5	F1,F2,F3,F4,F6,F7,F10,F11	0,1	13,3	16,9	76	12	64	12	56	24	120	144	17%
W/out F6	F1,F2,F3,F4,F5,F7,F10,F11	0,3	4,2	5,8	93	11	82	13	61	24	143	167	14%
W/out F7	F1,F2,F3,F4,F5,F6,F10,F11	0,2	4,2	5,1	91	10	81	14	69	24	150	174	14%
W/out F10	F1,F2,F3,F4,F5,F6,F7,F11	0,3	4,4	5,8	94	12	82	12	68	24	150	174	14%
W/out F11	F1,F2,F3,F4,F5,F6,F7,F10	0,3	3,8	4,8	91	11	80	10	70	21	150	171	12%
	All families	0.3	4.4	5.8	95	12	83	14	71	26	154	180	14%

Leave one out	Families considered	Relative Risk at 20 years	Relative Risk at 40 years	Relative Risk at 60 years	Nr of carriers	Nr. of affected carriers	Nr. of aymptomatic carriers	Nr. of affected unkown genotype cases	Nr. of asymptomatic unkown genotype cases	Total nr of affected cases	Total nr of asymptomatic cases	Total	GC cases %
W/out F1	F2, F3,F4,F5,F6,F7,F10,F11	45,5	114,1	12,9	78	10	68	13	67	23	135	158	15%
W/out F2	F1,F3,F4,F5,F6,F7,F10,F11	78	150	50,5	70	9	61	11	43	20	104	124	16%
W/outF3	F1,F2,F4,F5,F6,F7,F10,F11	216,4	59,7	31	76	10	66	13	63	23	129	152	15%
W/out F4	F1,F2,F3,F5,F6,F7,F10,F11	39,8	141,1	15,2	91	11	80	14	71	25	151	176	14%
W/out F5	F1,F2,F3,F4,F6,F7,F10,F11	16,5	454,2	47	76	12	64	12	56	24	120	144	17%
W/out F6	F1,F2,F3,F4,F5,F7,F10,F11	38,5	144,6	16	93	11	82	13	61	24	143	167	14%
W/out F7	F1,F2,F3,F4,F5,F6,F10,F11	31,5	142,3	14,3	91	10	81	14	69	24	150	174	14%
W/out F10	F1,F2,F3,F4,F5,F6,F7,F11	36,5	148,5	16,2	94	12	82	12	68	24	150	174	14%
W/out F11	F1,F2,F3,F4,F5,F6,F7,F10	41,7	131	13,3	91	11	80	10	70	21	150	171	12%
	All families	38.2	149.3	16,3	95	12	83	14	71	26	154	180	14%

LBC Non Parametric GRL Model
Age nodes: Q1=36, Median 45, Q3=60



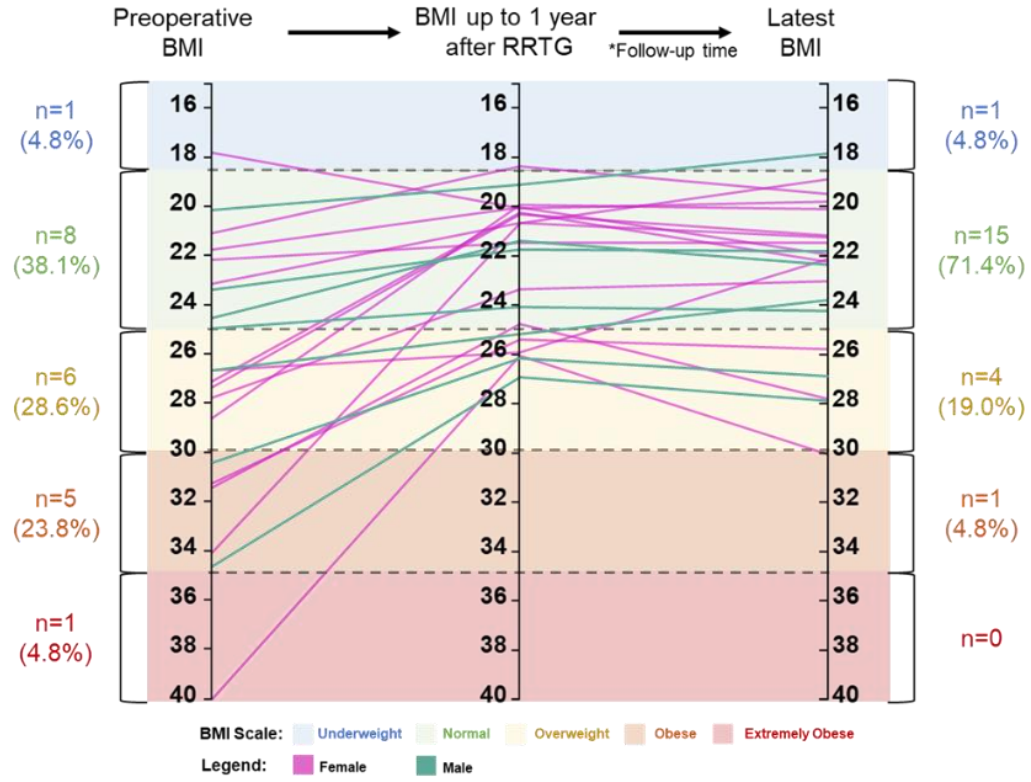
LBC Leave one-out Analysis:

Leave one out	Families considered	Cumulative Risk at 40 years	Cumulative Risk at 60 years	Nr of female carriers	Nr. of affected female carriers	Nr. of aymptomatic female carriers	Nr. of affected females w/ unkown genotype	Nr. of asymptomatic females w/ unkown genotype	Total nr of affected cases	Total nr of asymptomatic cases	Total	BC cases %
W/out F1	F2, F3,F4,F5,F6,F7,F10,F11	5,8	7,1	43	5	38	6	30	11	68	79	13,9%
W/out F2	F1,F3,F4,F5,F6,F7,F10,F11	3,5	42,5	35	4	31	3	19	7	50	57	12,3%
W/outF3	F1,F2,F4,F5,F6,F7,F10,F11	3,6	17,8	40	5	35	5	31	10	66	76	13,2%
W/out F4	F1,F2,F3,F5,F6,F7,F10,F11	4,5	19,8	48	6	42	6	32	12	74	86	14,0%
W/out F5	F1,F2,F3,F4,F6,F7,F10,F11	5	20,6	38	4	34	5	26	9	60	69	13,0%
W/out F6	F1,F2,F3,F4,F5,F7,F10,F11	4,2	18,9	49	6	43	5	28	11	71	82	13,4%
W/out F7	F1,F2,F3,F4,F5,F6,F10,F11	4	17,4	49	6	43	6	31	12	74	86	14,0%
W/out F10	F1,F2,F3,F4,F5,F6,F7,F11	4,4	18,9	50	6	44	6	29	12	73	85	14,1%
W/out F11	F1,F2,F3,F4,F5,F6,F7,F10	4,4	16,9	48	6	42	6	30	12	72	84	14,3%
	All families	4.5	19.1	50	6	44	6	31	12	75	87	14%

Leave one out	Families considered	Relative Risk at 40 years	Relative Risk at 60 years	Nr of female carriers	Nr. of affected female carriers	Nr. of aymptomatic female carriers	Nr. of affected females w/ unkown genotype	Nr. of asymptomatic females w/ unkown genotype	Total nr of affected cases	Total nr of asymptomatic cases	Total	BC cases %
W/out F1	F2, F3,F4,F5,F6,F7,F10,F11	73,4	18,4	43	5	38	6	30	11	68	79	13,9%
W/out F2	F1,F3,F4,F5,F6,F7,F10,F11	44,5	110,6	35	4	31	3	19	7	50	57	12,3%
W/outF3	F1,F2,F4,F5,F6,F7,F10,F11	46	46,4	40	5	35	5	31	10	66	76	13,2%
W/out F4	F1,F2,F3,F5,F6,F7,F10,F11	57,3	51,6	48	6	42	6	32	12	74	86	14,0%
W/out F5	F1,F2,F3,F4,F6,F7,F10,F11	63,1	53,7	38	4	34	5	26	9	60	69	13,0%
W/out F6	F1,F2,F3,F4,F5,F7,F10,F11	53	49,1	49	6	43	5	28	11	71	82	13,4%
W/out F7	F1,F2,F3,F4,F5,F6,F10,F11	51,1	45,3	49	6	43	6	31	12	74	86	14,0%
W/out F10	F1,F2,F3,F4,F5,F6,F7,F11	55,8	49,1	50	6	44	6	29	12	73	85	14,1%
W/out F11	F1,F2,F3,F4,F5,F6,F7,F10	55,4	44	48	6	42	6	30	12	72	84	14,3%
	All families	56.9	49.7	50	6	44	6	31	12	75	87	14%

SUPLEMENTAL FIGURE S4

A



Total: 21 individuals (14 Females + 7 Males)

*Follow-up time after RRTG varies between 1,5 to 10 years

Supplementary Table1: Incidence of diffuse gastric cancer (DGC) and lobular breast cancer (LBC) considered for reference population in lifetime risk estimates across different age intervals.

Registry	PT North Region					
Year	2000-2010					
Sex	Male	Female	Male	Female	Female	
Cancer	Stomach Cancer		DGC		Breast Cancer	LBC
[0;4]	0.00	0.00	0.00	0.00	0.00	0.00
[5;9]	0.00	0.00	0.00	0.00	0.00	0.00
[10;14]	0.00	0.00	0.00	0.00	0.00	0.00
[15;19]	0.20	0.00	0.06	0.00	0.3	0.03
[20;24]	0.50	1.10	0.15	0.33	1.80	0.18
[25;29]	1.30	2.40	0.39	0.72	9.90	0.99
[30;34]	4.60	4.30	1.38	1.29	27.60	2.76
[35;39]	10.60	8.60	3.18	2.58	60.30	6.03
[40;44]	22.00	14.10	6.60	4.23	118.30	11.83
[45;49]	38.10	19.30	11.43	5.79	164.60	16.46
[50;54]	56.20	23.40	16.86	7.02	172.10	17.21
[55;59]	70.00	35.10	21.00	10.53	177.50	17.75
[60;64]	104.50	47.10	31.35	14.13	192.70	19.27
[65;69]	142.80	72.70	42.84	21.81	188.20	18.82
[70;74]	199.00	101.20	59.70	30.36	178.30	17.83
[75;79]	245.70	127.20	73.71	38.16	177.70	17.77
[80;84]	266.50	139.50	79.95	41.85	169.60	16.96
85+	239.00	133.10	71.70	39.93	139.20	13.92

Supplementary Table 2: Relevant clinical Information for DGC and LBC patients per family.

	Overall (N=285)	F1 (N=41)	F2 (N=80)	F3 (N=48)	F4 (N=12)	F5 (N=54)	F6 (N=16)	F7 (N=13)	F10 (N=7)	F11 (N=14)
Gender (n) (%)										
Female	148 (51.9)	19 (46.3)	47 (58.8)	23 (47.9)	7 (58.3)	29 (53.7)	8 (50.0)	6 (46.2)	3 (42.9)	6 (42.9)
Carrier of *CDH1* c.1901C T Variant (n) (%)										
Yes	95 (33.3)	17 (41.5)	25 (31.3)	19 (39.6)	4 (33.3)	19 (35.2)	2 (12.5)	4 (30.8)	1 (14.3)	4 (28.6)
Unknown	85 (29.8)	5 (12.2)	31 (38.8)	9 (18.8)	0 (0)	17 (31.5)	11 (68.8)	2 (15.4)	5 (71.4)	5 (35.7)
Age of Last News (Years)										
Median [Q1,Q3]	44.0 [31.0, 57.0]	43.0 [30.0, 65.0]	45.5 [29.8, 57.0]	45.5 [31.0, 56.0]	62.5 [28.0, 76.5]	45.5 [34.8, 51.8]	37.5 [29.5, 44.2]	36.0 [30.0, 40.0]	45.0 [38.0, 80.0]	50.5 [38.5, 56.5]
Gastric Cancer Status (n) (%)										
Advanced DGC	10 (3.5)	2 (4.9)	2 (2.5)	2 (4.2)	1 (8.3)	0 (0)	1 (6.3)	1 (7.7)	0 (0)	1 (7.1)
Early DGC	26 (9.1)	6 (14.6)	4 (5.0)	10 (20.8)	2 (16.7)	3 (5.6)	0 (0)	1 (7.7)	0 (0)	0 (0)
GC Unknown Histology	16 (5.6)	1 (2.4)	4 (5.0)	1 (2.1)	0 (0)	2 (3.7)	1 (6.3)	1 (7.7)	2 (28.6)	4 (28.6)
Asymptomatic	233 (81.8)	32 (78.0)	70 (87.5)	35 (72.9)	9 (75.0)	49 (90.7)	14 (87.5)	10 (76.9)	5 (71.4)	9 (64.3)
Age of Advanced DGC Diagnosis (Years)										
Median [Q1,Q3]	35.0 [30.8, 41.0]	35.0 [35.0, 35.0]	42.0 [37.5, 46.5]	28.0 [23.0, 33.0]	30.0 [30.0, 30.0]	-	42.0 [42.0, 42.0]	23.0 [23.0, 23.0]	-	43.0 [43.0, 43.0]
Age of Early DGC Diagnosis (Years)										
Median [Q1,Q3]	39.5 [27.8, 49.5]	37.5 [31.2, 54.2]	45.5 [38.5, 54.2]	35.0 [20.5, 42.8]	32.5 [26.2, 38.8]	44.0 [35.5, 49.0]	-	35.0 [35.0, 35.0]	-	-
Age of Unknown Histology Gastric Cancer Diagnosis (Years)										
Median [Q1,Q3]	40.0 [34.5, 46.5]	27.0 [27.0, 27.0]	40.0 [38.2, 48.2]	36.0 [36.0, 36.0]	-	41.5 [36.2, 46.2]	44.0 [44.0, 44.0]	27.0 [27.0, 27.0]	40.0 [37.5, 42.5]	57.5 [50.2, 60.0]
Prophylactic Gastrectomy Performed (n) (%)	24 (8.4)	7 (17.1)	5 (6.2)	8 (16.7)	2 (16.7)	1 (1.9)	0 (0)	1 (7.7)	0 (0)	0 (0)
Age at Prophylactic Gastrectomy (Years)										
Median [Q1,Q3]	40 [31.0, 51.0]	40.0 [35.5, 59.0]	45.5 [40.0, 57.0]	35.0 [18.8, 42.2]	32.5 [26.2, 38.8]	27.0 [27.0, 27.0]	-	35.0 [35.0, 35.0]	-	-
Breast Cancer Status (n) (%)	**147 females***	**19 females**	**46 females***	**23 females**	**7 females**	**29 females**	**8 females**	**6 females**	**3 females**	**6 females**
LBC	6 (4.1)	1 (5.3)	2 (4.3)	1 (4.3)	0 (0)	2 (6.9)	0 (0)	0 (0)	0 (0)	0 (0)
BC Unknown Histology	6 (4.1)	0 (0)	3 (6.5)	1 (4.3)	0 (0)	1 (3.4)	1 (12.5)	0 (0)	0 (0)	0 (0)
Asymptomatic	135 (91.8)	18 (94.7)	41 (89.1)	21 (91.3)	7 (100)	26 (89.7)	7 (87.5)	6 (100)	3 (100)	6 (100)
Age of LBC Diagnosis (Years)										
Median [Q1,Q3]	48.5 [48.0, 56.5]	59.0 [59.0, 59.0]	55.0 [52.0, 58.0]	48.0 [48.0, 48.0]	-	45.0 [43.5, 46.5]	-	-	-	-
Age of Unknown Histology Breast Cancer Diagnosis (Years)										
Median [Q1,Q3]	47.5 [45.0, 50.8]	-	45.0 [37.5, 48.5]	50.0 [50.0, 50.0]	-	51.0 [51.0, 51.0]	45.0 [45.0, 45.0]	-	-	-
Prophylactic Mastectomy Performed (n) (%)	7 (4.8)	2 (10.5)	3 (6.5)	1 (4.3)	1 (14.3)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)
Age at Prophylactic Mastectomy (Years)										
Median [Min, Max]	43.0 [39.5, 45.5]	39.5 [38.2, 40.8]	43.0 [37.5, 43.5]	63.0 [63.0, 63.0]	47.0 [47.0, 47.0]	-	-	-	-	-
DGCLBC Criteria #	-	Yes	Yes	Yes	Yes	Yes	Yes	Yes	?	Yes

*One female had age of last news for DGC penetrance estimations, but not for LBC

According to Blair *et al.* 2020

Supplementary Table 3: Phenotypic data and procedures on carriers, unknown genotype and noncarrier groups

	Carriers (N=95)	Unknown Genotype (N=85)	Non-carriers (N=105)	Overall (N=285)
Gender				
Female	51 (53.7%)	38 (44.7%)	59 (56.2%)	148 (51.9%)
Family ID				
F1	17 (17.9%)	5 (5.9%)	19 (18.1%)	41 (14.4%)
F10	1 (1.1%)	5 (5.9%)	1 (1.0%)	7 (2.5%)
F11	4 (4.2%)	5 (5.9%)	5 (4.8%)	14 (4.9%)
F2	25 (26.3%)	31 (36.5%)	24 (22.9%)	80 (28.1%)
F3	19 (20.0%)	9 (10.6%)	20 (19.0%)	48 (16.8%)
F4	4 (4.2%)	0 (0%)	8 (7.6%)	12 (4.2%)
F5	19 (20.0%)	17 (20.0%)	18 (17.1%)	54 (18.9%)
F6	2 (2.1%)	11 (12.9%)	3 (2.9%)	16 (5.6%)
F7	4 (4.2%)	2 (2.4%)	7 (6.7%)	13 (4.6%)
Age of Last News				
Median [Q1,Q3]	42.0 [33.0, 54.0]	45.0 [31.0, 70.0]	46.0 [30.0, 56.0]	44.0 [31.0, 57.0]
Gastric Cancer Status				
Asymptomatic	57 (60.0%)	71 (83.5%)	105 (100%)	233 (81.8%)
GC Unknown Histology	2 (2.1%)	14 (16.5%)	0 (0%)	16 (5.6%)
Advanced DGC	10 (10.5%)	0 (0%)	0 (0%)	10 (3.5%)
Early DGC	26 (27.4%)	0 (0%)	0 (0%)	26 (9.1%)
Age of Advanced DGC Diagnosis				
Median [Q1,Q3]	35.0 [30.8, 41.0]	NA [NA, NA]	NA [NA, NA]	35.0 [30.8, 41.0]
Age of Early DGC Diagnosis				
Median [Q1,Q3]	39.5 [27.8, 49.5]	NA [NA, NA]	NA [NA, NA]	39.5 [27.8, 49.5]
Age of Unknown Histology Gastric Cancer Diagnosis				
Median [Q1,Q3]	36.0 [31.5, 40.5]	40.0 [35.2, 49.5]	NA [NA, NA]	40.0 [34.5, 46.5]
Prophylactic Gastrectomy Performed				
No	68 (71.6%)	85 (100%)	105 (100%)	258 (90.5%)
Yes	27 (28.4%)	0 (0%)	0 (0%)	27 (9.5%)
Age at Prophylactic Gastrectomy				
Median [Q1,Q3]	40.0 [31.0, 51.0]	NA [NA, NA]	NA [NA, NA]	40.0 [31.0, 51.0]

Supplementary Table 4: GRL risk and KM estimates for DGC and LBC at different age points.

GRL:

Cancer	DGC		LBC	
Risk	Cumulative Risk	Relative Risk	Cumulative Risk	Relative Risk
Age=20	0.3	38.2	4.0	4971.6
Age=25	0.5	74.6	4.3	952.3
Age=30	0.9	128.0	4.4	270.5
Age=35	2.0	163.6	4.4	110.6
Age=40	4.4	149.3	4.5	56.9
Age=45	5.4	82.7	5.1	38.1
Age=50	5.7	44.7	9.2	44.8
Age=55	5.8	26.0	15.6	53.9
Age=60	5.8	16.3	19.1	49.7
Age=65	5.8	10.8	19.3	40.0
Age=70	5.9	7.6	19.5	33.6
Age=75	5.9	5.6	19.6	29.2
Age=80	6.0	4.3	19.7	26.1

KM:

Cancer	Group	Cumulative risk at 20 (95% CI)	Cumulative risk at 40 (95% CI)	Cumulative risk at 60 (95% CI)
DGC	Carriers	0.01 (0.00, 0.03)	0.10 (0.03, 0.17)	0.19 (0.08, 0.29)
	Unknown Genotype	0.00 (0.00, 0.00)	0.13 (0.04, 0.21)	0.27 (0.13, 0.38)
LBC	Carriers	0.00 (0.00, 0.00)	0.03 (0.00, 0.08)	0.24 (0.00, 0.43)
	Unknown Genotype	0.00 (0.00, 0.00)	0.03 (0.00, 0.09)	0.27 (0.06, 0.44)

Supplementary Table 5: Short-and long-term events associated with prophylactic total gastrectomy.

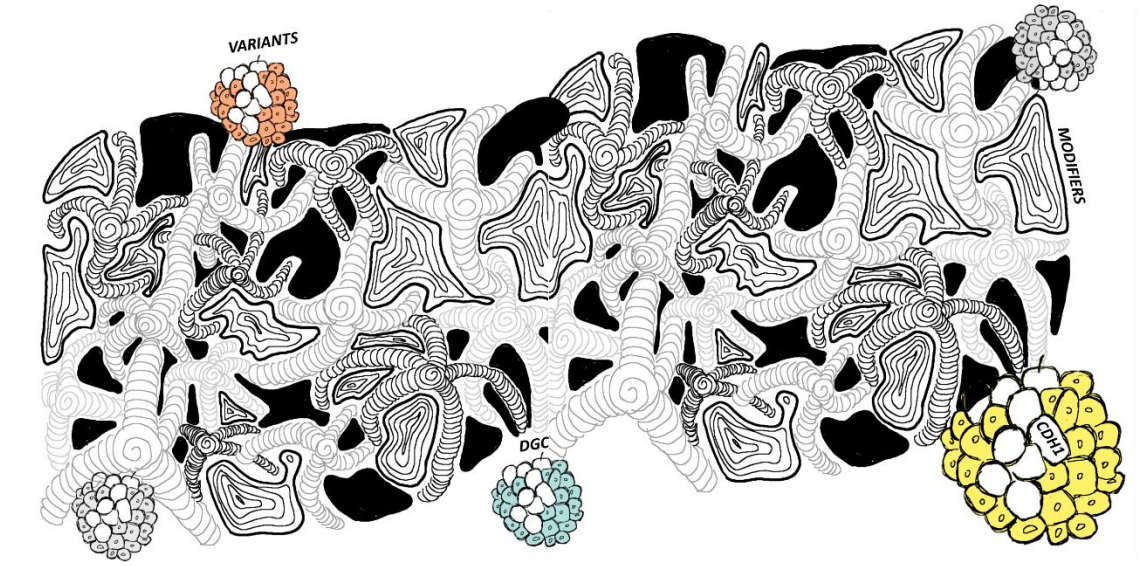
Cases submitted to prophylatic gastrectomy	Overall (n=25)	Asymptomatic Male (n=9)	Asymptomatic Female (n=12)	Females Diagnosed with LBC (n=2)
Age at Surgery (Years)				
Median [Q1,Q3]	39.0 [30.5, 48.0]	35.0 [30.0, 62.0]	38.0 [32.2, 43.5]	57.5 [54.2, 60.8]
Sucessful Recovery From Surgery (n) (%)				
Yes	21 (91.3)	8 (88.9)	11 (91.7)	2 (100)
Unknown	1 (4.3)	1 (11.1)	0 (0)	0 (0)
Reoperation During Post-Operative Recovery (n) (%)				
Yes	2 (8.7)	1 (11.1)	1 (8.3)	0 (0)
Unknown	1 (4.3)	1 (11.1)	0 (0)	0 (0)
Surgical Site Infection (n) (%)				
Yes	1 (4.3)	1 (11.1)	0 (0)	0 (0)
Unknown	1 (4.3)	1 (11.1)	0 (0)	0 (0)
Morbidity Association with Procedure (n) (%)				
Yes	2 (8.7)	1 (11.1)	1 (8.3)	0 (0)
Unknown	1 (4.3)	1 (11.1)	0 (0)	0 (0)
Overall Follow-up Time (Years)				
Median [Q1,Q3]	7.0 [5.5, 8.0]	7.00 [6.0, 8.0]	7.0 [5.8, 8.0]	6.0 [5.0, 7.0]
Clinical Symptoms and Complications Reported Within Six to 12 Months After Surgery (n) (%)				
Asthenia	4 (17.4)	1 (11.1)	2 (16.7)	1 (50.0)
Diarrhea	2 (8.7)	1 (11.1)	1 (8.3)	0 (0)
Nausea	7 (30.4)	2 (22.2)	4 (33.3)	1 (50.0)
Return to Normal Activities (n) (%)				
Work	14 (60.9)	6 (66.7)	8 (66.7)	0 (0)
Student	4 (17.4)	2 (22.2)	2 (16.7)	0 (0)
Retired	2 (8.7)	0 (0)	1 (8.3)	1 (50.0)
Unemployed	2 (8.7)	1 (11.1)	1 (8.3)	0 (0)
LBC patient	1 (4.3)	0 (0)	0 (0)	1 (50.0)
Unknown	5 (21.7)	3 (33.3)	2 (16.7)	0 (0)
Clinical Symptoms and Complications Reported More Than 12 Months After Surgery (n) (%)				
Hernia	2 (8.7)	1 (11.1)*	1 (8.3)	0 (0)
Nausea				
Surfeit	2 (8.7)	1 (11.1)	1 (8.3)	0 (0)
Obstipation	3 (13.0)	0 (0)	2 (16.7)	1 (50.0)
Dumping Syndrome	4 (17.4)	2 (22.2)	2 (16.7)	0 (0)
Muscle Cramps	4 (17.4)	3 (33.3)	1 (8.3)	0 (0)
Heartburn	2 (8.7)	1 (11.1)	1 (8.3)	0 (0)
Esophagitis	0 (0)	0 (0)	0 (0)	0 (0)
Reflux	1 (4.3)	0 (0)	1 (8.3)	0 (0)
Dysphagia	2 (8.7)	0 (0)	2 (16.7)	0 (0)
Micronutrients Deficiencies (n) (%)				
Low Iron	9 (39.1)	2 (22.2)	7 (58.3)	0 (0)
Low Vitamin B12	7 (30.4)	3 (33.3)	4 (33.3)	0 (0)
Low.Vitamin.D	10 (43.5)	5 (55.6)	5 (41.7)	0 (0)
Low.Magnesium	7 (30.4)	2 (22.2)	5 (41.7)	0 (0)
Total proteins	1 (4.3)	1 (11.1)	0 (0)	0 (0)

*Reported patient had two distinct hernias at two different timepoints

3.2. Original Paper 6, Chapter III

Combined whole-genome and transcriptome profiling of Early Onset Cancer Patients and Asymptomatic Carriers of a *CDH1* founder variant

R. Barbosa-Matos, J. Fonseca, M. Ferreira, L. Garrido, A. Gojanovich, M. Yang, A. Pedro, A. Dias, S. Doria, J. Pelaez, N. Martins, A. Brito, A. Moutinho, S. Seixas, J. Freixo, R. Oliveira, R. Quental, P. Louro, S. Fernandes, A. Grangeia, M. Baptista, I. Gullo, F. Carneiro, H. Pinheiro, S. Castedo, G. Mostovslasky, D. Ferreira, C.Oliveira



Author statement:

Rita Barbosa-Matos contributed to the conceptualization, investigation, data collection, methodology, figure design and writing (original draft)

In this study we explore the genetic and transcriptomic landscape of the families carrying the *CDH1* c.1901 C>T variant. From a combination of predicted deleterious germline alterations setting a genomic profile of disease susceptibility and by exploring the transcriptome of different tissues, we try to disclose the genetic modifiers behind *CDH1* incomplete penetrance. We identified at least 4 candidate genetic modifiers supported by other studies.

Abstract

CDH1 germline variants cause early-onset and deadly Hereditary Diffuse Gastric Cancer (HDGC). However, disease penetrance is incomplete, and it is currently impossible to distinguish carriers that will develop disease from those that will remain forever asymptomatic. The *CDH1* c.1901 C>T variant is the most recurrent in HDGC families across Europe, present in a population sharing the same geographical area for centuries. Our study aimed to identify genomic and transcriptomic molecular signatures that distinguish affected from asymptomatic carriers at the germline level.

Besides having whole genome sequencing (WGS) data from 21 individuals, we collected whole blood and extracted peripheral blood mononuclear cells (PBMCs) for RNA-seq studies, and to establish the induced-Pluripotent Stem Cells (iPSCs) from six *CDH1*-associated Diffuse Gastric Lobular Breast Cancer Syndrome carriers (3 young affected and 3 old asymptomatic carriers). Moreover, we had access to normal gastric mucosa formalin-fixed paraffin-embedded blocks from 2 young affected and 3 asymptomatic carriers. RNA was extracted from all samples and submitted to transcriptome analysis. Differential gene expression analysis identified at least 2 long non-coding RNAs and 2 protein-coding genes, shared between different tissue types. Additional candidate genes and pathways highlight a predisposing expression profile exclusive from young-affected carriers. Moreover, we identified a set of CNVs and SNVs that create a specific genomic signature characteristic of young-affected and old-asymptomatic carriers. Our study highlights the first candidate genetic modifiers of disease expressivity in *CDH1*-mutation carriers.

Introduction

Diffuse Gastric Lobular Breast Cancer Syndrome (DGLBCS) caused by CDH1 germline pathogenic, or likely-pathogenic variants (PV/LPV) predisposes to early onset diffuse gastric (DGC) and lobular breast cancer (LBC) (Blair et al., 2020). Intriguingly, disease penetrance is incomplete, and most families display individual phenotype heterogeneity, reflecting great variability in clinical behaviour or age of onset (Barbosa-Matos et al., 2021; Gullo et al., 2018). Since many carriers of a CDH1 mutation may never develop disease, it has been hypothesized that genetic variation and/or environmental interactions may influence lifetime cancer risk.

The existence of founder populations, characterized by individuals sharing a common ancestor and carrying the same disease-causing genetic variant represent the ideal study model to identify putative genetic modifiers (Neuhausen, 2000). In the North region of Portugal, multiple families carry the pathogenic CDH1 c.1901 C>T founder variant. This simplified context is a major advantage to identify HDGC genetic modifiers, since genetic heterogeneity is lower, a common haplotype was already identified in 7/11 (64.6%) families. Additionally, all individuals reside in a small municipality (70 km^2), reducing the impact of environmental factors.

Multiple combinations of single nucleotide variants (SNVs) and copy number variants (CNVs) alterations, converging in common molecular pathways, can lead to DGLBCS phenotypes. Millions of SNPs and CNVs detected at the germline level challenges interpretation of disease risk modifiers (Garcia-Pelaez et al., 2021; Lim et al., 2023). These modifiers may manifest as either rare, deleterious variants or common, low penetrance variants.

Therefore, we aim to compare for the first time the genome and transcriptome between CDH1 variant carriers affected by DGC and asymptomatic carriers within families harbouring the Northern Portuguese founder CDH1 c.1901C>T variant. Using next generation sequencing technologies, namely whole-genome sequencing and RNA-seq, our goal is to identify a set of candidate modifier genes and molecular pathways involved in DGC development.

Since the molecular impact of a DNA variant in different tissues can be highly variable, we decided to explore multiple contexts to compare gene expression in affected and non-affected carriers. For this, we will use several sources of material including whole blood, human peripheral blood mononuclear cells (PBMCs), induced-pluripotent stem cells (iPSCs) and formalin-Fixed Paraffin-Embedded (FFPE) tissue non-cancerous tissue from gastrectomies. Our goal is to identify gene

expression profiles that not only differentiate carriers affected by the DGC from those who remain asymptomatic, but also uncover common pathways or differentially expressed genes present across various tissue sources.

Material and Methods

Cases Selection for RNA extraction from whole-blood and PBMCs

Three young-affected carriers ant three old-asymptomatic carriers gave consent to donate blood for one PAXgene Blood RNA tube and one BD CPT™ glass tube. The former, was used for direct RNA extraction from whole blood and the latter tube, were aliquoted into four vials of isolated Peripheral blood mononuclear cells (PBMCs). One vial was used for RNA extraction, a second for iPSCs reprogramming and the remaining ones were stored in the biobank. PBMCs were collected and expanded as previously mentioned (C. A. Sommer et al., 2012).

Cases Selection for establishing an iPSCs library:

PBMCs for the abovementioned cases were used for iPSCs reprogramming using the Sendai virus integration-free method. These cells were reprogrammed to in the absence of serum or a xenogeneic feeder layer (Park & Mostoslavsky, 2018). Expression of TRA1-81, a marker of pluripotency, was assessed by immunofluorescence staining (Supplemental Figure 1).

iPSCs and PBMCs RNA extraction and sequencing

RNA was isolated from all samples using the RNeasy Kit (QIAGEN cat. no. 74014), either immediately after dissociation from tissue culture vessels, or after storage at −20 °C in RNAlater (ThermoFisher cat. no. 7020) following the manufacturer's **instructions.**

Whole blood, PBMCs, and iPSCs RNA-seq

Duplicated RNA samples were collected from patient-derived iPSCs, ribosomal RNA (rRNA) was removed, and samples were sequenced with Illumina Sequencing PE150 (12 G raw data per sample ~80M reads per sample). The same approach was used for the six PBMCs and whole blood samples. For the whole blood samples, first a globin-zero gold treatment and rRNA removal kit were applied.

Whole transcriptome sequencing analysis

Quality control (QC) of the raw fastq files was performed using FastQC, unique reads were mapped to the human genome (GRCh38) reference sequence using STAR. RSeQC was used to count and annotated the read counts using information from USCS GRCh38 (Genecode v41) and ensembl (v107 and v108). Differential expression analysis was performed using Deseq2 (v1.23.0) R package and canonical transcripts were selected considering |log2fold-change|≥1 and adjusted p-value<0.05. The heatmap of DEGs were generated using the ggplot2 plotting module (version 3.4.4) and PCA was performed using the prcomp module from stats package (version 3.6.2), both using normalized reads counts with vst function from Deseq2 (1.36.0) R package. Gene Ontology (GO) terms (biological processes and molecular functions) associated with each set of DEGs were explored using Enrichr(Chen et al., 2013).

Cases Selection for FFPE non-tumoral tissue extraction:

Histological samples from normal gastric mucosa of DGC patients followed and treated at the University Hospital Center of São João (CHUSJ) and asymptomatic carriers submitted to risk reducing gastrectomy (RRG) were selected for RNA extraction and subsequent RNA-sequencing analysis. In total, 5 individuals were selected and carriers of the CDH1 c.1901C>T variant.

FFPEs RNA extraction, sequencing, and analysis

MagMAX™ FFPE DNA/RNA Ultra Kit was used for RNA extraction. The Ion AmpliSeq™ Transcriptome Human Gene Expression Kit was used for sequencing. Data analysis was performed with Affymetrix Transcriptome Analysis Console (TAC), using e-Bayes ANOVA algorithm. Differentially expressed genes (DEGs) were defined as having a log2 Fold-Change below -1 or above 1 and a p-value <0.05. No p-value adjustment was introduced in the analysis, due to the low sample size and no revelatory output. Gene Ontology (GO) terms, biological processes, and molecular functions associated with each set of DEGs were explored using Enrichr(Chen et al., 2013).

Cases Selection for WGS analysis:

All individuals consented to donate blood for DNA extraction. In total, 21 individuals: 4 carriers with DGC ≤ 35 years old, 4 asymptomatic carriers older than 65 years old, 6 young asymptomatic carriers (≤ 35 years) and 7 asymptomatic

carriers between 39 and 64 years old were selected for WGS. Carriers detailed information is summarized in Supplemental Table 1. All samples were sequenced using a 30x coverage NovaSeq6000 150PE (2x150bp) 90Gb/Sample with TruSeq WGS Library and analysed with an in-house pipeline that proved consistent prioritization of true-positive and disease causing SNVs and CNVs. Briefly, after genomic DNA 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 and CNVs)

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). Identified SNVs were first filtered according to quality metrics, read depth (>=8) and genotype quality (>=20), and annotated with VEP (v.107). SNVs were prioritized called by Deep Variant only or if called by Deep Variant and HaplotypeCaller (high quality >=200). CNVs calling was performed with LUMPY and DELL, considering ≥ 90% overlap between both. We discarded structural variants ≥1Mb in length, inner-chromosomal events, and variants lacking ´PASS´ flag were discarded.

SNVs & CNVs variants prioritization

After SNV calling and annotation, we defined a strategy to select impactful variants based in specific criteria. First, we selected "HIGH IMPACT" variants, according to variant-effect predictor annotations (https://m.ensembl.org/info/genome/variation/prediction/predicted_data.html) independently of their Minor Allele Frequency (MAF) and genomic region. We then identified exclusive variants from each group, following the criterion that a variant must be present in at least three cases of the affected carriers' group and in no more than one old asymptomatic carrier, or vice-versa.

A similar approach was performed for CNVs prioritization, we considered as high impact variants, the copy-number deletions present in coding regions and identified exclusive variants from each group following the same selection criterion (≥3 young-affected carriers & ≤ 1 old asymptomatic carrier or ≥3 old asymptomatic carriers & ≤ 1 affected carrier).

Outputs from SNVs and CNVs selection for each group were merged to find candidate genes that establish specific genomic profiles for young-affected

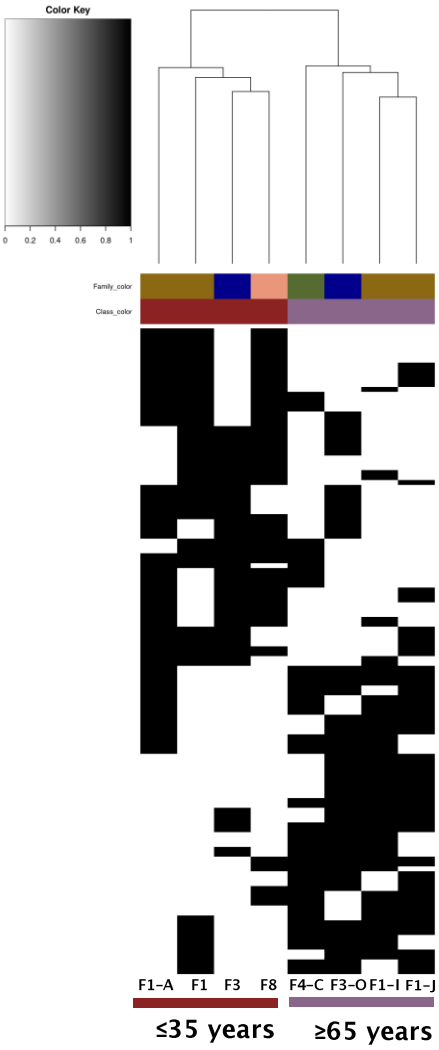
carriers and old asymptomatic carriers. Gene Ontology (GO) terms, biological processes and molecular functions associated with each set of "HIGH IMPACT" genes were explored using EnrichR.

Results

CDH1 c.1901 founder population germline landscape

We compared the germline landscape of carriers affected by DGC at very early age (≤ 35 years) with old asymptomatic carriers (≥ 65 years) who had not undergone any risk-reducing surgery or developed any cancer. Our goal was to unveil other genetic factors that might influence the predisposition to or protection against DGC. To achieve this, we focused on germline SNVs and CNVs that were present in at least 75% (3 out of 4) of cases within one group and found in only one case of the other group. A total of 52 SNVs across 37 genes characterized the affected carriers, with 16 SNVs being exclusive to this group. Meanwhile, 47 SNVs in 40 distinct genes were identified in the old asymptomatic carriers with 9 SNVs being exclusive to this group. Considering coding deleterious CNVs, in the young-affected carriers' group we identified 9 CNVs (with 2 of them being exclusive) encompassing 17 genes in total, while 15 CNVs (8 exclusive) across 25 genes were identified in the old asymptomatic group (**Figure 1A and 1B**). Only two genes, *MUC19* and *ANKRD36,* were shared between young and asymptomatic carriers with distinct high impact variants (**Figure 1B**). Interestingly, the *MUC3A* gene in chromosome 7 presents the highest number of high impact variants, presenting 8 different SNVs all in exon 2. From those, at least 3 SNVs are totally exclusive from affected carriers (**Figure 1B**). In total, 61 distinctive variants established a young-affected carrier genomic profile and other 62 variants defined an asymptomatic carrier genomic profile.

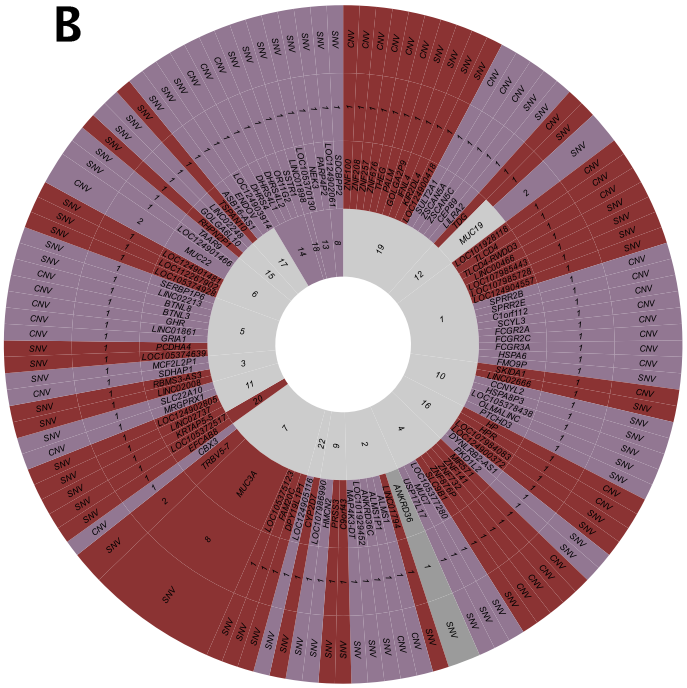
A



Legend:
- Affected Carriers (DGC)
- Old Asymptomatic Carriers

- F1 Family
- F3 Family
- F4 Family
- F8 Family

B



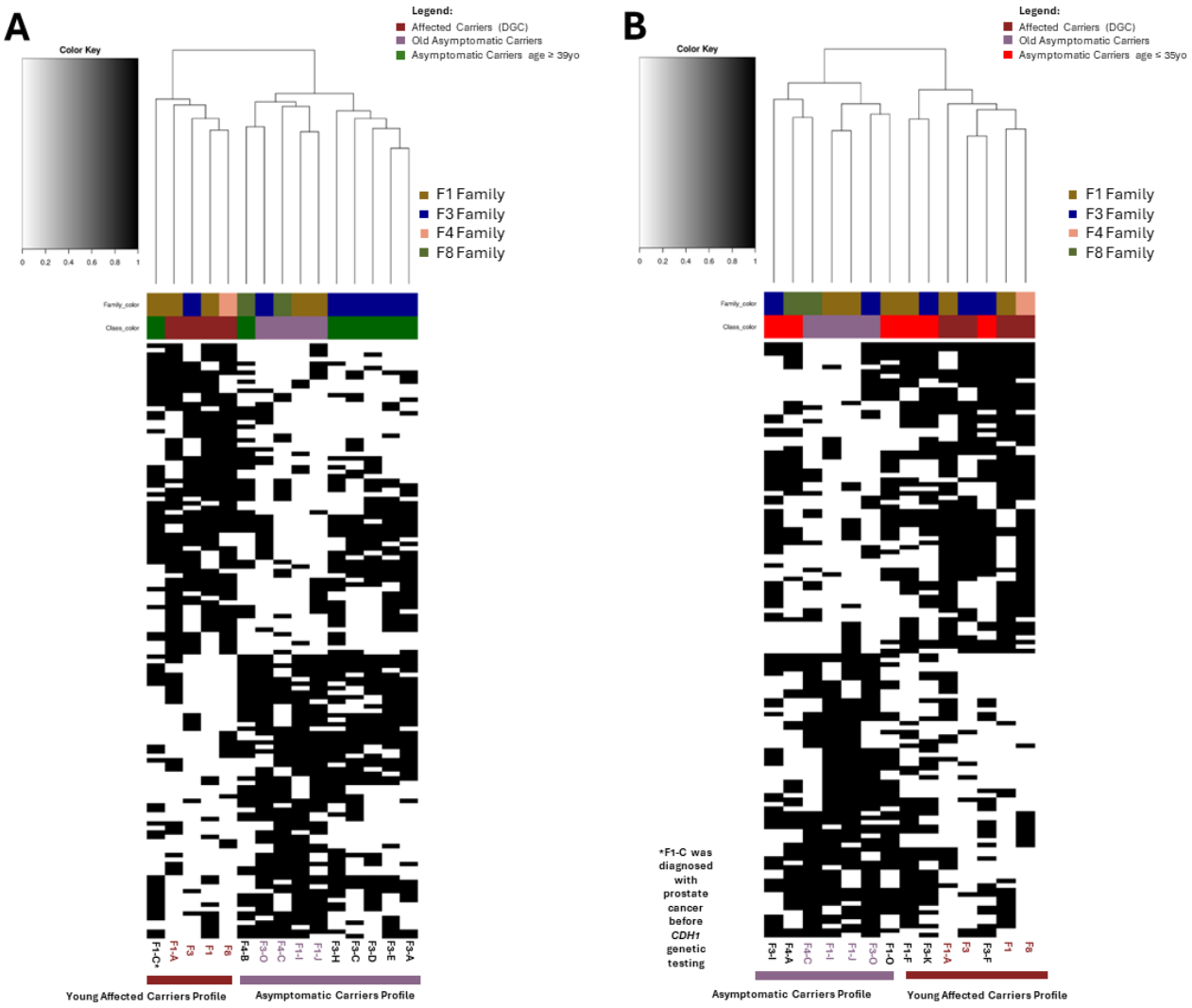
Legend:
- Affected Carriers Variant
- Old Asymptomatic Carrier Variant
- Shared Gene w/ Affected Carriers Variant & Old Asymptomatic Carrier Variant

Figure 1. WGS analysis of Young-affected and Old Asymptomatic groups. A) Heatmap of Young-affected Carriers diagnosed with DGC before 35 years old and Old Asymptomatic Carriers older than 65 years old for the total 123 SNVs and CNVs identified. **B)** Sunburst diagram representing the chromosome, genes, and number of variants (SNV or CNV) found by each group (affected carriers vs asymptomatic carriers).

Genomic Signatures: Comparative analysis across age groups of asymptomatic carriers

Given that the median age of DGC diagnosis among our *CDH1* c.1901 C>T carriers is 35 years old, and no cases of GC have been reported after 60 years of age across all 11 families (Paper 5), we focused on asymptomatic carriers who had risk-reducing gastrectomy either before (n=6, median [Q1, Q3], 23 years [19, 31]) or after (n=7, 45 years [42, 61]) turning 35. Our aim was to determine whether their genomic profiles more closely resemble with those of young-affected carriers or older asymptomatic carriers. (**Supplemental Table 1**, **Figure 2**). All the 7 asymptomatic carriers with RRG after 35 years shared an old asymptomatic carrier genomic signature except for F1-C, which developed pancreatic cancer at 57 years of age (**Figure 2A**). On the other hand, four of the young asymptomatic carriers aged 14,19, and 35 clustered in the young-affected profile (**Figure 2B**). Notably, case F3-F which performed risk-reducing gastrectomy at 14 years of age, shared several high impact variants with the proband of that family that perished from DGC at 18 years old. The remaining with three cases (F4-A, F1-O and F3-I) aged 20, 26 and 31, respectively, share the old asymptomatic genomic profile (**Figure 2B**).

Our rationale implied that both High Impact SNVs and coding CNVs deletions will impact gene regulation, supporting the identification of a specific expression profile for each group. Therefore, we explored in EnrichR the list of genes from both groups and their associated Gene Ontology (GO) terms. Despite both affected and asymptomatic groups had a similar total number of variants, their combined predicted impact is distinct. In fact, at more biological processes and molecular functions are expected to be affected by the young-affected genomic profile, being the top categories related to inflammation, gene expression regulation and cell death (**Figure 2C**). In old asymptomatic profile, associations with metabolic and enzymatic processes could be established.



C

Biological Process	Go Term	P-value	Adjusted p-value	Odds Ratio	Combined score
Young Affected Group	Regulation Of Nucleic Acid-Templated Transcription (GO:1903506)	0.0001200	0.01152	9.05	81.75
Young Affected Group	Acute Inflammatory Response (GO:0002526)	0.0009417	0.04520	50.36	350.87
Young Affected Group	Regulation Of Gene Expression (GO:0010468)	0.002966	0.08298	4.21	24.48
Young Affected Group	Positive Regulation Of Cell Death (GO:0010942)	0.003457	0.08298	25.15	142.52
Young Affected Group	Regulation Of Cell Death (GO:0010941)	0.004782	0.08329	21.17	113.10
Young Affected Group	Monoterpenoid Metabolic Process (GO:0016098)	0.008720	0.08329	146.77	696.01
Young Affected Group	Coumarin Metabolic Process (GO:0009804)	0.008720	0.08329	146.77	696.01
Young Affected Group	Regulation Of Natural Killer Cell Cytokine Production (GO:0002727)	0.01046	0.08329	117.41	535.47
Young Affected Group	Positive Regulation Of Natural Killer Cell Cytokine Production (GO:0002729)	0.01046	0.08329	117.41	535.47
Young Affected Group	Terpenoid Metabolic Process (GO:0006721)	0.01219	0.08329	97.84	431.21
Old Asymptomatic Group	Retinal Metabolic Process (GO:0042574)	0.000009064	0.001559	93.89	1090.13
Old Asymptomatic Group	Retinoid Metabolic Process (GO:0001523)	0.0001530	0.01316	32.95	289.44

Molecular Function	Go Term	P-value	Adjusted p-value	Odds Ratio	Combined score
Young Affected Group	Serine-Type Endopeptidase Activity (GO:0004252)	0.001348	0.02604	15.25	100.78
Young Affected Group	Serine-Type Peptidase Activity (GO:0008236)	0.001942	0.02604	13.37	83.50
Young Affected Group	Cis-Regulatory Region Sequence-Specific DNA Binding (GO:0000987)	0.002562	0.02604	4.32	25.81
Young Affected Group	RNA Polymerase II Cis-Regulatory Region Sequence-Specific DNA Binding (GO:0000978)	0.002893	0.02604	4.23	24.71
Young Affected Group	RNA Polymerase II Transcription Regulatory Region Sequence-Specific DNA Binding (GO:0000977)	0.004708	0.03390	3.85	20.62
Young Affected Group	Mismatched DNA Binding (GO:0030983)	0.008720	0.05232	146.77	696.01
Young Affected Group	Sodium Ion Binding (GO:0031402)	0.01046	0.05377	117.41	535.47
Young Affected Group	DNA N-glycosylase Activity (GO:0019104)	0.01564	0.06710	73.37	305.06
Young Affected Group	Chloride Ion Binding (GO:0031404)	0.01737	0.06710	65.22	264.33
Young Affected Group	Endopeptidase Activity (GO:0004175)	0.02052	0.06710	5.54	21.55
Old Asymptomatic Group	Carbonyl Reductase (NADPH) Activity (GO:0004090)	9.198e-7	0.00005886	244.20	3394.17
Old Asymptomatic Group	Oxidoreductase Activity, Acting On The CH-OH Group Of Donors, NAD Or NADP As Acceptor (GO:0016616)	0.001939	0.06203	13.21	82.53

Figure 2. WGS analysis of Young-affected and Old Asymptomatic groups. A) Heatmap of Young-affected Carriers diagnosed with DGC before 35 years old and Old Asymptomatic Carriers older than 65 years old for the total 123 SNVs and CNVs identified. **B)** Sunburst diagram representing the chromosome, genes, and number of variants (SNV or CNV) found by each group (affected carriers vs asymptomatic carriers).

Whole Blood, PBMCs, and iPSCs Transcriptomic Profiles

To infer the gene expression profiles distinguishing young-affected carriers from old asymptomatic carriers, first we had to consider the possible expression bias effect of age-related genes. Since iPSCs are generated from somatic tissues, but during reprogramming somatic cell identity is lost, aging is reverted. Therefore, we excluded age associated genes particularly in whole blood and PBMCs analysis,

First, we identified a list of 1578 genes, from different studies, reporting differentially expressed genes (DEGs) between young and old healthy cohorts in blood and PBMCs derived samples to be crossed with our final list of DEGs (**Supplemental Table 2**). Then, we excluded two DEGs overlapping with the age-related gene (*LRNN3* and *E2F2*) list identified in the whole blood analysis, whereas no overlap was found in the PBMCs analysis. Given the significance of long non-coding RNAs (lncRNAs) in our analysis, comprising between 23-33% of the identified DEGs, we also searched in the literature for lncRNAs associated with aging (**Supplemental Table 2**). No association was found with our differentially expressed lncRNAs. In total, RNA was extracted from blood samples, PBMCs and iPSCs in three old asymptomatic carriers, two carriers affected by DGC before 35 years and one asymptomatic carrier submitted to risk-reducing gastrectomy before 35 years that presented early lesions of DGC. All selected cases have WGS data available. By differential gene expression analysis between young and old asymptomatic carriers a total of 22 DEGs (30 transcripts) were identified in whole blood samples, 2 DEGs (3 transcripts) were found in PBMCs, and 21 DEGs (27 transcripts) were reported in iPSCs analysis (**Table 1; Figure 3A**).

No DEGs were commonly altered across all tissues. However, two transcripts were shared between whole blood and iPSCs, and another two were shared between PBMCs and iPSCs (**Figure 3B**). No CNV amplifications, deletions or SNVs were detected for these regions in the WGS of carriers. One of the shared differentially expressed lncRNA is in a highly regulatory region with several marks of chromatin regulation, enhancers, and methylation sites in the normal stomach. On the other

hand, the second lncRNA only overlaps a large CpG island with less regulatory features. Moreover, there was no overlap between the DEGs and the SNVs or CNVs found in the old asymptomatic genomic profile, while in the young-affected genomic profile one gene had a CNV detected in one allele present in 3 young-affected cases (also detected in one old asymptomatic case). This gene was upregulated in iPSCs from the same individuals.

Regarding the GO TERMS analysis, we could not identify overlapping biological processes or molecular functions across distinct tissue material (**Supplemental Table 3**).

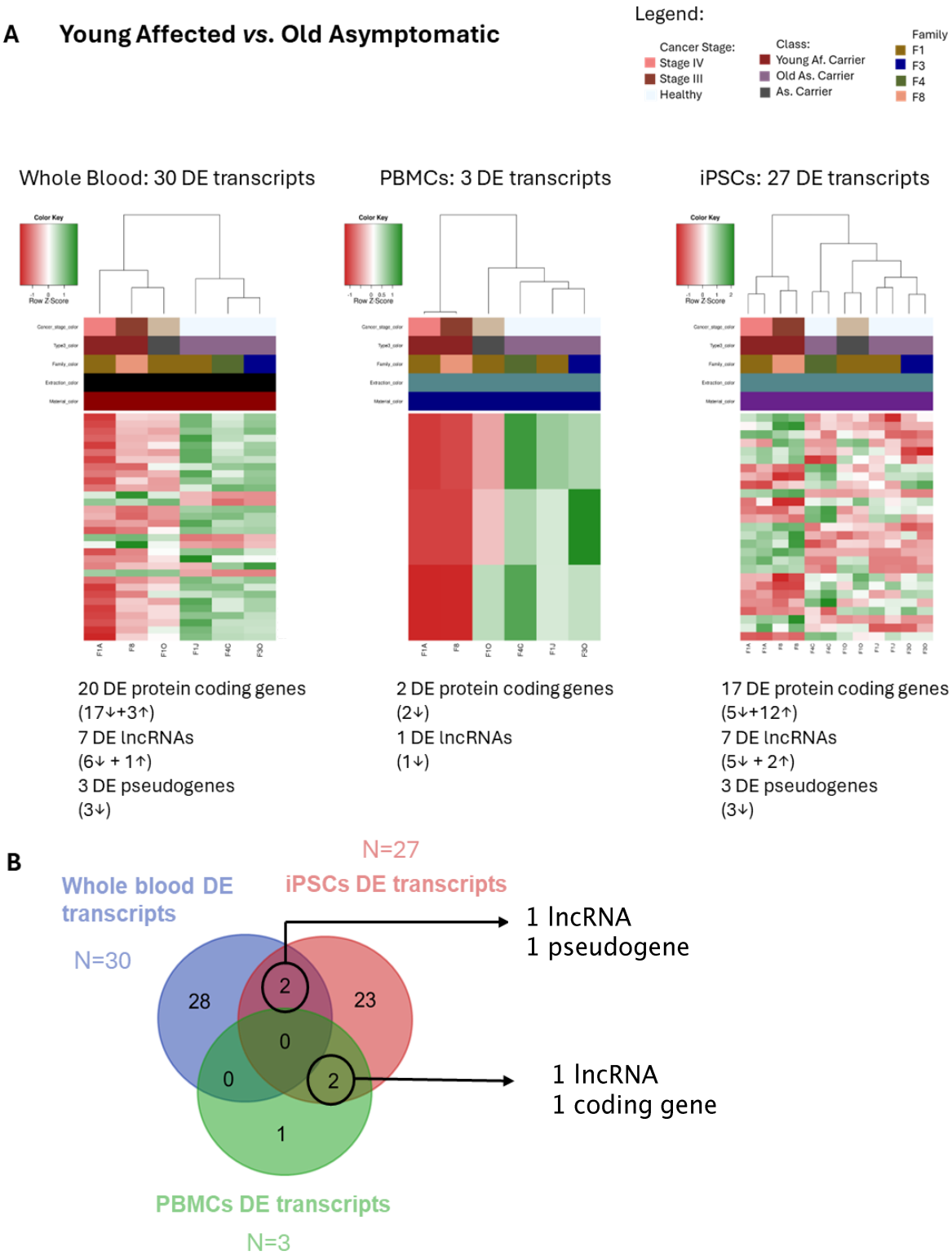


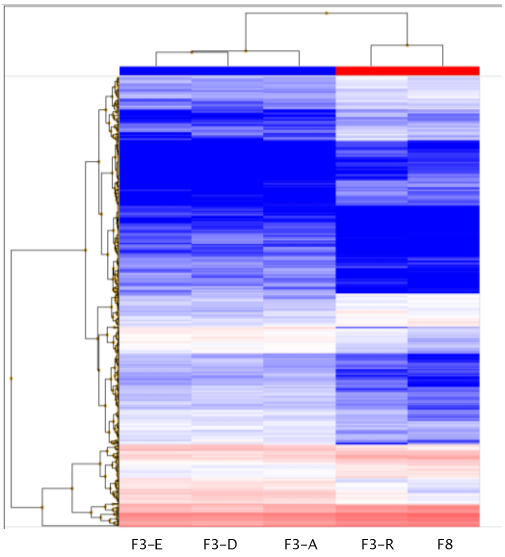
Figure 3. Analysis of gene differential expression between young-affected carriers and old asymptomatic carriers for Whole Blood, PBMCs and iPSCs. A) Differentially expressed transcripts and total genes up and downregulated for the three source materials. **B)** Venn diagram showing the interception of differentially

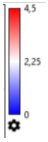
expressed genes between source materials and list of the four shared transcripts between iPSCs and blood or iPSCs and PBMCs.

Normal gastric mucosa transcriptomic analysis: Affected vs Asymptomatic carriers

We also wanted to compare the transcriptional differences between the gastric mucosa of young-affected cases and old asymptomatic individuals. However, for these individuals to be classified as old asymptomatic they did not remove their stomach as a risk reduction measure. Ultimately, there is no tissue available from these cases. Nevertheless, we had access to gastric FFPE blocks from asymptomatic carriers submitted to gastrectomy, and we selected individuals older than 35 years old that presented a genomic profile resembling old carriers identified in **Figure 2**. We extracted adjacent normal gastric mucosa from young-affected individuals (cases F3-R and F8) and compared it with normal gastric mucosa from asymptomatic carriers aged between 42 and 61 years old. In total, 648 DEGs were identified, being 303 of them upregulated in young-affected cases (**Figure 4A**). At least 50 DEGs were only expressed in young-affected carriers. From the 345 downregulated genes, 34 are only detected in asymptomatic samples. Interestingly, we identify clusters of genes that characterize young-affected carriers expression profile, these genes are all either upregulated such as human olfactory receptor genes and small nucleolar RNA genes or downregulated such as the FAM family of genes, solute carrier genes and zinc-finger proteins (**Figure 4B**). One coding gene is downregulated on FFPE tissue from affected carriers and it is also present in the young-affected carriers' genomic profile, its CNVs are absent from all old asymptomatic carriers (**Figure 4C**). From the 15 MUC genes detected by the Ion AmpliSeq™ Transcriptome, none reach statistically significance.

A

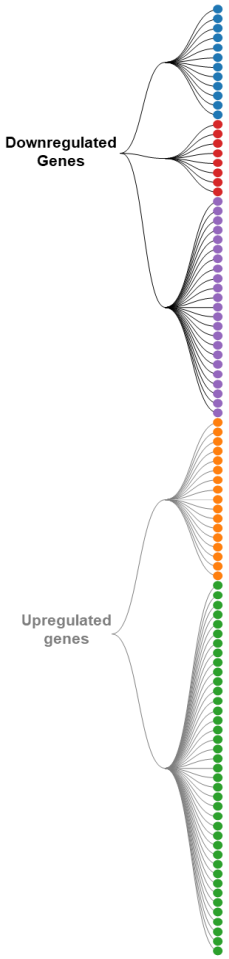




Class:
- Asymptomatic Carrier
- Young Affected Carrier

Filter criteria:
- Fold Change: ≤ -2 or ≥ +2
- P-val: < 0,05

B



FAM gene family n=12

Solute-carrier genes n=8

Zinc-finger protein genes N=23

Human olfactory receptor genes n=17

Small nucleolar RNA genes n=39

C

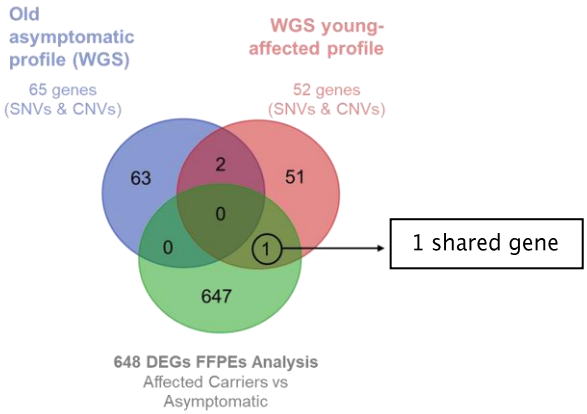


Figure 4. Analysis of normal gastric mucosa gene differential expression between young-affected carriers and old asymptomatic carriers. A) Differentially expressed genes represented in a heatmap. **B)** List of gene up and downregulated that belong to specific gene families/clusters establishing a young-affected expression profile. **C)** Venn diagram showing the interception of differentially expressed genes in FFPE and list of the genes present in the genomic profiles from young-affected and asymptomatic carriers based on SNVs and CNVs analysis

We also explored the associated GO TERMS but given the high number of genes we divided up and downregulated genes and analysed them separately. This allowed us to get a clearer picture on the different pathways impacted by the DEGs found in the normal mucosa analysis. Indeed, only one GO TERM was shared between up and down regulated genes namely "Regulation Of Glucose Transmembrane Transport (GO:0010827)", pinpointing the importance of this rate-limiting step for glucose metabolism in cancer development.

Integrative analysis of candidate genes and GO terms: insights across tissues

Finally, we decided to explore the overlap between all GO TERMS (biological processes and molecular functions) identified in the different analysis for each tissue, also including the GO TERMS associated with the young-affected and old asymptomatic genomic profiles (**Table 2**). Two GO TERMS associated with the upregulated genes from the FFPEs study were shared with analyses from other datasets. Specifically, the term "Epidermal Growth Factor Receptor Binding (GO:0005154)" term was found in the iPSCs analysis, while "Serine-Type Endopeptidase Activity (GO:0004252)" was shared with the young-affected genomic profile. Additionally, three other GO TERMS present in the downregulated genes from the FFPE analysis, namely "Cis-Regulatory Region Sequence-Specific DNA Binding (GO:0000987)", "RNA Polymerase II Cis-Regulatory Region Sequence-Specific DNA Binding (GO:0000978)" and "RNA Polymerase II Transcription Regulatory Region Sequence-Specific DNA Binding (GO:0000977)", were also shared with the young-affected genomic profile.

Table 2. GO TERMS (Biological Process and Molecular Functions) associated with the different DNA and RNA expression profiles analysis in whole blood, PBMCs, iPSCS, FFPEs and WGS data.

Blood (n=19 genes)	Blood BP	P-value	Adjusted p-value	Odds Ratio	Combined score
1	Regulation Of Megakaryocyte Differentiation (GO:0045652)	0.0001448	0.01477	138.16	1221.34
2	Positive Regulation Of Myeloid Cell Differentiation (GO:0045639)	0.0005284	0.02695	69.02	520.81
3	Positive Regulation Of Granulocyte Differentiation (GO:0030854)	0.004741	0.06434	277.46	1484.80
4	Mitochondrial Protein Catabolic Process (GO:0035694)	0.004741	0.06434	277.46	1484.80
5	Regulation Of Axon Guidance (GO:1902667)	0.005687	0.06434	221.96	1147.41
6	Tubulin Complex Assembly (GO:0007021)	0.005687	0.06434	221.96	1147.41
7	Negative Regulation Of T Cell Mediated Cytotoxicity (GO:0001915)	0.006632	0.06434	184.95	927.70
8	Mitochondrial Outer Membrane Permeabilization (GO:0097345)	0.006632	0.06434	184.95	927.70
9	Immune Response-Inhibiting Cell Surface Receptor Signaling Pathway (GO:0002767)	0.007576	0.06434	158.52	774.04
10	Positive Regulation Of Megakaryocyte Differentiation (GO:0045654)	0.007576	0.06434	158.52	774.04

Blood (n=19 genes)	Blood MF	P-value	Adjusted p-value	Odds Ratio	Combined score
1	[illegible]	0.004741	0.07823	277.46	1484.80
2	Myosin Light Chain Binding (GO:0032027)	0.004741	0.07823	277.46	1484.80
3	GTPase Activator Activity (GO:0005096)	0.01683	0.1111	11.13	45.46
4	Phosphatase Inhibitor Activity (GO:0019212)	0.02070	0.1111	52.80	204.75
5	Ubiquitin-Like Protein Transferase Activity (GO:0019787)	0.02144	0.1111	9.76	37.50
6	Ferrous Iron Binding (GO:0008198)	0.02349	0.1111	46.20	173.28
7	Protein Homodimerization Activity (GO:0042803)	0.02356	0.1111	5.50	20.61
8	MHC Protein Binding (GO:0042287)	0.03162	0.1302	33.58	115.78
9	Alpha-Tubulin Binding (GO:0043014)	0.03550	0.1302	29.95	99.96
10	Ubiquitin Ligase-Substrate Adaptor Activity (GO:1990756)	0.04100	0.1330	25.76	82.28

iPSCs (n=16 genes)	iPSCs BP	P-value	Adjusted p-value	Odds Ratio	Combined score
1	Regulation Of Transport (GO:0051049)	0.0006299	0.05720	63.30	466.55
2	Protection From Natural Killer Cell Mediated Cytotoxicity (GO:0042270)	0.004243	0.05720	312.17	1705.22
3	ERBB2-EGFR Signaling Pathway (GO:0038134)	0.004243	0.05720	312.17	1705.22
4	Galactolipid Biosynthetic Process (GO:0019375)	0.004243	0.05720	312.17	1705.22
5	Galactosylceramide Biosynthetic Process (GO:0006682)	0.004243	0.05720	312.17	1705.22
6	Plasma Membrane Raft Organization (GO:0044857)	0.004243	0.05720	312.17	1705.22
7	Glycosyl Compound Biosynthetic Process (GO:1901659)	0.004243	0.05720	312.17	1705.22
8	Vascular Endothelial Growth Factor Receptor-2 Signaling Pathway (GO:0036324)	0.004243	0.05720	312.17	1705.22
9	Wybutosine Biosynthetic Process (GO:0031591)	0.004243	0.05720	312.17	1705.22
10	Wybutosine Metabolic Process (GO:0031590)	0.004243	0.05720	312.17	1705.22

iPSCs (n=16 genes)	iPSCS MF	P-value	Adjusted p-value	Odds Ratio	Combined score
1	Vascular Endothelial Growth Factor Receptor Activity (GO:0005021)	0.005090	0.08267	249.72	1318.68
2	Cysteine-Type Endopeptidase Inhibitor Activity Involved In Apoptotic Process (GO:0043027)	0.01603	0.08267	69.32	286.51
3	Metal Cation:Proton Antiporter Activity (GO:0051139)	0.01771	0.08267	62.38	251.64
4	Zinc Ion Transmembrane Transporter Activity (GO:0005385)	0.01938	0.08267	56.71	223.63
5	Epidermal Growth Factor Receptor Binding (GO:0005154)	0.02105	0.08267	51.98	200.68
6	pre-mRNA Binding (GO:0036002)	0.02521	0.08267	43.00	158.28
7	Transition Metal Ion Transmembrane Transporter Activity (GO:0046915)	0.02769	0.08267	38.97	139.76
8	Hsp90 Protein Binding (GO:0051879)	0.02935	0.08267	36.67	129.39
9	Transmembrane Receptor Protein Kinase Activity (GO:0019199)	0.03100	0.08267	34.63	120.30
10	Transmembrane Receptor Protein Tyrosine Kinase Activity (GO:0004714)	0.04168	0.09627	25.43	80.80

PBMCs (n=2 genes)	PBMCs BP	P-value	Adjusted p-value	Odds Ratio	Combined score
1	Ammonium Homeostasis (GO:0097272)	0.0004999	0.004499	4998.50	37993.75
2	Ammonium Transmembrane Transport (GO:0072488)	0.0009998	0.004499	2221.00	15342.67
3	Cytoplasmic Translation (GO:0002181)	0.009279	0.02732	216.37	1012.62
4	Peptide Biosynthetic Process (GO:0043043)	0.01574	0.02732	126.38	524.67
5	Nitrogen Compound Transport (GO:0071705)	0.01623	0.02732	122.44	504.55
6	Macromolecule Biosynthetic Process (GO:0009059)	0.01822	0.02732	108.88	436.11
7	Translation (GO:0006412)	0.02326	0.02938	84.83	319.03
8	Inorganic Cation Transmembrane Transport (GO:0098662)	0.02820	0.02938	69.66	248.60
9	Gene Expression (GO:0010467)	0.02938	0.02938	66.79	235.59

PBMCs (n=2 genes)	PBMCs MF	P-value	Adjusted p-value	Odds Ratio	Combined score
1	Ammonium Transmembrane Transporter Activity (GO:0008519)	0.0009998	0.003999	2221.00	15342.67
2	Inorganic Cation Transmembrane Transporter Activity (GO:0022890)	0.003497	0.006924	587.18	3320.99
3	Channel Activity (GO:0015267)	0.005193	0.006924	391.12	2057.43
4	RNA Binding (GO:0003723)	0.1361	0.1361	13.18	26.29

FFPE Downregulated genes (n=345)	FFPE down BP	P-value	Adjusted p-value	Odds Ratio	Combined score
1	Regulation Of Glucose Transmembrane Transport (GO:0010827)	0.0001734	0.2848	17.72	153.49
2	Chemical Synaptic Transmission (GO:0007268)	0.0009697	0.5693	2.92	20.27
3	Modulation Of Chemical Synaptic Transmission (GO:0050804)	0.001335	0.5693	4.03	26.70
4	Neurotransmitter Transport (GO:0006836)	0.001984	0.5693	6.14	38.18
5	Signal Release From Synapse (GO:0099643)	0.004045	0.5693	6.77	37.30
6	Positive Regulation Of Icosanoid Secretion (GO:0032305)	0.004251	0.5693	28.65	156.42
7	Positive Regulation Of Mast Cell Activation (GO:0033005)	0.004251	0.5693	28.65	156.42
8	Phosphatidylglycerol Metabolic Process (GO:0046471)	0.004665	0.5693	10.13	54.39
9	Regulation Of Small Molecule Metabolic Process (GO:0062012)	0.004875	0.5693	6.39	34.03
10	Cellular Modified Amino Acid Biosynthetic Process (GO:0042398)	0.005884	0.5693	22.92	117.68

FFPE Downregulated genes (n=345)	FFPE down MF	P-value	Adjusted p-value	Odds Ratio	Combined score
1	Adrenergic Receptor Binding (GO:0031690)	0.001985	0.3001	14.36	89.35
2	N,N-dimethylaniline Monooxygenase Activity (GO:0004499)	0.002867	0.3001	38.20	223.62
3	RNA Polymerase II Cis-Regulatory Region Sequence-Specific DNA Binding (GO:0000978)	0.003733	0.3001	1.74	9.73
4	RNA Polymerase II Transcription Regulatory Region Sequence-Specific DNA Binding (GO:0000977)	0.004305	0.3001	1.69	9.23
5	Cis-Regulatory Region Sequence-Specific DNA Binding (GO:0000987)	0.004963	0.3001	1.72	9.13
6	G Protein-Coupled Glutamate Receptor Binding (GO:0035256)	0.005884	0.3001	22.92	117.68
7	Tubulin-Glutamic Acid Ligase Activity (GO:0070740)	0.007756	0.3390	19.10	92.79
8	G Protein-Coupled Receptor Binding (GO:0001664)	0.01091	0.4082	3.04	13.73
9	Ligand-Gated Monoatomic Ion Channel Activity Involved In Regulation Of Presynaptic Membrane Potential	0.01218	0.4082	14.32	63.12
10	Gamma-Tubulin Binding (GO:0043015)	0.01334	0.4082	6.62	28.59

FFPE Upregulated gens	FFPE up BP	P-value	Adjusted p-value	Odds Ratio	Combined score
1	Regulation Of Glucose Transmembrane Transport (GO:0010827)	0.002000	0.5542	14.06	87.37
2	Positive Regulation Of Lipid Storage (GO:0010884)	0.003242	0.5542	11.58	66.35
3	Positive Regulation Of Apoptotic Cell Clearance (GO:2000427)	0.004569	0.5542	26.17	141.01
4	Regulation Of Apoptotic Cell Clearance (GO:2000425)	0.004569	0.5542	26.17	141.01
5	Mitotic Nuclear Division (GO:0140014)	0.004722	0.5542	6.41	34.35
6	Cobalamin Transport (GO:0015889)	0.006031	0.5542	21.81	111.45
7	Ventricular Cardiac Muscle Tissue Development (GO:0003229)	0.006186	0.5542	8.94	45.48
8	Vesicle Fusion With Vacuole (GO:0051469)	0.007677	0.5542	18.69	91.01
9	Multivesicular Body-Lysosome Fusion (GO:0061763)	0.007677	0.5542	18.69	91.01
10	Glycolytic Process (GO:0006096)	0.009401	0.5542	7.57	35.31

FFPE Upregulated gens	FFPE up MF	P-value	Adjusted p-value	Odds Ratio	Combined score
1	Olfactory Receptor Activity (GO:0004984)	0.001327	0.3106	2.69	17.84
2	Hydrolase Activity, Acting On Carbon-Nitrogen (But Not Peptide) Bonds, In Linear Amidines (GO:0016813)	0.007677	0.6106	18.69	91.01
3	Estrogen Response Element Binding (GO:0034056)	0.009500	0.6106	16.35	76.15
4	Sodium-Independent Organic Anion Transmembrane Transporter Activity (GO:0015347)	0.02109	0.6106	10.06	38.82
5	Phosphatidic Acid Binding (GO:0070300)	0.03632	0.6106	7.26	24.08
6	Serine-Type Endopeptidase Activity (GO:0004252)	0.04175	0.6106	2.74	8.69
7	Epidermal Growth Factor Receptor Binding (GO:0005154)	0.05460	0.6106	5.68	16.53
8	Metalloexopeptidase Activity (GO:0008235)	0.05536	0.6106	3.64	10.53
9	Growth Factor Receptor Binding (GO:0070851)	0.05819	0.6106	2.85	8.11
10	Secondary Active Transmembrane Transporter Activity (GO:0015291)	0.06268	0.6106	3.45	9.54

Young Affected Genomic Profile n=54	Young Affected Genomic Profile BP	P-value	Adjusted p-value	Odds Ratio	Combined score
1	Regulation Of Nucleic Acid-Templated Transcription (GO:1903506)	0.0001200	0.01152	9.05	81.75
2	Acute Inflammatory Response (GO:0002526)	0.0009417	0.04520	50.36	350.87
3	Regulation Of Gene Expression (GO:0010468)	0.002966	0.08298	4.21	24.48
4	Positive Regulation Of Cell Death (GO:0010942)	0.003457	0.08298	25.15	142.52
5	Regulation Of Cell Death (GO:0010941)	0.004782	0.08329	21.17	113.10
6	Monoterpenoid Metabolic Process (GO:0016098)	0.008720	0.08329	146.77	696.01
7	Coumarin Metabolic Process (GO:0009804)	0.008720	0.08329	146.77	696.01
8	Regulation Of Natural Killer Cell Cytokine Production (GO:0002727)	0.01046	0.08329	117.41	535.47
9	Positive Regulation Of Natural Killer Cell Cytokine Production (GO:0002729)	0.01046	0.08329	117.41	535.47
10	Terpenoid Metabolic Process (GO:0006721)	0.01219	0.08329	97.84	431.21

Young Affected Genomic Profile n=54	Young Affected Genomic Profile MF	P-value	Adjusted p-value	Odds Ratio	Combined score
1	Serine-Type Endopeptidase Activity (GO:0004252)	0.001348	0.02604	15.25	100.78
2	Serine-Type Peptidase Activity (GO:0008236)	0.001942	0.02604	13.37	83.50
3	Cis-Regulatory Region Sequence-Specific DNA Binding (GO:0000987)	0.002562	0.02604	4.32	25.81
4	RNA Polymerase II Cis-Regulatory Region Sequence-Specific DNA Binding (GO:0000978)	0.002893	0.02604	4.23	24.71
5	RNA Polymerase II Transcription Regulatory Region Sequence-Specific DNA Binding (GO:0000977)	0.004708	0.03390	3.85	20.62
6	Mismatched DNA Binding (GO:0030983)	0.008720	0.05232	146.77	696.01
7	Sodium Ion Binding (GO:0031402)	0.01046	0.05377	117.41	535.47
8	DNA N-glycosylase Activity (GO:0019104)	0.01564	0.06710	73.37	305.06
9	Chloride Ion Binding (GO:0031404)	0.01737	0.06710	65.22	264.33
10	Endopeptidase Activity (GO:0004175)	0.02052	0.06710	5.54	21.55

Old Asymptomatic Genomic Profile n=50	Old Asymptomatic Genomic Profile BP	P-value	Adjusted p-value	Odds Ratio	Combined score
1	Retinal Metabolic Process (GO:0042574)	0.000009064	0.001559	93.89	1090.13
2	Retinoid Metabolic Process (GO:0001523)	0.0001530	0.01316	32.95	289.44

Old Asymptomatic Genomic Profile n=50	Old Asymptomatic Genomic Profile MF	P-value	Adjusted p-value	Odds Ratio	Combined score
1	Carbonyl Reductase (NADPH) Activity (GO:0004090)	9.198e-7	0.00005886	244.20	3394.17
2	Oxidoreductase Activity, Acting On The CH-OH Group Of Donors, NAD Or NADP As Acceptor (GO:0016616)	0.001939	0.06203	13.21	82.53

Finally, after generating various datasets and analysing the unique characteristics of each tissue samples, we compiled a table to summarize the genetic intersections observed among the different expression profiles (**Table 3**). We consider as more relevant candidate modifier, genes shared between different analysis with the same directional regulation (highlighted in gray).

Table 3. Summary of candidate modifiers genes

	Tissue Sample				WGS Analysis
Gene type	Whole Blood	PBMCs	iPSCs	Stomach FFPEs	Whole Blood
LncRNA	Upregulated	-	Upregulated	Not detected	No SNV/CNV detected
LncRNA	-	Downregulated	Downregulated	Not detected	No SNV/CNV detected
Protein coding gene	Upregulated	-	-	-	CNV in Younfg Profile
Protein coding gene	-	Downregulated	Downregulated	-	No SNV/CNV detected
Protein coding gene	-	-	Upregulated	Upregulated	-
Pseudogene name	Downregulated	-	Downregulated	Not detected	No SNV/CNV detected
LncRNA	Downregulated	-	-	Not detected	No SNV/CNV detected
LncRNA	Downregulated	-	-	Not detected	No SNV/CNV detected
LncRNA	Downregulated	-	-	Not detected	No SNV/CNV detected
LncRNA	Downregulated	-	-	Not detected	No SNV/CNV detected
LncRNA	Downregulated	-	-	Not detected	No SNV/CNV detected
LncRNA	Downregulated	-	-	Not detected	No SNV/CNV detected
LncRNA	-	-	Upregulated	Not detected	CNV in 2 young affected
LncRNA	-	-	Upregulated	Not detected	-
LncRNA	-	-	Downregulated	Not detected	-
LncRNA	-	-	Downregulated	Not detected	-
LncRNA	-	-	Downregulated	Not detected	CNV in 1 young affected
Protein coding gene	Downregulated	-	-	-	-
Protein coding gene	Downregulated	-	-	-	-
Protein coding gene	Downregulated	-	-	-	-
Protein coding gene	Downregulated	-	-	-	-
Protein coding gene	Downregulated	-	-	-	-
Protein coding gene	Downregulated	-	-	-	-
Protein coding gene	Downregulated	-	-	-	-
Protein coding gene	Downregulated	-	-	-	-
Protein coding gene	Downregulated	-	-	-	-
Protein coding gene	Downregulated	-	-	-	-
Protein coding gene	Downregulated	-	-	-	-
Protein coding gene	Downregulated	-	-	-	-
Protein coding gene	Downregulated	-	-	Not detected	-
Protein coding gene	Downregulated	-	-	Not detected	-
Protein coding gene	Downregulated	-	-	-	-
Protein coding gene	Downregulated	-	-	-	-
Protein coding gene	Upregulated	-	-	-	-
Protein coding gene	Upregulated	-	-	Not detected	-
Protein coding gene	-	Downregulated	-	-	-
Protein coding gene	-	-	Upregulated	-	-
Protein coding gene	-	-	Upregulated	-	-
Protein coding gene	-	-	Upregulated	-	-
Protein coding gene	-	-	Upregulated	-	-
Protein coding gene	-	-	Upregulated	Not detected	-
Protein coding gene	-	-	Upregulated	-	-
Protein coding gene	-	-	Upregulated	-	-
Protein coding gene	-	-	Upregulated	-	-
Protein coding gene	-	-	Upregulated	-	-
Protein coding gene	-	-	Upregulated	-	-
Protein coding gene	-	-	Upregulated	-	-
Protein coding gene	-	-	Downregulated	-	-
Protein coding gene	-	-	Downregulated	-	-
Protein coding gene	-	-	Downregulated	Not detected	-
Protein coding gene	-	-	Downregulated	-	-
Pseudogene name	Downregulated	-	-	Not detected	-
Pseudogene name	Downregulated	-	-	Not detected	-
Pseudogene name	-	-	Downregulated	Not detected	-
Pseudogene name	-	-	Downregulated	Not detected	-

Discussion

Our search for genetic modifiers was initially focused on germline alterations that could either predispose or protect from DGC development. We tried to combine transcriptomic with genomic data to unveil putative molecular pathways that could be associated with young affected or old asymptomatic carriers. One limitation of this work is the limited number of samples, which is typically a constraint in rare genetic disease studies. A higher number of young-affected and asymptomatic carriers would allow us to perform a more robust differential gene expression analysis and to further detail the genomic profiles herein identified. Nevertheless, by comparing DEGs and transcripts between blood, PBMCs, iPSCs and normal stomach we identified 5 candidate modifers: 2 lncRNAs, 2 protein coding genes and 1 pseudogene.

Regarding our lncRNAs, one of them is a very interesting candidate, since its highly expressed in blood and distinguishes young-affected individuals from old asymptomatic carriers, a strong candidate as a biomarker for risk stratification. This lncRNA is antisense to a coding gene in a region of the genome highly enriched for regulatory elements of gene expression (Talotta, 2023). This locus was associated as a prognostic signature of gastric distant metastasis, reinforcing the relevance of exploring the role of this lncRNA (Hu et al., 2019). Less is known about the second lncRNA, but since it is located on chromosome 16 (as the *CDH1* gene), we can hypothesize that its expression might be associated with a protective effect in *cis*.

One coding gene and one pseudogene are interesting hits as well. In a study looking for circulating biomarkers for early-stage gastric cancer, the pseudogene mRNA was one of the hits (Lin et al., 2018). While the coding gene has also been associated with carcinogenesis. Its downregulation was associated with worse overall survival in invasive bladder cancer and predicted to be a biomarker for disease recurrence (Piao et al., 2022). It is also described that its genetic variants can impar ribosome function and cellular metabolism (Dlamini et al., 2021)

Finally, the shared gene between FFPEs and iPSCs analysis is involved in many developmental and pathological mechanisms. In normal conditions, its expression is important for growth and repair of the gastrointestinal tract (Kosone et al., 2006). In breast cancer, its overexpression promotes cell migration and angiogenesis (Wolczyk et al., 2016). In 1995, it was already reported as a useful biomarker in the serum of gastric cancer patients for prognosis and follow-up

(Moskal et al., 1995). Additionally, it has been reported that SNVs n the 3′ untranslated region (UTR) of this gene might predispose to gastric cancer and it has been proposed as a candidate for cleft lip/palate and cleft palate (Machida et al., 1999; Wu et al., 2014)

The comparison of normal gastric mucosa between young-affected cases and asymptomatic individuals also highlighted candidate genes and molecular pathways relevant for DGC pathogenesis. Indeed, the cluster of genes differentially expressed that we identified, have been described as strong candidate biomarkers in GC or validated as biomarkers in other cancers. The solute carrier family genes were already proposed as prognostic biomarkers in GC, the olfactory receptor (OR) gene transcripts abundance have been validated as prostate cancer biomarkers and their upregulation is described among invasive breast cancer patients (L. Liu et al., 2020; Masjedi et al., 2019). Additionally, a risk profile assessment based on the expression of FAM family genes validated novel biomarkers that help to stratify endometrial cancer patients (Chi et al., 2023). Regarding small nucleolar RNAs, it is known that many are upregulated in GC tumors compared with normal tissue, and their clinical significance for GC staging and metastic status has been described (Pourghasem et al., 2023; Wang et al., 2019).

Finally, we cannot ignore the zinc-finger protein (ZFP) genes identified, not only by being frequently deleted at the germline level in young-affected carriers than in old asymptomatic cases, but also due to their frequent appearance in the differential expression analysis. ZFPs may act as tumor suppressors or oncogenes in GC development, involved in many steps of disease including inflammation, apoptosis, invasion, and metastasis (S. Liu et al., 2023).

Conclusion

The improvement of a genetic susceptibility expression profile that allows to distinguish at-risk carriers from those likely to remain asymptomatic is urgent. In this work, we were able to find interesting targets for future validation that likely increase the risk of DGC development and may be tested as stratification biomarkers. To our knowledge this is the first report of iPSCs lines derived from multiple members of a founder population with a *CDH1* pathogenic variant. This new model will be a powerful tool for functionally validating the herein identified genetic modifiers.

References:

Barbosa-Matos, R., Silva, R. L., 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., … Oliveira, C. (2021). The cdh1 c.1901c>t variant: A founder variant in the portuguese population with severe impact in mrna splicing. Cancers, 13(17), 4464. https://doi.org/10.3390/CANCERS13174464/S1

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. di, … Guilford, P. (2020). Hereditary diffuse gastric cancer: updated clinical practice guidelines. The Lancet Oncology, 21(8), e386–e397. https://doi.org/10.1016/S1470-2045(20)30219-9

Chen, E. Y., Tan, C. M., Kou, Y., Duan, Q., Wang, Z., Meirelles, G. V., Clark, N. R., & Ma'ayan, A. (2013). Enrichr: Interactive and collaborative HTML5 gene list enrichment analysis tool. BMC Bioinformatics, 14(1), 1–14. https://doi.org/10.1186/1471-2105-14-128/FIGURES/3

Chi, H., Gao, X., Xia, Z., Yu, W., Yin, X., Pan, Y., Peng, G., Mao, X., Teichmann, A. T., Zhang, J., Tran, L. J., Jiang, T., Liu, Y., Yang, G., & Wang, Q. (2023). FAM family gene prediction model reveals heterogeneity, stemness and immune microenvironment of UCEC. Frontiers in Molecular Biosciences, 10. https://doi.org/10.3389/FMOLB.2023.1200335/FULL

Dlamini, Z., Marima, R., Hull, R., Syrigos, K. N., Lolas, G., Mphahlele, L., & Mbita, Z. (2021). Genomics and molecular analysis of RPL9 and LIAS in lung cancer: Emerging implications in carcinogenesis. Informatics in Medicine Unlocked, 25, 100698. https://doi.org/10.1016/J.IMU.2021.100698

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(11), 2841–2867. https://doi.org/10.1002/1878-0261.12948

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. do C., Amil Dias, J., Figueiredo, C., … Carneiro, F. (2018). Phenotypic heterogeneity of hereditary diffuse gastric cancer: report of a family with early-onset disease. Gastrointestinal Endoscopy, 87(6), 1566–1575. https://doi.org/10.1016/j.gie.2018.02.008

Hu, B. li, Xie, M. zhi, Li, K. zhi, Li, J. lin, Gui, Y. chang, & Xu, J. wen. (2019). Genome-wide analysis to identify a novel distant metastasis-related gene signature predicting survival in patients with gastric cancer. Biomedicine & Pharmacotherapy = Biomedecine & Pharmacotherapie, 117. https://doi.org/10.1016/J.BIOPHA.2019.109159

Kosone, T., Takagi, H., Kakizaki, S., Sohara, N., Horiguchi, N., Sato, K., Yoneda, M., Takeuchi, T., & Mori, M. (2006). Integrative roles of transforming growth factor-α in the cytoprotection mechanisms of gastric mucosal injury. BMC

Gastroenterology, 6(1), 1–14. https://doi.org/10.1186/1471-230X-6-22/FIGURES/7

Lim, H. J., Zhuang, L., & Fitzgerald, R. C. (2023). Current advances in understanding the molecular profile of hereditary diffuse gastric cancer and its clinical implications. Journal of Experimental and Clinical Cancer Research, 42(1), 1–12. https://doi.org/10.1186/S13046-023-02622-3/TABLES/3

Lin, L. Y., Yang, L., Zeng, Q., Wang, L., Chen, M. L., Zhao, Z. H., Ye, G. D., Luo, Q. C., Lv, P. Y., Guo, Q. W., Li, B. A., Cai, J. C., & Cai, W. Y. (2018). Tumor-originated exosomal lncUEGC1 as a circulating biomarker for early-stage gastric cancer. Molecular Cancer, 17(1), 1–6. https://doi.org/10.1186/S12943-018-0834-9/FIGURES/3

Liu, L., Yang, J., & Wang, C. (2020). Analysis of the prognostic significance of solute carrier (SLC) family 39 genes in breast cancer. Bioscience Reports, 40(8). https://doi.org/10.1042/BSR20200764

Liu, S., Liu, X., Lin, X., & Chen, H. (2023). Zinc Finger Proteins in the War on Gastric Cancer: Molecular Mechanism and Clinical Potential. Cells, 12(9), 1314. https://doi.org/10.3390/CELLS12091314

Machida, J., Yoshiura, K. I., Funkhauser, C. D., Natsume, N., Kawai, T., & Murray, J. C. (1999). Transforming Growth Factor-α (TGFA): Genomic Structure, Boundary Sequences, and Mutation Analysis in Nonsyndromic Cleft Lip/Palate and Cleft Palate Only. Genomics, 61(3), 237–242. https://doi.org/10.1006/GENO.1999.5962

Masjedi, S., Zwiebel, L. J., & Giorgio, T. D. (2019). Olfactory receptor gene abundance in invasive breast carcinoma. Scientific Reports 2019 9:1, 9(1), 1–12. https://doi.org/10.1038/s41598-019-50085-4

Neuhausen, S. L. (2000). Founder populations and their uses for breast cancer genetics. Breast Cancer Research, 2(2), 77–81. https://doi.org/10.1186/BCR36/TABLES/1

Piao, X. M., Kim, Y. U., Byun, Y. J., Zheng, C. M., Moon, S. M., Kim, K., Lee, H. Y., Jeong, P., Kang, H. W., Kim, W. T., Lee, S. C., Kim, W. J., & Yun, S. J. (2022). Expression of RPL9 predicts the recurrence of non-muscle invasive bladder cancer with BCG therapy. Urologic Oncology: Seminars and Original Investigations, 40(5), 197.e1–197.e9. https://doi.org/10.1016/J.UROLONC.2021.12.009

Pourghasem, N., Ghorbanzadeh, S., & Nejatizadeh, A. (2023). Expression and Regulatory Roles of Small Nucleolar RNA Host Gene 4 in Gastric Cancer. Current Protein & Peptide Science, 24(9), 767–779. https://doi.org/10.2174/1389203724666230810094548

Serum levels of transforming growth factor alpha in gastrointestinal cancer patients. | Cancer Epidemiology, Biomarkers & Prevention | American Association for Cancer Research. (n.d.). Retrieved March 30, 2024, from https://aacrjournals.org/cebp/article/4/2/127/156864/Serum-levels-of-transforming-growth-factor-alpha

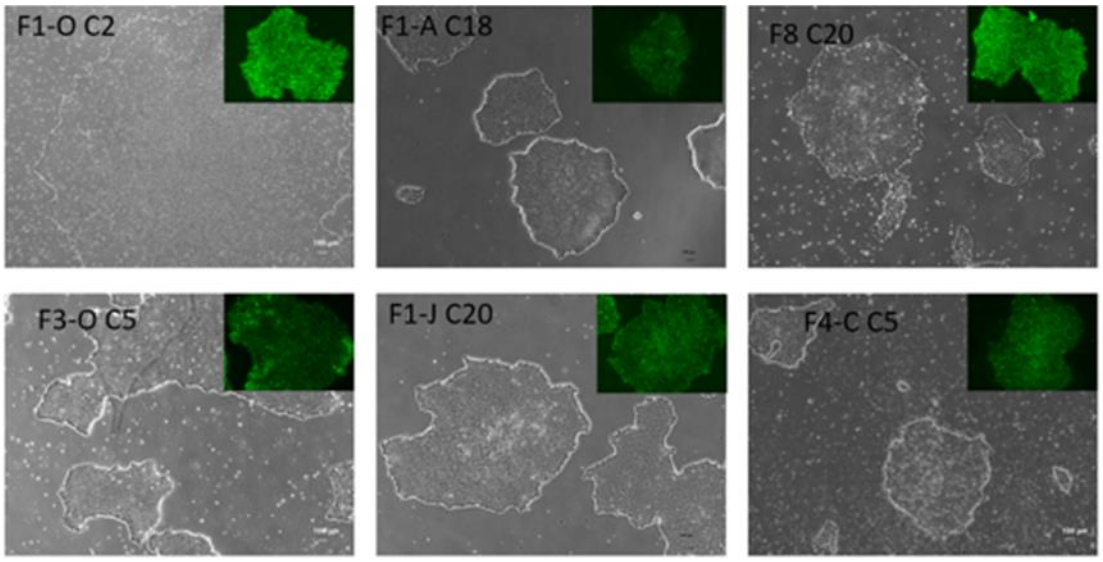
Talotta, R. (2023). Interaction between Long Noncoding RNAs and Syncytin-1/Syncytin-2 Genes and Transcripts: How Noncoding RNAs May Affect Pregnancy in Patients with Systemic Lupus Erythematosus. International Journal of Molecular Sciences, 24(3). https://doi.org/10.3390/IJMS24032259

Wang, X., Xu, M., Yan, Y., Kuang, Y., Li, P., Zheng, W., Liu, H., & Jia, B. (2019). Identification of Eight Small Nucleolar RNAs as Survival Biomarkers and Their Clinical Significance in Gastric Cancer. Frontiers in Oncology, 9, 467340. https://doi.org/10.3389/FONC.2019.00788/BIBTEX

Wolczyk, D., Zaremba-Czogalla, M., Hryniewicz-Jankowska, A., Tabola, R., Grabowski, K., Sikorski, A. F., & Augoff, K. (2016). TNF-α promotes breast cancer cell migration and enhances the concentration of membrane-associated proteases in lipid rafts. Cellular Oncology (Dordrecht), 39(4), 353–363. https://doi.org/10.1007/S13402-016-0280-X

Wu, J., Tong, S., Zhan, Z., Zhang, J., Chen, C., Hu, S., & Wang, R. (2014). TGF-α gene variations and increased susceptibility of gastric cancer in an eastern Chinese Han population. Biomarkers, 19(1), 9–15. https://doi.org/10.3109/1354750X.2013.859736

Supplementary Material



Supplemental Figure 1. Bright field microscopy images and immunocytochemistry for TRA-1-81 (GFP signal) from representative colonies of iPSCS generated for 6 individuals.

Supplemental Table 1. Cohort characteristics

Family	Sample ID	Class	Case ID	Age	Sex	WGS Available	Whole Blood RNA-seq Available (n=1)	PBMCs RNA-seq Available (n=1)	iPSCs RNA-seq Available (n=2)	FFPE Stomach
F1	1	Young DGC	F1	35	F	Yes	No	No	No	No
F1	5	Young DGC	F1-A	43	M	Yes	Yes	Yes	Yes	No
F1	2	Asymptomatic carrier ≥ 39 years	F1-C	64	M	Yes	No	No	No	No
F1	3	Asymptomatic carrier≤ 35 years	F1-F	35	F	Yes	No	No	No	No
F1	35	Old Asymptomatic Carrier (OAC)	F1-I	89	M	Yes	No	No	No	No
F1	4	Old Asymptomatic Carrier (OAC)	F1-J	71	M	Yes	Yes	Yes	Yes	No
F1	37	Asymptomatic carrier≤ 35 years	F1-O	26	F	Yes	Yes	Yes	Yes	No
F3	10	Young DGC	F3	18	M	Yes	No	No	No	No
F3	11	Asymptomatic carrier ≥ 39 years	F3-A	61	F	Yes	No	No	No	Yes
F3	13	Asymptomatic carrier ≥ 39 years	F3-C	39	M	Yes	No	No	No	No
F3	16	Asymptomatic carrier ≥ 39 years	F3-D	42	M	Yes	No	No	No	Yes
F3	17	Asymptomatic carrier ≥ 39 years	F3-E	43	F	Yes	No	No	No	Yes
F3	15	Asymptomatic carrier≤ 35 years	F3-F	14	F	Yes	No	No	No	No
F3	24	Asymptomatic carrier ≥ 39 years	F3-H	58	F	Yes	No	No	No	No
F3	25	Asymptomatic carrier≤ 35 years	F3-I	31	M	Yes	No	No	No	No
F3	18	Asymptomatic carrier≤ 35 years	F3-K	19	M	Yes	No	No	No	No
F3	21	Old Asymptomatic Carrier (OAC)	F3-O	90	F	Yes	Yes	Yes	Yes	No
F3	0	Young DGC	F3-R	35	M	No	No	No	No	Yes
F4	31	Asymptomatic carrier≤ 35 years	F4-A	20	M	Yes	No	No	No	No
F4	32	Asymptomatic carrier ≥ 39 years	F4-B	45	F	Yes	No	No	No	No
F4	33	Old Asymptomatic Carrier (OAC)	F4-C	84	F	Yes	Yes	Yes	Yes	No
F8	40	Young DGC	F8	33	F	Yes	Yes	Yes	Yes	Yes

Supplemental Table 2. List of age-related genes from literature search

Authors:	Magalhaes et al 2009	Harries et al 2011	Glass 2013	Peters 2015									Yang Hu et al 2022
PMID:	PMID: 19189975	PMID: 21668623	PMID: 23889843	PMID: 26490707									PMID: 35822054
Source material:	Meta-analysis (differente tissues) n=71	PBMCs (n=50)	Blood (n=2)	PBMCs (n=1497 genes)									PBMCs (n=7)
Genes	APOD	LRRN3	ASPA	CD248	SPRY1	KLRF1	GNL3	URI1	RASSF1	TCF3	MIR600HG	HIC2	TCERG1	NUDT7
	FCGR2B	CD248	OTUD7A	LRRN3	ATP2B4	PDK4	RPS6	MBOAT1	ZNF302	SERTAD2	MPO	UBXN1	WDR1	CLPB
	C3	LRRN3		NELL2	FOXO1	NMUR1	BATF	LRPPRC	DNASE1L3	CD3D	PSIP1	SCIMP	PODXL2	OXNAD1
	CLU	LEF1		LEF1	MMP28	BCAS4	IRF8	RCAN2	EIF4B	GIMAP7	NOD2	USP4	IQCG	MLLT3
	CTSS	LEF1		CCR7	LGMN	IARS	CD79B	TRMT13	TOP1MT	NOLC1	GNPTAB	PTTG1	PTK2B	ADCY4
	LYZ	NELL2		ABLIM1	CD84	C6orf48	SRPRB	PPRC1	C6orf108	C22orf29	PIK3C2B	TAP1	SPPL2A	PIK3CA
	IGJ	ABLIM1		GZMH	DENND2D	RSL1D1	GPA33	PDCD2L	RBM26	EIF3G	TCN2	MRPL35	GNGT2	NEO1
	C1QA	FAM102A		MYC	CLUAP1	VEGFB	NOL11	BCKDHB	SLC04C1	MRPS27	C3orf37	POLG2	IDH1	
	S100A6	CCR6		CCR7	OGAD2	PTGDR	TESPA1	AFF3	SRSF6	GBP1	MIF	ANXS1A	DCAF4	
	GFAP	ZSCAN18		FAM102A	GZMA	IMPI3	GPR137B	PARP1	SERPINB6	DIMT1	PHC2	HSPA6	RBFA	
	MT1F	BZW2		SERPINE2	CD79A	SEPT6	IRAK3	RPL8	ENTPD1	TYSND1	USP3	NLRC5	ALG13	
	C1QB	CCR7		SLC16A10	PDGFRB	LGR6	PUS1	FRMD3	PRDM1	SFTPD	GSTO1	PPWD1	NEDD9	
	LGALS3	LTB		FCGBP	RCAN3	POLR1C	SET	CD19	DSC2	PITPNC1	FGR	NSUN2	CHORDC1	
	MGST1	GRAP		GPR56	TSPAN13	PLAG1	ASCL2	EBF1	FOXJ3	TNFRSF21	LSM5	GDE1	UBAC1	
	ANXA3	HGFL		BACH2	HOOK1	AXIN2	MKI67IP	NOP2	H2AFY2	ZNF101	TCF4	NCBP2	PRMT1	
	ANXA5	PASK		SYT11	KIAA1147	TTC3	LPCAT1	CHGL2	E2F2	ANAPC1	ELP2	KRI1	TMEM5	
	S100A4	KIAA1147		PDE9A	THNSL1	ZNF135	CYP4F22	B4GALT5	WRN	ZNF480	PAFAH1B3	USP44	RPL11	
	GPNMB	ABLIM1		NOG	ATIC	FAM129C	MAFB	EIF2D	PUS7	IFNG	LPAR6	NIN	TRIB1	
	GSTA1	SYPL1		FLNB	NAA16	SPOCK2	ITK	SLC26A8	C12orf23	BRIX1	NONO	ZNF175	SLC39A1	
	C1QC	CD27		NT5E	IGFBP3	CNN3	NRIP1	WARS	SPON1	SYNE2	GOLGA3	C1orf216	TLN1	
	VWF	DQ894085		FGFBP2	SUSD1	ZNF439	FAM214B	SLC22A15	RBL2	FAM46A	LAT	TRIM21	ZNF581	
	DERL1	DKC1		TGFBR3	CD28	MTA3	PDE3B	MARS2	EFHD2	RCL1	FOXD2	GRWD1	DZIP3	
	NDRG1	FAIM3		ITM2C	RORC	C2orf40	LINC00152	NME1	ITFG2	SLAIN1	CCNB1IP1	LOC338799	CLNS1A	
	RatcAnia4	CDR2		ATF7IP2	RAD54B	GBP5	GRPEL2	CBLL1	KIAA0355	RABGGTB	USP11	THAP11	ACTN4	
	FCGR2A	ATM		CR2	MAN1C1	EPB41L4A-AS1	C9orf123	CPA5	ARHGEF40	FAR2	TNIP3	LARP4	SURF6	
	GBP2	VAMP5		FAIM3	STRBP	EFNA1	CMKLR1	HSD17B8	GLTSCR2	ACTA2	EHD4	RPS29	GPR97	
	GNS	GBP5		PHGDH	TARBP1	BHLHE40	SH2D1B	C12orf57	TPS3I3	TMEM8B	NSMCE1	TCIRG1	AGPAT5	
	IL33	STAT1		LDHB	BANK1	APOBEC3G	ANXA2	PPA1	ATL3	WDR67	PRSS30P	TDP1	VPS33A	
	MPEG1	WARS		SIRPG	KLF6	ZNF256	SEH1L	ABI3	CHCHD10	PRKCA	CRLF3	VARS	BMX	
	NPC2	WARS		FCRL6	KLRD1	LPP	TNFRSF17	RPS13	RAP2A	ESYT1	SLC4A4	CYCS	RPAIN	
	ADIPOR2	MT2A		PDE7A	TRA@	C1orf162	ABCE1	ANKRD22	LSM2	ZNF329	RPF2	IFT74	EPB41L4A	
	B2M	CTSA		NOSIP	FAM159A	NIPAL3	MORC2	RPL5	EBAG9	SMARCD3	UBIAD1	SORCS3	PTPN13	
	LITAF	SLC27A3		PAICS	NPM3	P2RX7	TRAF3IP2	RBM17	CTSW	DKFZP586I1420	VCL	HNRNPH1	PLOD3	
	SERPING1	FBXO6		BZW2	ZEB2	EXOSC5	ZNF671	SH3YL1	ANG	HMOX1	PPP2R5C	RPL17	PRPF19	
	HCST	VCAN		OXNAD1	VAMP5	SLC25A20	DBP	CD58	AKR1C3	ZZZ3	OCIAD1	BAG3	EBPL	
	RNF213	EIF4G3		CX3CR1	ZNF154	GAB3	TMX4	RASGRP2	SLC27A3	ARMCX2	ZNF548	LRRC33	SLC20A1	
	C4A	DUSP3		SCML1	DGKA	PAQR8	ALKBH2	SRM	ZCCHC11	PCNX	IL12RB1	SIAH2	TCTN1	
	SPP1	GZMH		RPL22	KAT2A	ENOSF1	ATP6V0E2	FAM50B	ADK	TRMT112	SMAD7	FAM134A	SMARCC1	
	CLIC4	PSMB9		LDLRAP1	POU2AF1	EIF3E	MAPKAPK5-AS1	CCDC50	TSEN2	DDX18	CCDC41	ERGIC1	IFIT3	
	EFEMP1	TTC38		RHOC	EPHA1	IGLL1	MT1E	TCEA3	GPX7	NRG1	TSPAN17	TOMM70A	TNFRSF1B	
	LAPTM5	SYT11		LTB	TPH2	EOMES	RPS4X	AGPAT4-IT1	CCT4	YTHDC1	AKR1B1	ACADVL	UPF1	
	VAT1	FCGR1A		FAM134B	C1QBP	TEX10	PLBD1	CYP2J2	NOA1	SDCCAG3	CSRP1	MFN2	ENPP4	
	HIST1H1C	WIPI1		LBH	LTBP3	MAGED1	PPAT	FAM153B	ATXN1	CYB561	ANXA4	SPECC1L	FBXO15	
	KIAA0494	P2RX7		PRSS23	PLCG1	KLF10	KIF1B	RNASE2	NOP58	LGALS3BP	MVP	PRRG4	ARRB1	
	PTGES3	TTC38		SUSD3	IGFBP7	UBASH3A	KIAA0125	AKTIP	UBR2	VPS51	CEBPA	HSPA9	SSH1	
	Rat:LOC360231	H3F3A		PIK3IP1	ITGA6	SPTBN1	GPR114	ALG8	SREBF1	CCDC170	LCMT2	FAM160A2	EXOSC8	
	RASA3	STOM		MFGE8	F5	MTSS1	NCAM1	NREP	EIF4A2	SLC7A7	MTHFD1L	IRF9	TMEM230	
	PSMD11	STAT1		AGMAT	B3GAT1	HSF2	TNFRSF13B	NCL	EIF2A	ZNF683	MFSD10	ERGIC3	ZNF121	
	SGK1	MT1A		NKG7	TMEM204	PTPN7	SLC31A1	RPUSD2	PALLD	TNFAIP6	TTYH3	CRELD1	NLRP3	
	TXNIP	B3GAT1		PPP2R2B	CD96	ARHGEF18	LY9	HLA-DOB	MFNG	DMXL2	PFKFB2	TBRG4	FAM129B	
	Mouse:Igh-VJ558			GZMB	VCAN	TRAP1	WDR4	TPST1	NPC2	CD86	DOK2	SEPT9	METTL25	
	HBA1			SLC7A6	RRS1	TSHZ2	N6AMT1	TIGIT	TNF	ATG7	RPL4	LEO1	RPL27A	
	HLA-G			TTC38	HSPH1	EIF3H	CDK6	CERS6	SLFN13	KIF13B	CLSTN3	FUCA2	ZNF234	
	PCSK6			IMPDH2	SESN1	SNTB2	MAP3K5	GTPBP4	OGDH	VIM	SLAMF7	HIBADH	UBE2G2	
	TMED10			GRAP	CRIP1	ZNF550	MAL	NPAT	THEM156	CAPNS1	MRPL3	MTERFD1	NSUN5	
	TFRC			NUCB2	TBC1D4	BCL11A	UBAS2	RAP1GAP2	APBA2	TMEM123	TCN1	DHRS9	CYB5A	
	COL3A1			FBL	ANXA2R	CTNNA1	PRMT3	RPLP0	STX11	GAL3ST4	ACAA2	ANAPC16	KIAA0586	
	COL1A1			PRRSL	HAVCR2	HLA-DOA	UXT	CD163	RAB11FIP5	AIMP1/p43	P2RX5	TAGLN	SETBP1	
	COL4A5			SATB1	CD7	IL10RA	RPL18	MRPL45	CTSL1	PDIA5	RASGRP3	AHSA1	USP13	
	FABP3			IL7R	STMN1	HSP90B1	SPIB	HADH	SERBP1	MR1	PRPSAP2	PNPLA6	TOX	
	ATP5G3			ZSCAN18	GNLY	SCD	BNIP3	C3AR1	ZBTB25	HDAC2	HAPLN3	MZF1	CCR9	
	NDUFB11			ADRB2	APEX1	POLR1E	BEX2	SRPK1	MAML2	PRDM8	DUSP3	SLC25A15	HSPA8	
	CX3CL1			KLRB1	ICOS	SNRPN	ARG1	LHFPL2	AHCY	ST6GAL1	HNRNPM	IFI30	RASGRP1	
	CA4			HSP90AB1	GK5	SORT1	TTLL3	EIF4G3	MGST3	PLA2G16	EIF3D	GPX1	SOX8	
	CALB1			VPREB3	SAE1	ADSL	BLK	HMGN1	RGCC	SH3YKD2A	JAKMIP1	ABCF1	GLMN	
	GHITM			TCF7	FLT3LG	LARS	CPT1A	CSTF3	C19orf59	GNS	RAB31	REPS2	METAP2	
	U0CR0			OSBPL5	ERBB2	SYPL1	LMNA	CDKN1C	RRP15	ZCCHC7	TRAT1	QTRT1	CALML4	
	C5orf13			IGJ	PYHIN1	PTGER2	FCGR1A	SLAMF6	ANKRD13A	KCNG1	CMSS1	CFH	PLCB2	
	ACSS2			C1orf21	NME8	SLC38A1	TTC27	MS4A6A	RPSAP58	TRIM28	LUC7L	RAB7L1	DHPS	
	U0CRFS1			BCL11B	FAM13A	CIRH1A	GIMAP5	RALA	GALC	VNN3	DDX21	SIRT4	CELA1	
	DIABLO			NMT2	PMEPA1	GSPT2	IL1R2	ZNF529	NUP160	RPL19	DNMT3A	PTP4A2	PPIL3	
				TNFRSF25	RAPGEF6	APH1B	RPL13	ACVR2A	TSPAN3	GTPBP3	ZFAND1	GNB2L1	TGFBR2	
				PTPRK	ZNHIT6	E2F5	WDR12	WIPI1	NDST1	MMP9	LRRC8C	CBY1	SRRM1	
				BTBD11	PASK	SERINC5	EIF2M	NBEAL2	DNMT1	HVCN1	PLAC8	CCR2	ELOVL6	
				EDAR	AGPAT4	EIF3L	KIF5C	ZNF22	OSBPL10	RBPJ	CD151	GYPE	TAF4	
				ELOVL4	STOM	CHMP7	CCT8	RALGDS	TMTC1	MAP3K4	ABCB1	CTPS1	CYBB	
				LGALS1	TRIB2	MARCH1	DSC1	UPP1	CDC42EP4	SLCO3A1	SLC1A7	SPON2	IL10RB	
				KLHL3	AHNAK	DDX10	PARP3	GAR1	C16orf58	RPP40	CAST	XPO4	CEPT1	
				CD8B	DDX47	LAGE3	ASGR2	ADPRM	SMARCA1	PROS1	SGPL1	IPO4	C4orf34	
				N8EA	CDCA7L	ITGB1	CAPN2	DDX55	MS4A4A	PDCD4	RAB37	BLOC1S1	ABHD3	
				NUAK1	S100A10	WDR75	STAT1	BMS1	SLC38A11	ARHGAP26	MRPL24	DEDD	MTRF1	
				GTSCR1	GNG7	CCDC109B	CPEB3	TMEM66	PARP8	LTA	ID2	LPHN1	STAT2	
				PDGFD	GEMIN4	PFAS	TMEM97	APMAP	RPS2	EPRS	TIMD4	PDZK1IP1	GTF3A	
				ANXA1	TRAF5	FAM213A	RRAS	ALOX5AP	KDM1A	PAG1	FGL2	PLIN2	AOAH	
				APOBEC3H	CD3E	TIGD7	LRRFIP2	SERPINB8	ADAP2	METRNL	PVRIG	CACYBP	ICAM4	
				ATM	CLDND2	PKIA	KLHL6	METTL1	TRAFD1	MAPK14	GLT25D2	STAT3	NAGA	
				ID3	CXCR5	CD70	ROPN1L	ANTXR2	NPC1	TTC39B	PADI4	SUPT3H	MSH2	
				PCED1B	FBLN2	ARHGAP32	PACSIN1	GDPD5	RPL13P5	MRI1	NXT1	HNMT	TRIM44	
				COBLL1	PAPSS2	ZNF365	KIAA0020	ZNF30	MSC	PLOD1	C7orf23	CPNE8	SAMD9L	
				MZB1	RNF144A	EEF2	DPP4	LTV1	HK3	C11orf1	MYOF	OPN3	RARRES3	
				GPR18	KLF11	MT1A	GFI1	LYPD2	CRISPLD2	LGALSL	CFLAR	CPQ	TSPYL5	
				CEP41	GPR141	PDZD4	PLEK	SLC31A2	SHMT2	GOLM1	DUSP5	FKBP11	STARD3NL	
				ETS1	CD22	DPH5	CCL5	GAMT	SERPINF1	RRP1B	THEM6	MTR	KANSL2	
				FAM117B	ZBP1	OBSCN	MTUS1	VSIG4	LILRB1	GBA	S1PR5	ZNF721	INPP5D	
				EPHX2	CHN1	DOCK10	VAMP3	GABARAPL1	HNRPDL	ITGA2B	CD97	IRF1	PCSK5	
				CAMK4	SLFN11	NAT10	GLIPR2	NUP107	NOC3L	ZNF146	F13A1	ZNF260	SLITRK4	
				MCOLN2	CD244	CKAP4	DYRK2	CCT3	WDR59	RPS16	C1orf109	GLA	KIAA0913	
				CCR6	ILK	ZNF544	FBXO6	CDK4	C10orf55	CAPN1	PAQR4	AKAP7	ARHGAP24	
				TXK	CST7	TC2N	RPS23	TMEM99	TIPARP	HNRNPD	NACC2	ILF2	RIOK2	
				LEPROTL1	AK5	BTLA	IDS	FCRLA	CMTM8	HERPUD2	PLEKHB1	UBTF	ARF3	
				KYNU	PTGDS	TLE2	GPATCH4	TCEA2	NOL8	TOP2B	MRPS26	NOB1	CDKN1A	
				F2R	EIF2S3	CCT2	MEGF9	CD63	LINGO2	NHP2L1	EEF1B2	TADA3	ATF3	
				PRF1	HSPD1	C10orf2	TSPAN2	PHF14	TIMM21	CTSG	SRBD1	BPI	FCGR2B	
				DKC1	DNAJA3	EVL	PNPO	MGC24103	C5orf51	DUSP18	CXCR7	PSMB9	SSBP4	
				CACNA2D2	TXNDC5	FASLG	DPH2	AARS	CYBASC3	DCXR	LCK	DENND1A	C11orf75	
				SEPHS1	ADAM17	WDR74	YPEL1	NARS2	ZDHHC9	HMGB2	MORC2-AS1	MPHOSPH9	TMEM209	
				CLEC2D	ZNF239	HNRNPH3	PIGN	NLRP12	DPPA4	EIF4A3	COLQ	OASL	TBCB	
				TPP2	ORC5	PPAPDC1B	ATP5G2	PIK3CG	ZNF428	FGF9	GART	CTSC	RIPK3	
				APRT	MRPS18B	NFKB1	MAF	C3orf62	SRSF5	BIVM	TPT1	LAPTM4B	MAD2L2	
				SERPINB1	CCT7	YIPF1	ANXA5	THBS1	TNFSF13B	ATP5G1	SUMF2	EPG5	CSNK2A2	
				QSOX2	C12orf65	NSUN6	ZNF470	RNMTL1	GCNT1	FAM165B	SRFBP1	MYD88	ACADM	
				SESTD1	MSH6	CD300C	TCL1A	FUT8	DRAXIN	U2AF1L4	CDC16	BOD1	FAM70A	
				DIAPH1	RNMT	NCEH1	TFAM	DGKG	RIOK1	ATP1B3	RBM11	ACOT9	DDX1	
				RAB27A	ITM2A	CXCR6	VAV3	NAA25	DAAM2	LARP1B	HLA-B	AGPAT3	MGLL	
				USP47	PPP1R3E	S100A8	ZMYND19	SLC22A4	ZNF830	DENR	ABHD5	CDK14	ITPRIP	
				THOC1	TSPYL2	C21orf7	LAIR2	DENND1B	BCL6	MRPS9	HDAC1	MEOX1	EPM2A	
				RPAP2	NET1	NARG2	VIPR1	BCL9	SLC24A6	C11orf31	ORM1	CALCOCO1	PGRMC2	
				MXRA7	IGIP	PEBP1	SOX13	APLP2	FXR1	CYP2E1	TCTN3	ARHGAP15	RPL30	
				TRAPPC6A	CXCL16	MARCH3	SLC9A9	FLVCR2	AFG3L2	IPCEF1	MOAP1	PAM	TSGA10	
				EML4	PPAN	NKRF	WDR89	VOPP1	MRTO4	NCR1	PFKFB4	RBM14	EDARADD	
				ZNF573	PLEKHF1	C21orf2	NGDN	PTPLAD1	P2RY10	RRN3P1	GUSB	ITPK8	ENDOD1	
				HLTF	TOMM20	ST13	CTPS2	DYSF	KIAA0922	DCK	PHOD1	RPSA	USF1	
				ZNF827	RCN2	LINC00260	POLR3H	CLPP	CTSD	SRSF1	SAAL1	EWSR1	PGD	
				CREB3L2	UBE2F	NAP1L1	ZBED5	RASSF4	KLF12	TGDS	ETFDH	RPS6KB1	PRIM1	
				PLAUR	RPS20	TLR10	UBE2D2	CDC14B	ANP32B	HSPE1	PARP14	BCCIP	EIF4EBP2	
				LAMP5	PNN	HNRNPA0	DOLPP1	ZNF385A	PNOC	NACA	GP6	TF82M	FARSB	
				MRPS31	C12orf75	B3GALT2	TSR1	ERCC6L2	AGPAT2	CXXC5	ZRANB2	GPR68	GALNT10	
				TNFAIP8	ADCY9	RPUSD4	STX6	THEM220	MAK16	GRB10	NUP62	DHX15	EMP3	
				ITGAL	NUPB8	CECR5	MCCC1	CUTA	TASP1	PPIL1	MIR17HG	C1QTNF6	NSG1	
				LILRA5	FAM126B	DDX31	SEL1L3	MARCO	RPL28	CEP135	ARSG	GORASP2	BNC2	
				FKBP4	LIG1	MST4	TMEM42	DLGAP4	PSAT1	PIGP	CCDC34	FAM43A	IPO11	
				TBXAS1	VNN1	BYSL	CLEC4E	UTP14A	IL15	HP	CMTM3	SH2D2A	ABR	
				ARHGAP19	CPM	TAGAP	ST6GALNAC6	AKIRIN2	YTHDF1	ELF4	ABCF2	PRKRA	ACSL1	
				LYRM4	ADAMTSL4	TIMM9	PARP9	NOP56	MIS12	ZNF706	TK2	MBTPS1	ZNF567	
				CYFIP1	ADHFE1	PRMT6	TIMP1	TRIM69	ABCA6	SMARCAD1	SLC26A6	MGC57346	CCNC	
				DIRC2	WDFY2	EPHA4	KLHL34	EXOSC2	EFHD1	ZNF10	RPS6KA1	RUVBL1	SNX11	
				ABCC3	C22orf32	MAP3K3	TTC16	AKAP13	GAS7	RUNX3	DUSP10	STIP1	FBXO10	
				IGHV5-78	AGFG1	EMP1	SH3BP2	CSF1R	RAB24	OAS1	RPL23AP64	CD37	EBP	
				BIN1	TRMT5	RDH14	CLIC1	ZEB1	MAPK1	NGLY1	LARS2	POLE3	CD72	
				APOBEC3C	GRAMD3	CNOT7	VRK1	PHTF1	P2RX1	RPL35	TBC1D2	FAM220A	ZNF189	
				AUTS2	LUC7L3	TSPO	S100A12	IL18BP	PDCL3	EIF2C4	OXCT1	SFXN4	SH2D3A	
				ATP5L	KLF4	RPS14	GNG2	BTN3A1	LPGAT1	TRA2B	FANCA	RPP21	HIVEP2	
				NDFIP1	SNRPF	TCTA	FAM53B	MS4A1	NUDT9	QSOX1	MAGEE1	PRKACB	SH3BP5L	
				ITGAM	LOXL3	FEZ1	QPCT	MPHOSPH10	UQCC	EI24	TMBIM1	TRIM38	CCDC57	
				GYG1	PARM1	LINC00324	DAP3	NOL6	IL27RA	CYSTM1	RIPK2	LCLAT1	ABHD16A	
				SEPT1	UBAP2	TUT1	BLMH	HCG27	MYL6	MGAM	HLA-E	ADAR	POLRMT	
				DPYSL2	PLXDC2	WBP1L	HIST1H3H	FBXL5	PAK1IP1	ETV6	ZNF337	MATN2		
				RPL36	DEFA4	EIF2AK3	MRPS33	RPL13AP3	DYNLT1	N6AMT2	CCT6A	PJA1		
				SALL2	CYP1B1	KDSR	SH3BGRL3	TP53I11	INPPL1	ZNF800	BCS1L	AGAP1		
				CD5	PCBD1	SFXN2	KIAA0319L	ADORA2A	MAPKAPK3	C20orf197	RPS18	ATAD1		

Supplemental Table 3. List of GO TERMS (Biological processes and Molecular functions) associated with DEGs in blood, PBMCs and iPSCs.

Blood (n=19 genes)	Blood BP	P-value	Adjusted p-value	Odds Ratio	Combined score
1	Regulation Of Megakaryocyte Differentiation (GO:0045652)	0.0001448	0.01477	138.16	1221.34
2	Positive Regulation Of Myeloid Cell Differentiation (GO:0045639)	0.0005284	0.02695	69.02	520.81
3	Positive Regulation Of Granulocyte Differentiation (GO:0030854)	0.004741	0.06434	277.46	1484.80
4	Mitochondrial Protein Catabolic Process (GO:0035694)	0.004741	0.06434	277.46	1484.80
5	Regulation Of Axon Guidance (GO:1902667)	0.005687	0.06434	221.96	1147.41
6	Tubulin Complex Assembly (GO:0007021)	0.005687	0.06434	221.96	1147.41
7	Negative Regulation Of T Cell Mediated Cytotoxicity (GO:0001915)	0.006632	0.06434	184.95	927.70
8	Mitochondrial Outer Membrane Permeabilization (GO:0097345)	0.006632	0.06434	184.95	927.70
9	Immune Response-Inhibiting Cell Surface Receptor Signaling Pathway	0.007576	0.06434	158.52	774.04
10	Positive Regulation Of Megakaryocyte Differentiation (GO:0045654)	0.007576	0.06434	158.52	774.04

Blood (n=19 genes)	Blood MF	P-value	Adjusted p-value	Odds Ratio	Combined score
1	Oxidoreductase Activity, Acting On Paired Donors, With Incorporation Or	0.004741	0.07823	277.46	1484.80
2	Myosin Light Chain Binding (GO:0032027)	0.004741	0.07823	277.46	1484.80
3	GTPase Activator Activity (GO:0005096)	0.01683	0.1111	11.13	45.46
4	Phosphatase Inhibitor Activity (GO:0019212)	0.02070	0.1111	52.80	204.75
5	Ubiquitin-Like Protein Transferase Activity (GO:0019787)	0.02144	0.1111	9.76	37.50
6	Ferrous Iron Binding (GO:0008198)	0.02349	0.1111	46.20	173.28
7	Protein Homodimerization Activity (GO:0042803)	0.02356	0.1111	5.50	20.61
8	MHC Protein Binding (GO:0042287)	0.03182	0.1302	33.58	115.78
9	Alpha-Tubulin Binding (GO:0043014)	0.03550	0.1302	29.95	99.96
10	Ubiquitin Ligase-Substrate Adaptor Activity (GO:1990756)	0.04100	0.1330	25.76	82.28

iPSCs (n=16 genes)	iPSCs BP	P-value	Adjusted p-value	Odds Ratio	Combined score
1	Regulation Of Transport (GO:0051049)	0.0006299	0.05720	63.30	466.55
2	Protection From Natural Killer Cell Mediated Cytotoxicity (GO:0042270)	0.004243	0.05720	312.17	1705.22
3	ERBB2-EGFR Signaling Pathway (GO:0038134)	0.004243	0.05720	312.17	1705.22
4	Galactolipid Biosynthetic Process (GO:0019375)	0.004243	0.05720	312.17	1705.22
5	Galactosylceramide Biosynthetic Process (GO:0006682)	0.004243	0.05720	312.17	1705.22
6	Plasma Membrane Raft Organization (GO:0044857)	0.004243	0.05720	312.17	1705.22
7	Glycosyl Compound Biosynthetic Process (GO:1901659)	0.004243	0.05720	312.17	1705.22
8	Vascular Endothelial Growth Factor Receptor-2 Signaling Pathway	0.004243	0.05720	312.17	1705.22
9	Wybutosine Biosynthetic Process (GO:0031591)	0.004243	0.05720	312.17	1705.22
10	Wybutosine Metabolic Process (GO:0031590)	0.004243	0.05720	312.17	1705.22

iPSCs (n=16 genes)	iPSCS MF	P-value	Adjusted p-value	Odds Ratio	Combined score
1	Vascular Endothelial Growth Factor Receptor Activity (GO:0005021)	0.005090	0.08267	249.72	1318.68
2	Cysteine-Type Endopeptidase Inhibitor Activity Involved In Apoptotic Process	0.01603	0.08267	69.32	286.51
3	Metal Cation:Proton Antiporter Activity (GO:0051139)	0.01771	0.08267	62.38	251.64
4	Zinc Ion Transmembrane Transporter Activity (GO:0005385)	0.01938	0.08267	56.71	223.63
5	Epidermal Growth Factor Receptor Binding (GO:0005154)	0.02105	0.08267	51.98	200.68
6	pre-mRNA Binding (GO:0036002)	0.02521	0.08267	43.00	158.28
7	Transition Metal Ion Transmembrane Transporter Activity (GO:0046915)	0.02769	0.08267	38.97	139.76
8	Hsp90 Protein Binding (GO:0051879)	0.02935	0.08267	36.67	129.39
9	Transmembrane Receptor Protein Kinase Activity (GO:0019199)	0.03100	0.08267	34.63	120.30
10	Transmembrane Receptor Protein Tyrosine Kinase Activity (GO:0004714)	0.04168	0.09627	25.43	80.80

PBMCs (n=2 genes)	PBMCs BP	P-value	Adjusted p-value	Odds Ratio	Combined score
1	Ammonium Homeostasis (GO:0097272)	0.0004999	0.004499	4998.50	37993.75
2	Ammonium Transmembrane Transport (GO:0072488)	0.0009998	0.004499	2221.00	15342.67
3	Cytoplasmic Translation (GO:0002181)	0.009279	0.02732	216.37	1012.62
4	Peptide Biosynthetic Process (GO:0043043)	0.01574	0.02732	126.38	524.67
5	Nitrogen Compound Transport (GO:0071705)	0.01623	0.02732	122.44	504.55
6	Macromolecule Biosynthetic Process (GO:0009059)	0.01822	0.02732	108.88	436.11
7	Translation (GO:0006412)	0.02326	0.02938	84.83	319.03
8	Inorganic Cation Transmembrane Transport (GO:0098662)	0.02820	0.02938	69.66	248.60
9	Gene Expression (GO:0010467)	0.02938	0.02938	66.79	235.59

PBMCs (n=2 genes)	PBMCs MF	P-value	Adjusted p-value	Odds Ratio	Combined score
1	Ammonium Transmembrane Transporter Activity (GO:0008519)	0.0009998	0.003999	2221.00	15342.67
2	Inorganic Cation Transmembrane Transporter Activity (GO:0022890)	0.003497	0.006924	587.18	3320.99
3	Channel Activity (GO:0015267)	0.005193	0.006924	391.12	2057.43
4	RNA Binding (GO:0003723)	0.1361	0.1361	13.18	26.29

4. DGLBCS experimental resources

4.1. iPSCs Biobank Establishment

In this part of the project, we aimed to generate nine human derived induced pluripotent stem cells (iPSCs) from blood samples and establish for the first time a cellular model using iPSCs for families with DGLBC syndrome. The knowledge herein obtained was extremely helpful to implement this technique at the Expression in Regulation in Cancer group and to help in the establishment of an induced-pluripotent stem cell core at i3s. This is the protocol established for the reprogramming of patient-derived iPSCs.

Keywords: Diffuse gastric cancer and lobular breast cancer syndrome, *CDH1*, human derived induced pluripotent stem cells,

INTRODUCTION

Induced pluripotent stem cells (iPSCs) research achieved a lot of success since their discovery in 2006 by Shinya Yamanaka [1]. Besides being an excellent model of disease research due to their differentiation ability giving rise to different tissues, they have been considered a promising strategy for regenerative therapy, personalized human genetics, pharmacogenomics, and drug-screening studies [2], [3].

The interest in iPSCs research and their potential led to the establishment of several iPSCs-biobanks in highly recognized institutes such as the National Institutes of Health Center for Regenerative Medicine (NIH CR), the Coriell's Stem Cell Laboratory, the RIKEN BioResource Research Center in Japan, and spread through many others European universities.

The use of iPSCs is a valuable resource to overcome many of the challenges associated with studying rare diseases, such as the DGCLBC syndrome. We can create cell lines that mimic the genetic background of affected individuals, establishing an infinity supply of cells from a small non-invasive patient sample. This not only avoids some of the ethical concerns associated with the use of embryonic stem cells but also decreases the use of animal models for research purposes [4].

In the context of my PhD project, I was involved in the establishment of the first biobank for families with DGLBC syndrome from the Northern region of Portugal carrying the same causative germline variant in the *CDH1* gene defined as a founder event [5]. Until this moment, we collected 36 DNA samples that were analysed by whole genome sequencing, 23 Archived Formalin-Fixed Paraffin-Embedded (FFPE) Blocks from which we extracted DNA and RNA, and 13 Blood samples from which we extracted the Peripheral Blood Mononuclear Cells (PBMCs) and selected 9 cases for reprogramming.

Various protocols have been developed for deriving iPSCs, differing in the source tissue for cell collection, the cocktail of genes used to achieve reprogramming and the methodology employed to address off-target effects or to enhance efficiency. An important consideration is that irrespective of the reprogramming technique used, substantial changes in the epigenetic profile of donor cells occur during iPSC induction.

Several studies have compared the reprogramming efficiency of different methods [6], [7]. One of the most widely used is the Sendai virus system, a Viral non-integrating reprogramming strategy that requires the expression of all four Yamanaka factors (i.e., Oct4, Sox2, Klf4, and c-Myc).

The Sendai virus system has shown to be highly efficient to generate iPSCs from peripheral blood samples and a rapid clearance of RNA vectors [8]. Moreover, several studies report that reprogramming efficiency was not affected by the age of the donors using PBMCs as the starting cell material [6]. Consequently, it stands as our method of choice the reprogramming of PBMCs using the CytoTune iPS 2.0 Sendai Reprogramming Kit.

MATERIAL

1. Cells and vectors:

- Frozen PBMCs extracted from fresh peripheral blood at i3S

- CytoTune™ 2.0 Sendai reprogramming vectors (Invitrogen, ThermoFisher, Cat# A16517)

2. Cells' media

- PBMCs (Erythroblast) Expansion Medium (feeder-free):

Table 1. Reagents needed to prepare expansion medium

Expansion Medium	**Working Concentration**	**Final Concentration**	**Volume (5 mL)**
Primocin	500X	100 µg/mL	10 µL
Ascorbic Acid (AA)	5 mg/mL	50 µg/mL	50 µL
SCF	100 µg/mL	50 ng/mL	2.5 µL
IL-3	50 µg/mL	10 ng/mL	1.0 µL
EPO	2 U/µL	2 U/mL	5.0 µL
IGF-1	25 µg/mL	40 ng/mL	8.0 µL
Dexamethasone	$5x10^{-5}$ M	1 µM	0.1 mL
QBSF-60 (221ortu free 221ortug)	-	-	up to 5 mL

- QBSF-60 serum free medium (Fisher Cat# NC0823508)
- Primocin (InvivoGen Cat# ant-pm-2)
- Ascorbic Acid (SIGMA Cat# A4544)
- SCF/c-kit Ligand (R&D Systems Cat# 255-SC)
- IL-3 (R&D Systems Cat# 203-IL)
- EPO (R&D Systems Cat# 287-TC)
- IGF-1 (R&D Systems Cat# 291-G1-01M)
- Dexamethasone (SIGMA Cat# D4902)

-iPSC Medium (feeder-free):

- Complete ReproTeSR™ medium (STEMCELL Tech. Cat# 05920)
-Complete mTeSR1 (STEMCELL Tech. Cat# 05850)

3. Other reagents and specific cell culture material

- Corning TM Matrigel TM hESC-qualified Matrix (Corning Cat# 354277)
- Rock Inhibitor (Reprocell Cat# 04001202; Fisher Cat# NC0454602)
- Polybrene (SIGMA Cat# H9268)
- 12-well plates (Fisher Cat# 07-200-84)
- 6-well plates (Fisher Cat# 07-200-83)
- ReLeSR (StemCell Techn. Cat# 05782)

- Cryovials (Corning Cat# 430487)
- Freezing Container (ex.: Biocision Coolcell; Fisher Cat# 13900856)
-mFreSR™(STEMCELL Tech. Cat# 05855)
- DPBS (Life Tech. Cat# 14190-144)

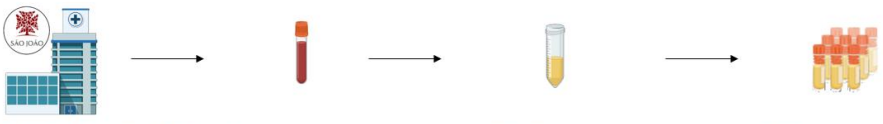
METHODS

Before beginning: Frozen PBMCs collected at CHUSJ as starting material.

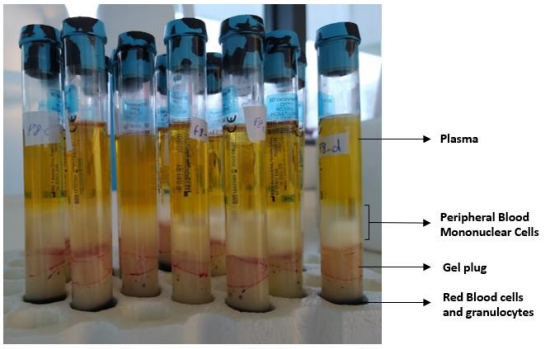
This study was conducted in accordance with the European General Data Protection Regulation. The project was reviewed and approved by the Ethics Committee at Centro Hospital de São João and I3S (Process 469-20 and Ref CHUSJ-AIPD-I&D). All participants provided informed consent, and their confidentiality and rights were strictly protected throughout the study.

Fresh peripheral blood samples were collected from living *CDH1*-carriers, both affected and asymptomatic, as well as from healthy controls (non-carriers), during a visit to Centro Hospital Universitario de São João (CHUSJ). The blood samples were subsequently transported to i3S for the isolation of Peripheral Blood Mononuclear Cells (PBMCs), cell quantification, and freezing in preparation for the reprogramming protocol to be performed at the Center for Regenerative Medicine in Boston (CreM) (Figure 1).

Single-step, standardized method for the isolation of Peripheral Blood Mononuclear Cells (PBMCs):



* PBMCs isolation protocol:



(Representative figure of isolated PBMCs at step 2)

1. Collect a minimum of 4 mL of peripheral blood into two BD Vacutainer Cell Preparation Tubes (CPTs)
2. Centrifuge at 1,800 x g for 30 minutes at room temperature. Ideally, this step should be done within two hours of collection.
3. Collect the mononuclear cells (MCs) by pipetting the buffy coat (cell layer between gel barrier and plasma) into a sterile 15 mL conical centrifuge tube.
4. Bring total volume to 10 mL with sterile DPBS and pipette gently or invert several times to mix.
5. Centrifuge at 300 x g for 15 min at RT and aspirate supernatant.
6. Resuspend cell pellet in 10 mL of sterile DPBS and perform cell count.
7. Centrifuge at 300 x g for 10 min at RT and aspirate supernatant.
8. Resuspend cell pellet in fetal bovine serum (FBS) containing 10% DMSO to freeze (aliquot ~ $2x10^6$ cells/cryovial in 1 mL volume)

Figure 1. Overview of PBMCs collection procedure before iPSCs reprogramming at the CreM

The PBMCs pellet were carefully resuspended in fetal bovine serum (FBS) containing 10% DMSO. Approximately two million cells were aliquoted into each cryovial, with a volume of 1 mL per vial. This constitutes the starting material for the iPSCs reprogramming protocol, and the remaining PBMCs were used for subsequent DNA and RNA extraction.

Progenitor Cells/Erythroblasts expansion

Upon thawing of each PBMCs vial to a 12-well plate well, plated cells are grown during nine days in a 5% CO_2 incubator at 37 °C. During this period, we aim to enrich our culture in $CD34^+$ hematopoietic stem and progenitor cells (HSPCs), the target cells of reprogramming. PBMCs are cultured in a particular expansion media (Table 1) with QBSF-60 used as basal medium combined with erythroid cytokines as growth factors to stimulate the growth of erythroid precursors (erythroblasts). Every three days non-adherent and adherent cells were recovered, centrifuged, and the cells in the supernatant were resuspended in fresh erythroblasts expansion medium and transferred to a new well. Daily morphological analyses confirmed that, after nine days, over 80% of mononuclear cells were progenitors, indicating that cells are ready to be transduced.

Reprogramming of expanded progenitor cells using CytoTune™-iPS 2.0 Sendai Reprogramming Kit

On day 10 in culture, cells are transduced with reprogramming vectors that include the four Yamanaka factors (Oct3/4, Sox2, Klf4 and c-Myc) using the Sendai virus delivery and expression system (Cytotune-iPS 2.0 Sendai Reprogramming Kit). The critical steps at this point are the calculation of volume of each virus needed to reach the target MOI. This is dependent on the live cells available and the titer information on the Certificate of Analysis (CoA) that is lot-specific CoA. The volumes are calculated in the following way:

$$\text{Volume of virus (µL)} = \frac{\text{MOI (CIU/cell)} \times \text{number of cells}}{\text{titer of virus (CIU/mL)} \times 10^{-3}\ \text{(mL/µL)}}$$

For transduction cells are collected, centrifuged and resuspended in fresh expansion medium with polybrene to increase transduction efficiency. Nearly 2x 10^5 cells are plated in one well of a 12 well dish. Each viral vector is thawed and the correspondent volume previously calculated is added to each well, assuring gentle shaking to evenly distribute virus. Another important step is to perform spinnoculation at 2250 RPM at 25°C for 90 min. By the end of the day, our cells will receive an extra 1mL of expansion media and return to plate incubator at 37 °C.

By day 11, we need to prepare Matrigel-coated wells to plate our cells after washing them with QBSF-60 to remove virus. We plate serial dilutions of cells into at least three Matrigel-coated wells (500µL of cell suspension, 300µL of cell suspension and 200 µL of cell suspension) in a final volume of 2mL of expansion media. Each well was supplemented with 1 mL of expansion medium on day 12 post-transduction.Complete ReproTeSR™ medium is added on day 13 and 15 without aspirating original medium. The use of this specific medium during the induction phase of reprogramming facilitates generating human induced pluripotent stem (hiPS) cells from CD34+ and erythroid precursor cells. After day 15 medium has to be changed daily.

iPSCs colonies isolation

Typically, iPSC-like colonies appeared after 15 days post-transduction. We must monitor the cells until hiPS cell colonies appear. Once they are observed, we let them grow for 3 to 4 more days ensuring that colonies borders do not merge with each other.Before colony picking, we have to prepare multiple Matrigel-coated wells of a 12-well plate supplemented with 1m of complete mTeSR1 medium supplemented with 10 µM ROCK inhibitor to avoid cell death. If necessary, we need to carefully scrape away any partially reprogrammed and/or differentiated cells surrounding the iPS cell colony that we want to isolate. We must manually pick large cell colonies and transfer them to the new individual well. Each colony is broken into small fragments by gently and slowly pipetting with a P1000 pipet tip, and its fragments are distributed across the well by rocking the plate and returning it to the incubator. If isolated fragments attached to the well, iPSC colony isolation was successful and ROCK inhibitor should be removed from cell medium after 24 hours by adding fresh complete mTeSR1 medium.

iPSCs maintenance

Typically, six clones per PBMCs sample are picked and the three best clones are further expanded and banked. Human iPSC cultures were maintained on plates coated with Matrigel with fresh mTESR1media daily. All cells were cultured at 37 °C containing 5% CO_2. For the newly reprogrammed iPSCs, cells were washed with DPBS to removed cell debris and passed enzymatically with ReLeSR (Stem Cell Technologies) to larger plates (6-wells) based on colony size for subculturing. At 70-80% confluency, expanded iPSCs clones were cryopreserved using mFreSR™ which is a cryopreservation medium specific for iPSC following manufactures instructions.

RESULTS

Workflow and timeline of iPSCs reprogramming using Peripheral Blood Mononuclear Cells (PBMCs)

In this work, during an 11-month period we used previously collected PBMCs samples and reprogrammed them into induced pluripotent stem cells using a feeder-free system.

Three cell lines were reprogrammed at each time, and this protocol was performed three times to establish a total of nine cell lines. A workflow and timeline of this process is represented in Figure 2, to demonstrate that despite highly efficient this a time and resource demanding set of experiments.

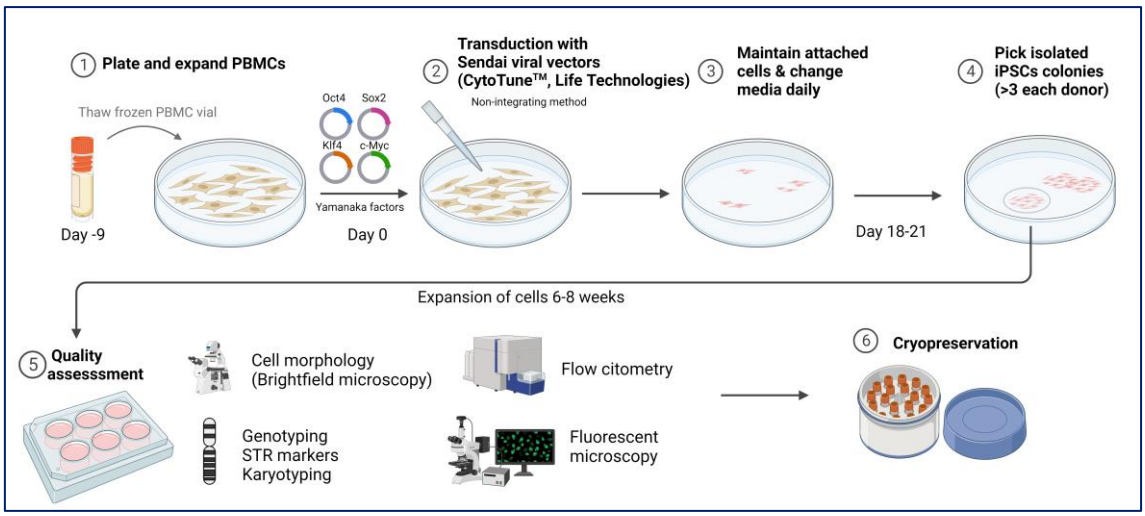
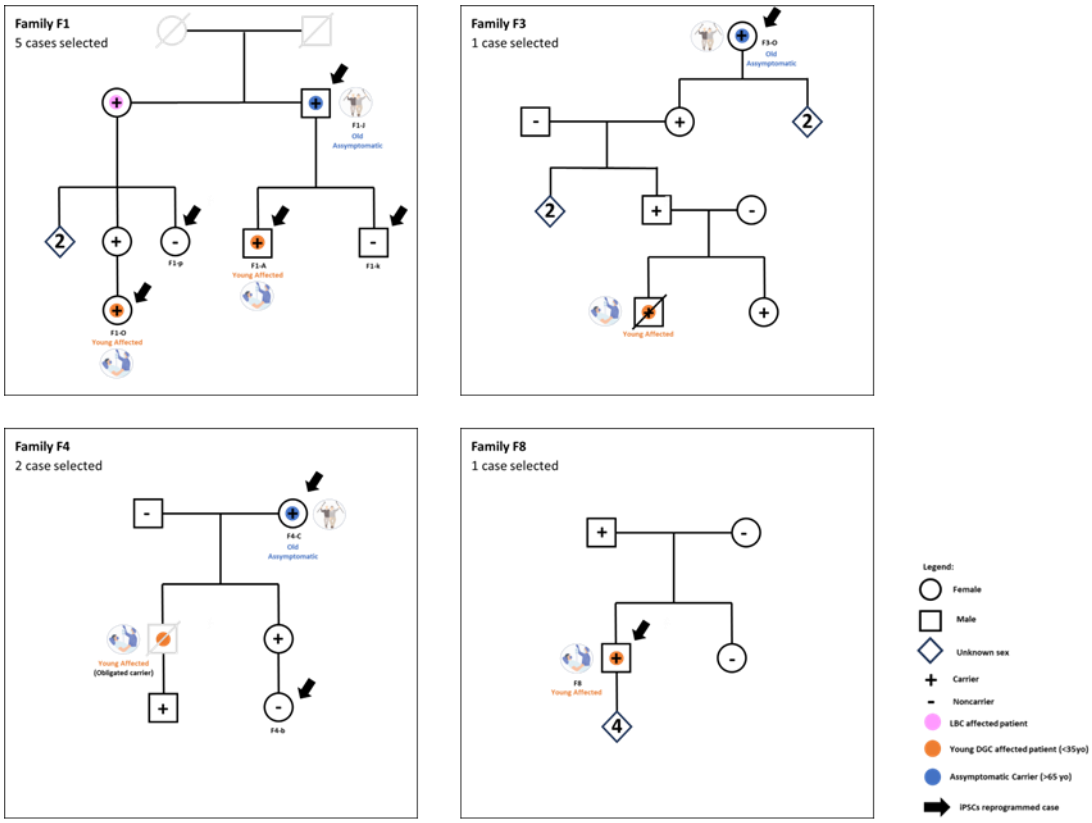


Figure 2. Workflow for iPSCs reprogramming.

iPSCs generation

We successfully generated iPSCs for all cases selected and confirmed that senescent cells or cells obtained from the elderly asymptomatic carriers were not more difficult to reprogram at least when PBMCs are used as the cell of origin. Since we are interested in creating a novel model of study for DGLBC syndrome, we specifically selected three carriers that developed disease at an early age (before 35 years old), three carriers that remain asymptomatic at older ages (above 65 years old) and three noncarriers selected as healthy controls from the same region. These cases belong to four distinct families and detailed information on pedigrees and ages of blood collection and cancer diagnosis are available on Figure 3.



Case Selected	Family	Age at blood collection	Age at DGC Diagnosis	Classification
F1-J	F1	70	NA	Old assymptomatic carrier
F1-A	F1	42	35	Young affected carrier
F1-O	F1	25	22	Young affected carrier
F1-p	F1	47	NA	Noncarrier
F1-k	F1	36	NA	Noncarrier
F3-O	F3	89	NA	Old assymptomatic carrier
F4-C	F4	83	NA	Old assymptomatic carrier
F4-b	F4	23	NA	Noncarrier
F8	F8	33	32	Young affected carrier

Legend: DGC, Diffuse Gastric Cancer; NA, Not Applicable

Figure 3. Pedigrees and detailed information on cases selected for reprogramming.

From the six clones picked for each case, the clone with the best colony morphology was selected for further characterization and validation. The maintenance of healthy colonies requires careful monitoring and handling to expansion of undifferentiated cells over long periods of time. Cultures were examined daily using a light microscope under phase contrast and quality of the cells were inferred based on round colonies shape, a defined border, a phase bright center without areas of differentiation represented in Figure 4.

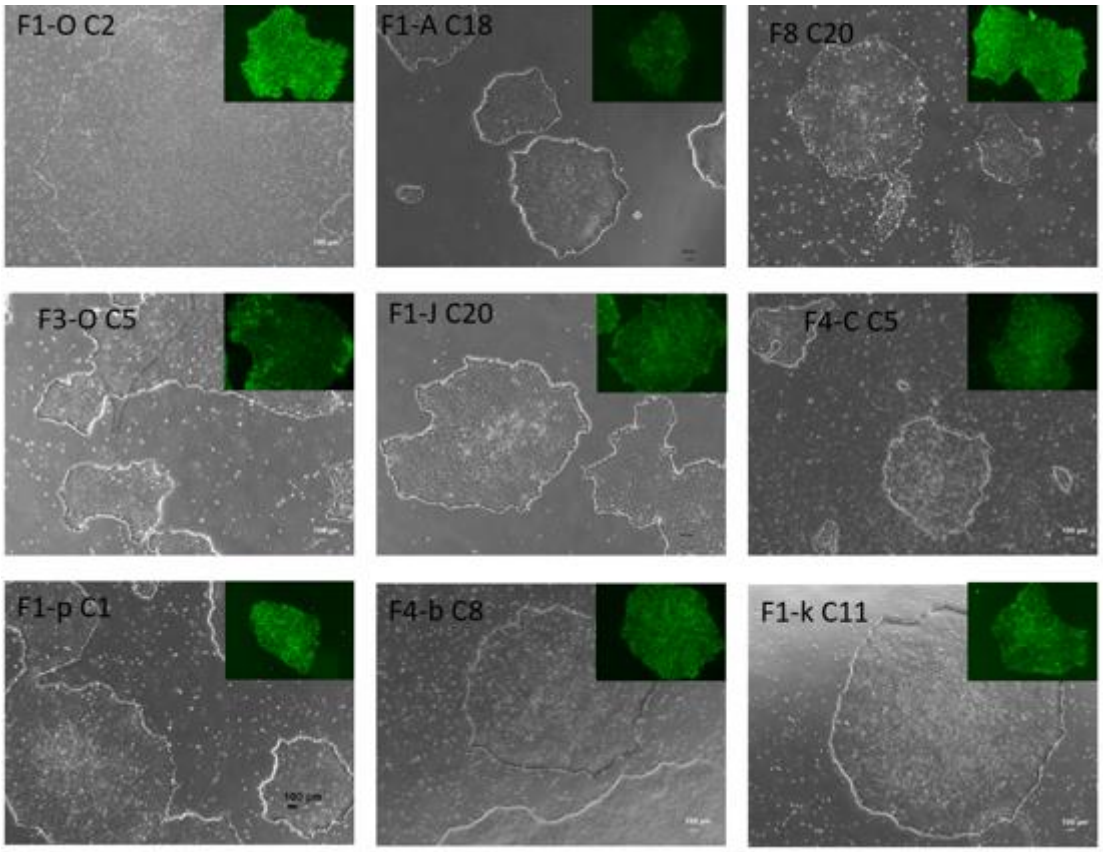


Figure 4. Bright field microscopy images from representative colonies of iPSCS generated for 9 individuals.

The detailed protocol described in this report encompassed the entire process from the expansion of PBMCs, generation of iPSCs by transduction of viral vector and maintenance of iPSCs. We make use of a feeder-free system compatible with good manufacturing practises. The knowledge obtained at the CreM is extremely valuable to bring back to my host institution since there are few researchers working at i3S with iPSCs cells and most only used commercially available cell lines. We are now at a valuable position to collaborate with other research groups and share our expertise. Moreover, iPSC-derived organoids have a unique ability to differentiate into functional different lineages and to 228ortuguese228ee tissues self-renewal and architectural features which we are now at the point of exploring.

CONCLUSION AND FUTURE PERSPECTIVES

In conclusion, our findings support the feasibility of building a biobank iPSC clones from families with rare disease syndromes. This is an important step in the development of novel research and clinical application towards personalized medicine. After my stay at the Center for Regenerative Medicine, I was able to validate the quality of the generated iPSCs by karyotyping, short tandem repeats (STR) profiling, immunofluorescent and flow cytometry staining with stemness markers. In a near future, we will differentiate our iPSCs lines into stomach and breast organoids, which are the target organs of disease in DGLBC syndrome.

REFERENCES

[1] K. Takahashi and S. Yamanaka, “Induction of Pluripotent Stem Cells from Mouse Embryonic and Adult Fibroblast Cultures by Defined Factors,” Cell, vol. 126, no. 4, pp. 663–676, 2006, doi: 10.1016/j.cell.2006.07.024.

[2] V. E. J. M. Palasantzas et al., “iPSC-derived organ-on-a-chip models for personalized human genetics and pharmacogenomics studies,” Trends Genet., vol. 39, no. 4, pp. 268–284, 2023, doi: 10.1016/j.tig.2023.01.002.

[3] M. A. M. Aboul-soud, A. J. Alzahrani, and A. Mahmoud, “Therapies , Disease Modelling and Drug Screening,” Cells, vol. 10, no. 2319, pp. 1–23, 2021.

[4] R. H. Anderson and K. R. Francis, “Modeling rare diseases with induced pluripotent stem cell technology,” Mol. Cell. Probes, vol. 40, pp. 52–59, 2018, doi: 10.1016/j.mcp.2018.01.001.

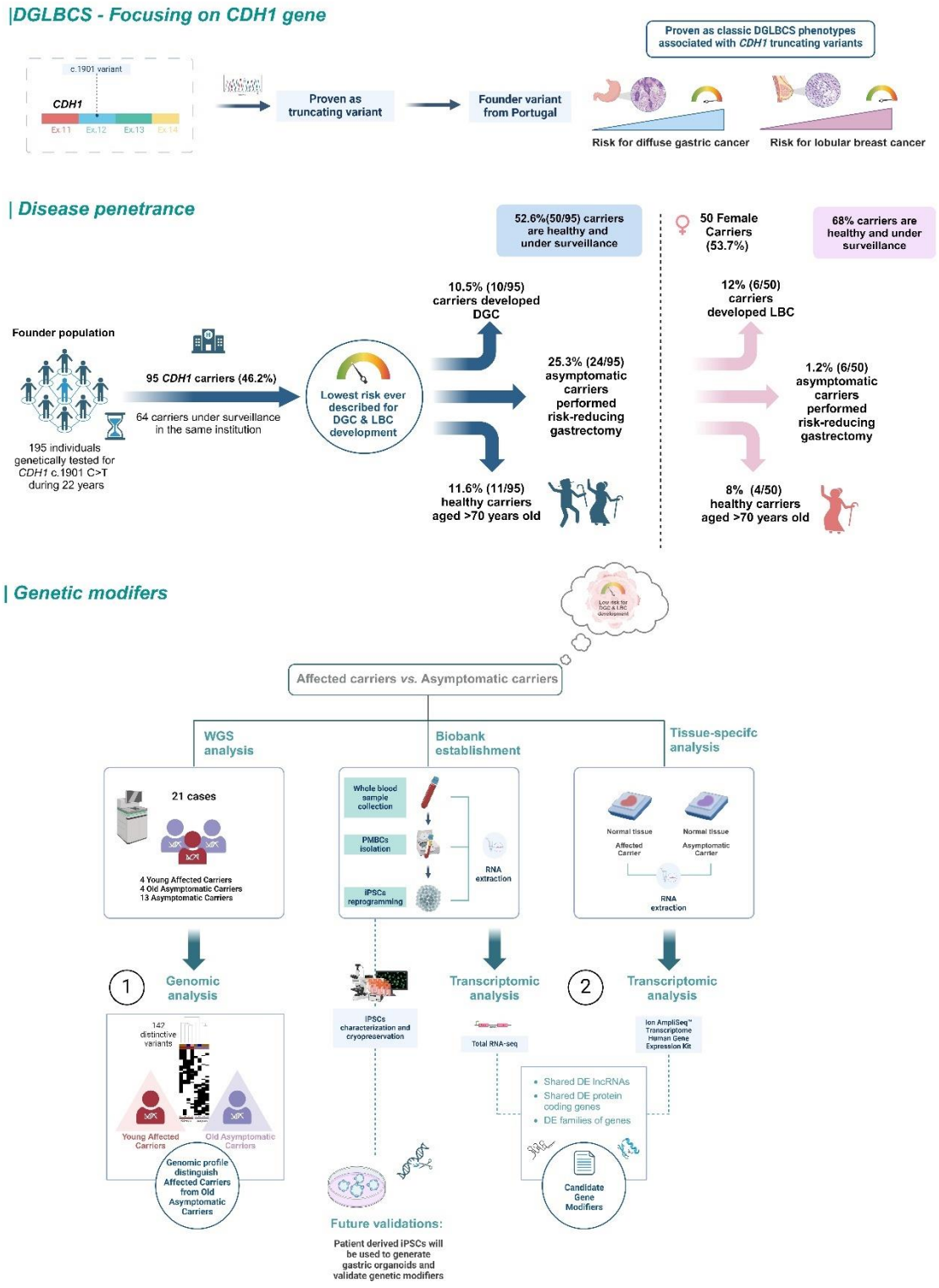
[5] R. Barbosa-Matos et al., “The cdh1 c.1901c>t variant: A founder variant in the 229ortuguese population with severe impact in mrna splicing,” Cancers (Basel)., vol. 13, no. 17, pp. 1–18, 2021, doi: 10.3390/cancers13174464.

[6] G. Scesa, R. Adami, and D. Bottai, “iPSC preparation and epigenetic memory: Does the tissue origin matter?,” Cells, vol. 10, no. 6, 2021, doi: 10.3390/cells10061470.

[7] T. M. Schlaeger et al., “Performed Reprogramming and Cell Culture Experiments,” Nat Biotechnol, vol. 33, no. 1, pp. 58–63, 2015, doi: 10.1038/nbt.3070.A.

[8] T. Okumura et al., “Robust and highly efficient hiPSC generation from patient non-mobilized peripheral blood-derived CD34+ cells using the auto-erasable Sendai virus vector,” Stem Cell Res. Ther., vol. 10, no. 1, pp. 4–6, 2019, doi: 10.1186/s13287-019-1273-2.

Chapter IV|
Discussion & Conclusion



The research presented in this thesis delves into the complex genetic landscape of Diffuse Gastric and Lobular Breast Cancer Syndrome (DGLBCS), aiming to address crucial challenges in disease management and advance our understanding of risk factors and modifiers. The exploration of genomics and transcriptomics is pivotal, as it offers insights into potential causative genes, may identify modifiers of *CDH1* penetrance, and contribute to risk stratification for personalized disease management.

Genetic Landscape and Familial Burden:

The study commences by examining the genetic landscape of families with hereditary GC (Paper 1) with a particular focus on families that pose a challenge due to the absence of a recognized germline causative event. This subset of families, defined as DGLBCS-like, remains burdensome for disease management, making it imperative to revisit studies with overlapping disease spectra. The identification of candidate genes such as *TP53*, *ATR*, *SBDS*, and others becomes important, potentially shedding light on the elusive genomics basis of DGC and LBC development.

Somatic Mutational Profile:

The search for the somatic mutational profile is equally vital, as frequently mutated genes in tumours may also be associated with hereditary forms of the disease. Therefore, in Paper 2, our exploration not only aids in identifying candidate genes for molecularly undiagnosed cases, but also addresses a critical question – which genes could be modifiers of *CDH1* penetrance. Understanding why some individuals remain healthy until old age while others succumb to deadly cancer at a young age becomes central to unravelling disease mechanisms and providing targeted risk management. and the somatic mutational profile of sporadic DGC

The founder effect:

To identify genes causing incomplete penetrance with rare frequency or low and moderate risk, the study emphasizes the importance of a founder population

carrying a truncating *CDH1* variant in Portugal. Our discovery in Paper 3 challenges initial assumptions about missense variants and highlights the impact of variant re-classification on clinical management. Furthermore, the largest genotype-phenotype study in Paper 4 demonstrates the association of *CDH1* truncating variants with DGC and LBC development, prompting a proposal to expand clinical criteria for improved surveillance of affected families.

Cancer Lifetime Risk Assessment:

The thesis extends its investigation into the risk estimations for cancer development in DGLBCS families by exploring a Portuguese cohort in detail in Paper 5. The more controlled genetic and environmental context of 11 families carrying the same *CDH1* variant allows for a unique calculation of disease risk. Contrary to previous reports, the cancer incidence in this cohort does not align with expectations, prompting a reflection on the impact of different *CDH1* variants on disease penetrance.

Exploring Gene Expression Landscape:

The final paper of the thesis explores the genomic and transcriptomic landscape of affected and asymptomatic carriers. Surprising findings, such as the identification of specific variants capable of distinguishing young-affected from old asymptomatic carriers, highlight the intricate interplay between distinct germline SNV and CNVs. Despite the difficulties to integrate our genomics findings with the transcriptomic data, the identification of differentially expressed lncRNAs and protein-coding genes between affected and asymptomatic carriers opens avenues for future research, including the use of iPSCs-derived organoids to validate the candidate modifiers as well to explore therapeutic targets.

Conclusion:

In conclusion, this thesis contributes significantly to the understanding of DGLBCS by unravelling its genetic and somatic landscape, proposing expanded clinical criteria, assessing cancer lifetime risk, and exploring the gene expression landscape of affected and asymptomatic carriers. The establishment of a novel

model with the largest iPSCs library from *CDH1* carriers sets the stage for future research on genetic modifiers and potential therapeutic targets.