2020

U.PORTO

FMUP FACULDADE DE MEDICINA UNIVERSIDADE DO PORTO

E-cadherin and CD44v6 in Gastric Cancer: Role, crosstalk and clinical implications

CARLA LUÍS RIBAS COSTA PEREIRA

TESE DE DOUTORAMENTO APRESENTADA

À FACULDADE DE MEDICINA DA UNIVERSIDADE DO PORTO EM BIOMEDICINA

E-cadherin and CD44v6 in Gastric Cancer:

Role, crosstalk and clinical implications

Doctoral Thesis

Doctoral Programme in Biomedicine

Faculty of Medicine of the University of Porto (FMUP)

2019/2020

Carla Luís Ribas Costa Pereira

HOST INSTITUTIONS

FINANCIAL SUPPORT

This study was supported by Fundação para a Ciência e Tecnologia (FCT) through a PhD grant (SFRH/BD/113031/2015) financed by Programa Operacional and União Europeia.

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

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

SUPERVISOR

Doutora Carla Oliveira, PhD, Affiliated Professor

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

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

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

CO-SUPERVISOR

Doutora Gabriela Almeida, PhD, Affiliated Professor

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

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

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

LIST OF SCIENTIFIC PUBLICATIONS

Ao abrigo do Art. 8º do Decreto-Lei nº 388/70, fazem parte desta dissertação as seguintes publicações:

1. Solorzano L, **Pereira C**, Carneiro F, Almeida R, Almeida GM, Oliveira C, Wälhby C, *Towards automatic protein co-expression quantification in immunohistochemical TMA slides.* IEEE Journal of Biomedical and Health Informatics. (2020) DOI: 10.1109/JBHI.2020.3008821
   Original Paper 1, Chapter 3

2. **Pereira C**, Park J L, Campelos S, Gullo I, Lemos C, Solorzano L, Martins D, Gonçalves G, Leitão D, André A, Wälhby C, Almeida R, Kim WH, Carneiro F, Yang H, Almeida GM, Oliveira C, ***Comparison of East Asian and Western European gastric cancer cohorts explains disparities in survival outcomes and highlights predictive biomarkers for surgery extension in early disease stages.*** *Ready for submission.*
   Original Paper 2, Chapter 3

3. **Pereira C***, Ferreira D*, Mendes N***,*** Granja P, Almeida GM, Oliveira C, ***Expression of CD44v6 containing isoforms influences cisplatin response in gastric cancer cells.*** Cancers. (2020) 12(4), 858. DOI: 10.3390/cancers12040858
   Original Paper 3, Chapter 4

   *These authors contributed equally to this work.

4. Lobo S*, **Pereira C***, Almeida GM, Oliveira C, ***Skipping exon-v6 from CD44v6-containing isoforms influences chemotherapy response and self-renewal capacity of gastric cancer cells.*** Cancers. (2020) 12(9), 2378. DOI: 10.3390/cancers12092378
   Original Paper 4, Chapter 4

   *These authors contributed equally to this work.

5. **Pereira C**, Ferreira M, Martins D, Oliveira P, Carneiro F, Almeida GM, Oliveira C, ***The cast of characters behind E-cadherin dysfunction in gastric cancer.*** Manuscript in preparation.
   Original Paper 5, Chapter 5

COMPLEMENTARY PUBLICATIONS IN RELATED FIELDS

1. Korsak B, Almeida GM, Rocha S, **Pereira C**, Mendes N, Osório H, Pereira PMR, Rodrigues JMM, Schneide RJ, Sarmento B, Tomé JPC, Oliveira C. *Porphyrin modified trastuzumab improves efficacy of HER2 targeted photodynamic therapy of gastric cancer.* International Journal of Cancer. (2017) 141(7):1478-1489. DOI: 10.1002/ijc.30844

2. Kennedy PJ, Sousa F, Ferreira D, **Pereira C**, Nestor M, Oliveira C, Granja PL, Sarmento B. *Fab-conjugated PLGA nanoparticles effectively target cancer cells expressing human CD44v6.* Acta Biomaterialia. (2018) 81:208-218. DOI: 10.1016/j.actbio.2018.09.043

PUBLICATIONS IN CONFERENCE PROCEEDINGS

1. **Pereira C**, Ferreira F, Oliveira P, Carneiro F, Almeida GM, Oliveira C. Proceedings of the 23rd Annual Meeting of the Portuguese Society of Human Genetics, Coimbra, Portugal, 14–16 November 2019. *The prognostic significance of E-cadherin in Gastric Cancer: an integrative approach based on patients' cohort and CRISPR-Cas9 engineered cell models.* Medicine (Baltimore). (2020) 99(9): e19291. DOI: 10.1097/MD.0000000000019291

THESIS COMMITEE

PRESIDENTE

Doutora Maria de Fátima Machado Henriques Carneiro

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

VOGAIS

Doutora Marilia Lopes Cravo

Professora Associada Convidada da Faculdade de Medicina da Universidade de Lisboa (Lisboa, Portugal)

Doutor Peter Jordan

Investigador Principal do Instituto Nacional de Saúde Doutor Ricardo Jorge (Porto, Portugal)

Doctor Arne Östman

Professor at Karolinska Institutet (Stockholm, Sweden)

Doutora Raquel Almeida

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

Doutora Filipa Abreu Gomes de Carvalho

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

Doutora Carla Isabel Gonçalves de Oliveira

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

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

AGRADECIMENTOS / ACKNOWLEDGMENTS

Em primeiro lugar, quero agradecer à Carla Oliveira, uma grande lider de um grande grupo, e a minha orientadora. Obrigada por ter acreditado em mim, mas principalmente, por me ter ensinado a pensar e a compreender as questões científicas de uma forma tão clara e transparente.

À Gabriela, a minha segunda orientadora, obrigada pelos momentos constantes de discussão científica, pela paciência e compreensão. Vocêm sabem, mas preciso de vos dizer, que sem vocês nada disto teria sido possível.

À Sara Rocha, Joana Melo e Tânia Fernandes quero agradecer todos os conselhos, tontarias e risadas. Vocês são a melhores amigas que levo destes quatro anos.

À Sofia Valente, a minha primeira companheira de bancada, obrigada pelos momentos cientificos e não cientificos que partilhamos. Fizeste-me falta.

À Ana Carmo, a menina mais doce e bonita que já conheci, obrigada por me teres escolhido para te ensinar. És uma amiga que levo para a vida.

Ao grupo ERiC, aos membros actuais e aos que por cá passaram, tenho que agradecer todos os momentos de discussão científica, mas também a partilha diária das casualidades da vida.

Ao Daniel Ferreira, Silvana Lobo, Marta Ferreira, Sofia Campelos, Hugo Osório e Diana Martins, obrigada pelo trabalho que desenvolvemos juntos e por estarem sempre disponíveis para ajudar em todas as ocasiões.

Aos nossos colaboradores do Hospital São João, à Gilza Gonçalves, Dina Leitão, D. Conceição, Irene Gullo e Professora Fátima Carneiro, obrigada pela vossa preciosa contribuição.

I also would like to thank Carolina Wälhby for her kindness and scientific input during my visit to her laboratory in Uppsala (Sweden), and to Leslie Solorzano for the enthusiastic discussions, hospitality and friendship.

Aos meus catequistas e catequisandos, obrigada por me terem ajudado a construir um grupo tão bonito e me terem dado o prazer de vos ensinar que podemos genuinamente acreditar, e por em prática, o melhor que a ciência e religião têm.

Ao grupo de jovens, Tozé, Melissa, Tiago, Carolina, Mário, Mariana, José Guilherme, Pedro Mêda, Inês Mêda, Pedro Valente, Mônica, Marco, Estreia e Mimosa, obrigada pela

amizade, mas acima de tudo, por me ter ajudado a crescer e me terem transmitido os melhores valores que tenho.

Ao grupo das Calinadas, Sofia, Joana, Hugo, Diogo, Francisco, Tiago, João, Diana, Antony, Bárbara, Daniela e Maria, obrigada por viverem, tão intensamente como eu, todas as etapas da minha vida. Apesar de já terem passado 10 anos, ainda temos muitos momentos para colecionar.

À Regina, Joana, Neiva, Capelo, Branquinho e Diva, as *girls* com quem fui mais feliz a viajar. Somos todas diferentes e por isso somos tão especiais juntas. Obrigada pela força, o amor, a amizade e diversão que trouxeram à minha vida.

Ao João Pedro, Denise, André, Luís, Marina e José, obrigada pelas noites de descontração e amizade únicas.

À Sofia Neto, obrigada pela confiança que depositamos uma na outra e pela amizade que sei que guardamos para a vida.

À Joana Silva, que me conheçe melhor que eu própria, agradeço-te por todos os momentos de telepatia e pelas conversas sem fim. És a minha segunda irmã.

Aos meus tios e primos, quero simplesmente agradecer os ensinamentos e vivências que partilhamos ao longo da vida.

Aos meus avôs paternos e ao meu tio Jorge, que infelizmente não me vão poder ver a tornar-me “doutora”. Irei sempre guardar-vos vivamente no meu coração.

Aos meus avôs maternos, Joaquim e Augusta, obrigada por me terem criado com tanta alegria e carinho. Vocês são o melhor exemplo que tenho na minha vida, a prova viva de amor.

À minha irmã Catarina, que conheçe todos os meus defeitos e feitios e, que nunca deixou de me apoiar. Este é um grande ano para as duas, mas principalmente para mim, que vou poder ver-te casar. Desejo-te tudo o que desejo a mim, o melhor do mundo.

Ao Francisco, obrigada por te dedicares verdadeiramente a mim todos os dias e por me lembrares constantemente que sou capaz. És o melhor de mim. O futuro está mesmo à nossa frente.

À minha mãe e ao meu pai, que tornaram possivel tudo o que sou hoje. Obrigada por me terem dado asas e por não me deixarem ter medo daquilo que a vida me reserva.

Por fim, à ciência, mais do que o conhecimento, transformou a minha maneira de pensar.

À minha familia e
a todos os que guardo no meu coração,

“The real voyage of discovery
consists not in seeking new landscapes
but in having new eyes”

Marcel Proust

Table of contents

Thesis Outline

This thesis is composed by seven chapters as outlined below. The structure of this thesis reflects the research work featured in five original pieces of work, three of them published.

Chapter 1 introduces the general aspects crucial to the development of this work, by covering the landscape of gastric cancer (GC) disease and treatment and by exploring the existing knowledge on E-cadherin and CD44v6, the central players of this thesis. Last but not least, it gives a taste of the ongoing developments in the new era of digital and computational pathology.

Chapter 2 presents the general and specific aims within the scope of this work.

Chapters 3, 4 and 5 bring into question the role of E-cadherin and CD44v6 in tumor aggressiveness, explores the importance of CD44v6 in GC pathogenesis and therapy response and finally, investigate surrogate markers for E-cadherin loss-of-function (LoF). Within chapters 3, 4 and 5, specific introduction, material and methods, results, discussion, references and appendix sections were included.

In ***Chapter 6***, the transversal findings and clinical implications of this work in the management of GC patients are discussed.

At last, ***Chapter 7*** encloses one additional research work (Original Paper 6) disclosing the prognostic and predictive value of CD44v6 in GC disease (**Appendix A**) and a poster illustrating the preliminary work presented in the Original Paper 5 (**Appendix B**).

Abstract

Despite the significant progress in diagnosis and treatment, gastric cancer (GC) remains a major clinical concern, ranking as the third leading cause of cancer death worldwide. This is partially due to its late detection, which hampers a timely intervention in the majority of patients. A better understanding of molecular alterations determining GC aggressiveness and poor response to therapy will likely contribute to find better ways to deal with this disease.

E-cadherin loss-of-function (LoF) is the most well-established alteration in GC initiation and progression, influencing local invasion, cancer dissemination and metastases establishment. CD44v6 is overexpressed in GC and is correlated with cancer stemness and therapy resistance. Our previous data showed that aggressive inherited CDH1-deficient cancers expressed *de novo* CD44v6 in regions that had lost E-cadherin expression. Herein, we used series of well characterized GC patients, their tumour material, clinicopathological and survival information, combined with cell line models to better understand the clinical implications of E-cadherin and CD44v6 expression dysfunction in GC. For this, we: 1) develop a strategy to characterize the level of co-expression of E-cadherin and CD44v6 in GC tumors; 2) assessed the impact of E-cadherin and CD44v6 co-expression in tumor aggressiveness and patients' outcome; 3) explored the role of CD44v6 expression in the response to therapy; 4) unveiled the influence of exon-v6 itself in modulating the chemotherapeutic response, and; 5) identified surrogate markers of E-cadherin LoF.

Tissue biomarker scoring by pathologists is crucial to study the clinical implications of specific molecular alterations in patients. However, inter-pathologist variability in the interpretation of ambiguous cases, complex expression patterns, and analysis of multiple biomarkers can affect analytical accuracy and overall results. A solution is the development of computational tools to assist pathologist's analysis and register large and complex amounts of data. In the first study of this thesis (**Chapter 3, Original Paper 1**), we established an image analysis pipeline to automatically align immunohistochemical images from consecutive tissue microarray slides and quantify E-cadherin and CD44v6 co-expression. Through cross-slide image registration, we observed that the co-occurrence of E-cadherin LoF and *de novo* CD44v6 expression is very frequent, however the degree of co-expression of both proteins varies widely.

Given the likelihood of this co-occurrence in the same tumor and the myriad of events in which E-cadherin and CD44v6 are involved, we explored the impact of these co-occurring events in GC aggressiveness. In the second study of this thesis (**Chapter 3, Original Paper 2**), we correlated the expression profiles of E-cadherin and CD44v6 with the survival and clinicopathological features of GC patients from two epidemiologically distinct cohorts from Western Europe (WE) (Portugal) and East Asia (EA) (South Korea). We found combined widespread E-cadherin dysfunction and CD44v6 very high expression in ~12% of GC in both cohorts. Such tumors were particularly aggressive in both cohorts. However, only for the WE cohort (WE-C) an association with poor survival was observed, particularly at early stages of disease. The comparison of both series revealed that early stage EA patients' cohort (EA-C) with tumors bearing E-cadherin dysfunction and CD44v6 very high

expression were diagnosed in population GC surveillance screenings, on average 8-years earlier than WE-C patients, and were submitted to more radical surgical procedures with extended lymphadenectomy, when compared to WE-C patients. This likely explains the poorer prognosis in the WE-C. However, early stage tumors from this subset of both cohorts were significantly associated with features of aggressiveness related to invasion and dissemination, as compared to all tumors in each series. EA-C tumors permeated the lymphatic vessels, promoting lymph node (LN) metastasis, while tumors from this WE-C patient subset invaded deeper into the gastric wall, more often permeating the vasculature and nerves. While in the EA-C, the potential consequences of nodal metastization were counteracted by an extensive LN dissection, explaining the impressive outcomes, in the WE-C it would not be simple to neutralize tumor aggressiveness surgically as cancer appears to have spread for more than one route.

Notwithstanding, survival outcomes are not only dependent on how earlier is the tumor detected and whether it can be completely surgically removed, but also whether it responds to chemotherapy (CT). In addition to the aforementioned mechanisms, CD44v6-containing isoforms have been widely associated to drug resistance in several cancers, including GC. In this sense, we aimed at exploring the role of CD44v6 in GC, not only on response to therapy, but also in the transformative process and/or progression of GC (**Chapter 4, Original Paper 3**). To do so, we generated isogenic cell lines by stably expressing a CD44 standard isoform (CD44s) and CD44v6-containing isoform highly expressed in GC cells, and tested the different cancer cell hallmarks and the response to a conventional chemotherapeutic drug, cisplatin. They all presented identical performance in terms of growth rate, invasion, cell-cell aggregation or migration capacity, suggesting that CD44v6 is not a driver of GC development. Interestingly, upon cisplatin treatment, CD44v6+ cells presented lower apoptosis levels compared to CD44v6- cells, indicating that *de novo* expression of CD44v6 in GC cells allows tolerating cisplatin treatment. Driven by these findings, it became important to investigate whether exon-v6 itself is responsible for conferring these properties to CD44v6-containing isoforms (**Chapter 4, Original Paper 4**). We induce specific exon-v6 skipping using a CRISPR/Cas9 and Phosphorodiamidate Morpholino oligomers (PMOs) approach in two GC cell lines endogenously expressing CD44v6. Overall, we demonstrate that exon-v6 specific removal from CD44v isoforms increases cell sensitivity to cisplatin and impairs GC cells self-renewal. We strongly believe that using a PMO approach (a strategy that has been implemented in the clinics) in CD44v6 expressing tumors would lead to tumor sensitization to CT and delayed tumor progression.

Studying E-cadherin LoF is not trivial neither for the pathologist, nor for the computer, due to the difficulty of estimating multiple expression patterns. In the last study of this thesis, we aimed to identify surrogate markers for E-cadherin LoF with potential clinical utility (**Chapter 5, Original Paper 5**). E-cadherin depleted cell models, permanent (CRISPR/Cas9) to mimic the diffuse GC type; and transient (RNAi) to mimic intestinal GC were established and characterized them by label-free quantitative proteomics profiling followed by bioinformatics analysis. We found that permanent E-cadherin LoF dramatically affects cell-cell and cell-matrix adhesion and leads to the overexpression of PEG10 in E-cadherin negative cells. The transient E-cadherin LoF model was associated with a cadherin-switch, from E- to P-cadherin.

Collectively, the work developed under the scope of this thesis has contributed to generate new tools to deal with the analysis of multiple protein biomarkers in GC, revealed a combination of biomarkers that impact early dissemination of GC and may be used to identify patients with poorer outcomes, improved the knowledge on the role of CD44v6 in therapy resistance and GC cells self-renewal, and finally contributed to the identification of surrogate markers for E-cadherin LoF.

Sumário

Apesar do progresso significativo no diagnóstico e tratamento, o cancro gástrico (CG) continua a ser uma grande preocupação clínica, sendo a terceira causa de morte por cancro a nível mundial. Isso deve-se em parte à sua deteção tardia, o que dificulta uma intervenção atempada na maioria dos pacientes. Uma melhor compreensão das alterações moleculares que determinam a agressividade do CG e a fraca resposta à terapia contribuirá provavelmente para encontrar melhores maneiras de lidar com esta doença.

A disfunção da caderina-E é a alteração mais bem estabelecida na iniciação e progressão do CG, influenciando a invasão local, disseminação do cancro e estabelecimento de metástases. O CD44v6 está sobre-expresso no CG e correlaciona-se com a estaminalidade do cancro e a resistência à terapia. Os nossos dados anteriores mostraram que cancros hereditários agressivos com deficiência em CDH1 expressavam CD44v6 *de novo* em regiões que perderam a expressão da caderina-E. Assim, usamos uma série de pacientes de CG bem caracterizados, o seu material tumoral, informações clínico-patológicas e de sobrevida, combinados com modelos de linha celular para melhor compreender as implicações clínicas da disfunção de expressão da caderina-E e do CD44v6 no CG. Para esse efeito: 1) desenvolvemos uma estratégia para caracterizar o nível de co-expressão de caderina-E e CD44v6 em tumores de CG; 2) avaliamos o impacto da co-expressão da caderina-E e CD44v6 na agressividade do tumor e na sobrevida dos pacientes; 3) exploramos o papel da expressão de CD44v6 na resposta à terapia; 4) desvendamos a influência do exão-v6 na modulação da resposta quimioterapêutica e; 5) identificamos marcadores alternativos de disfunção da caderina-E.

A avaliação de biomarcadores nos tecidos por patologistas é crucial para estudar as implicações clínicas de alterações moleculares específicas em pacientes. No entanto, a variabilidade entre patologistas na interpretação de casos ambíguos, os padrões de expressão complexos e análise de múltiplos biomarcadores pode afetar a precisão analítica e os resultados no geral. Uma solução é o desenvolvimento de ferramentas computacionais para auxiliar a análise do patologista e registar grandes e complexas quantidades de dados. No primeiro estudo desta tese (**Capítulo 3, Artigo Original 1**), estabelecemos uma pipeline de análise de imagem para alinhar automaticamente images imuno-histoquímicas de lâminas consecutivas de um microarray de tecidos e quantificar a co-expressão de caderina-E e CD44v6. Através do alinhamento destas imagens, observamos que a co-ocorrência da disfunção de caderina-E e expressão de CD44v6 *de novo* é muito frequente, no entanto, o grau de co-expressão de ambas as proteínas varia amplamente.

Dada a probabilidade dessa co-ocorrência no mesmo tumor e a miríade de eventos em que a caderina-E e o CD44v6 estão envolvidos, exploramos o impacto desses eventos co-ocorrentes na agressividade do CG. No segundo estudo desta tese (**Capítulo 3, Artigo Original 2**), correlacionamos os perfis de expressão da caderina-E e do CD44v6 com a sobrevivência e características clínico-patológicas de pacientes com CG de duas coortes epidemiologicamente distintas da Europa Ocidental (EO) (Portugal) e Ásia Oriental (AO) (Coreia do Sul). Descrobrimos

que 12% dos pacientes com CG apresentavam disfunção generalizada de caderina-E combinada com expressão muito alta de CD44v6 em ambas as coortes. Esses tumores eram particularmente agressivos em ambas as coortes. No entanto, foi observada uma associação com baixa sobrevida apenas para a coorte da EO, particularmente nos estadios iniciais da doença. A comparação de ambas as séries revelou que pacientes do coorte da AO em estadio inicial, com tumores portadores de disfunção de caderina-E e expressão muito alta de CD44v6 foram diagnosticados em testes de screening efectuados à população com CG, em média 8 anos mais cedo do que os pacientes da coorte da EO, e foram submetidos a procedimentos cirúrgicos mais radicais com linfadenectomia extensa em comparação com os pacientes da coorte da EO. Isso provavelmente explica o pior prognóstico na coorte da EO. No entanto, os tumores em estadio inicial deste subconjunto de ambas as coortes, foram significativamente associados a características de agressividade relacionadas com a invasão e disseminação, em comparação com todos os tumores em cada série. Os tumores da coorte da AO permearam os vasos linfáticos, promovendo metástases nos nódulos linfáticos, enquanto os tumores deste subconjunto de pacientes na coorte da EO invadiram mais profundamente a parede gástrica, permeando mais frequentemente a vasculatura e os nervos. Enquanto que na coorte da AO, as potenciais consequências da metastização nodal foram neutralizadas com uma dissecção extensa dos nódulos linfáticos, explicando os resultados impressionantes, na coorte da EO não seria simples neutralizar a agressividade do tumor cirurgicamente, pois aparentemente o cancro propagou-se usando mais do que uma via.

Não obstante, os resultados de sobrevivência não dependem apenas de quão precocemente o tumor é detetado e se é completamente removido cirurgicamente, mas também se responde à quimioterapia (QT). Além dos mecanismos mencionados anteriormente, as isoformas que contém CD44v6 foram amplamente associadas à resistência a fármacos em vários tipos de cancro, incluindo o CG. Nesse sentido, pretendemos explorar o papel do CD44v6 no CG, não apenas na resposta à terapia, mas também no processo transformador e/ou na progressão do CG (**Capítulo 4, Artigo Original 3**). Para isso, geramos linhas celulares isogénicas que expressam de forma estável a isoforma padrão de CD44 (CD44s) e a isoforma que contém CD44v6 mais amplamente expressa em células do CG e testamos as diferentes caracteristicas associados a células cancerígenas e a resposta a um fármaco quimioterapéutico convencional, a cisplatina. Todos as linhas celulares apresentaram desempenhos idênticos em termos de taxa de crescimento, invasão, agregação célula-célula ou capacidade de migração, sugerindo que CD44v6 não é um impulsionador do desenvolvimento de CG. Curiosamente, após o tratamento com cisplatina, as células CD44v6+ apresentaram níveis de apoptose mais baixos em comparação com as células CD44v6-, indicando que a expressão de CD44v6 *de novo* em células de CG permite tolerar o tratamento com cisplatina. Impulsionado por essas descobertas, tornou-se importante investigar se o próprio exão-v6 é responsável por conferir essas propriedades às isoformas que contém CD44v6 (**Capítulo 5, Artigo Original 4**). Para isso, induzimos especificamente o *skipping* do exão-v6, usando uma abordagem baseada no CRISPR/Cas9 e Oligómero Morfolino Fosforodiamidato (OMPs) em duas linhas celulares de CG que expressam endogenamente CD44v6. No geral, demonstramos que a remoção específica do exão-v6 das isoformas CD44v aumenta a sensibilidade das células à cisplatina e

prejudica a autorrenovação das células de CG. Acreditamos fortemente que o uso de uma abordagem OMPs (uma estratégia que foi implementada nas clínicas) em tumores que expressam CD44v6 levaria à sensibilização do tumor à QT e retardaria a progressão tumoral.

O estudo da disfunção da caderina-E não é trivial nem para o patologista, nem para o computador devido à dificuldade de estimar múltiplos padrões de expressão. No último estudo desta tese, pretendemos identificar marcadores substitutos para a disfunção de caderina-E com potencial utilidade clínica (**Capítulo 5, Artigo Original 5**). Foram estabelecidos modelos de células sem expressão de caderina-E, permanentes (CRISPR/Cas9), para imitar o CG do tipo difuso, e transiente (RNAi), para imitar o CG do tipo intestinal, e caracterizados por análise proteómica quantitativa sem rótulo, seguindo para análise bioinformática. Descobrimos que a disfunção permanente da caderina-E afeta dramaticamente a adesão célula-célula e célula-matriz extracellular e leva à sobre-expressão de PEG10 em células negativas para caderina-E. O modelo transitório de disfunção da caderina-E foi associado a uma troca de caderinas, de caderina-E para caderina-P.

Coletivamente, o trabalho desenvolvido no âmbito desta tese contribuiu para gerar novas ferramentas para lidar com a análise de múltiplos biomarcadores proteicos em CG, revelou uma combinação de biomarcadores com impacto na disseminação precoce de CG e que podem ser utilizados para identificar pacientes com piores resultados clinicos, melhorou o conhecimento sobre o papel do CD44v6 na resistência à terapia e na autorrenovação das células de CG e, finalmente, contribuiu para a identificação de marcadores substitutos para a disfunção da caderina-E.

Acronyms and Abbreviations

5-FU - 5-fluorouracil

ACRG - Asian Cancer Research Group

AJCC - American Joint Committee on Cancer

CA - Carbohydrate antigen

CAE - Carcinoembryonic antigen

CCC - Cytoplasmic cell-adhesion complex

CD44 - Cluster of differentiation 44

CNS - Central Nervous System

CP - Cytoplasmic

CRT - Chemoradiotherapy

CT - Chemotherapy

DGC - Diffuse Gastric Cancer

DGCT - Dutch Gastric Cancer Trial

ECM - Extracellular matrix

ECX - Epirubicin, cisplatin and capecitabine

EGFR - Epidermal Growth Factor Receptor

EMR - Endoscopic mucosal resection

EMT - Epithelial-mesenchymal transition

ESD - Endoscopic submucosal dissection

ESD - Endoscopic submucosal dissection

EUS - Endoscopic ultrasound

FGFR2 - Fibroblast Growth Factor Receptor

FIGC - Familial Intestinal Gastric Cancer

FISH - Fluorescence in situ hybridization

GAPPS - Gastric Adenocarcinoma and Proximal Polyposis of the Stomach

GC - Gastric Cancer

GERD - Gastroesophageal Reflux Disease

HA - Hyaluronic Acid

HDGC - Hereditary Diffuse Gastric Cancer

HER2 - Human Epidermal Growth factor 2

HGFR - Hepatocyte Growth Factor Receptor

IGC - Intestinal Gastric Cancer

IHC - Immunohistochemistry

IL - Interleukin

IOC - Intra-operative pathology consultation

LG - Laparoscopic gastrectomy

LN - Lymph Node

LoF - Loss of function

LOH - Loss of Heterozygosity

mAb - Monoclonal antibody

MDR - UK Medical Research Council

mIHC - Multiplex immunohistochemical

MRI - Magnetic resonance imaging

NPs - Nanoparticles

OG - Open gastrectomy

OGJ - Oesophagogastric junction

PDT - Photodynamic therapy

PEG - polyethylene glycol

PET - Positron emission topography

PG - Pepsinogen

PS - Photosensitizer

ROI - Region Of Interest

RT - Radiotherapy

TCF - T cell factor

TCGA - Cancer Genome Atlas

TM - Transmembrane

TNF - Tumour necrosis factor

TOGA - Trastuzumab for Gastric Cancer

VEGF - Vascular endothelial growth factor

VEGFA - Vascular Endothelial Growth Factor A

WHO - World Health Organization

WNT - Wingless signalling pathway

WSI - Whole Slide Imaging

Foreword

Cancer imposes a global health burden, being the second leading cause of death worldwide estimated to affect over 18.1 million of individuals worldwide, of whom, 9.6 million died. As our lifestyle habits boost cancer and global population is growing and aging, cancer burden will continue to increase and the number of individuals affected will rise to 43.8 million over the next 20 years [1]. These facts exert a tremendous physical, emotional and financial strain on individuals, families and health systems. Low and middle-income countries are less prepared to deal with this burden and therefore patients unavoidably suffer with the lack of a timely and high quality diagnosis and treatment [2].

Cancer is not just one disease but a complex group of diseases that can start in almost any tissue of the body. All of them start when the normal cellular processes fail and a rogue cell originates a group of cells that share its abnormal behaviours or capabilities. If these cells are uncontrollably efficient, they grow, surpass their usual boundaries, invade the adjoining parts of the body and disseminate to other organs [3].

According to the latest statistics, lung, breast, colorectum, prostate and stomach occupy the top 5 in the most common cancer types, regardless of gender, although, lung, colorectum, stomach, liver and breast are the deadliest (**Figure 1**) [4].

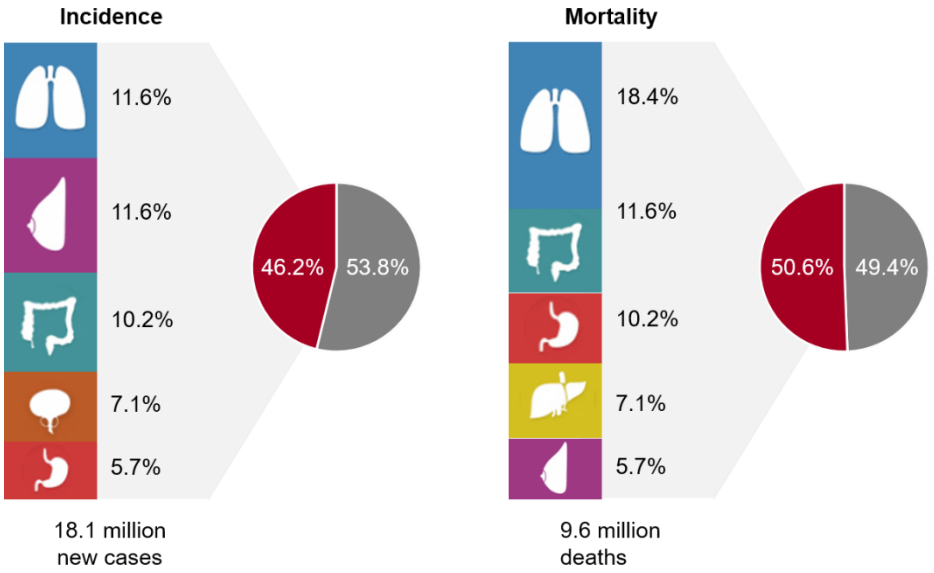


Figure 1. The top 5 cancers with highest incidence and mortality worldwide in 2018. Estimated percentages of new cancers cases and cancer deaths cover all ages, all cancers and both sexes. Image adapted from [5].

Each type of cancer presents different molecular and phenotypic features, which brings up different clinical challenges. It is clear that when cancer is identified early, it is more likely to respond

to a given treatment, increasing the probability of patient survival and decreasing disease and treatment-associated morbidities. However, many cancer patients often experience symptoms late in the development of the disease, delaying the diagnosis at an initial stage and narrowing the treatment options and their efficacy [6]. Such is the case of gastric cancer (GC), to which this thesis is dedicated.

Chapter 1

General Introduction

1. The Clinical and Pathological Aspects of Gastric Cancer Disease

Nearly 90% of GC cases arise in sporadic setting and result of the complex interplay between epigenetic, genetic and environmental factors, without any inherited contribution. The remaining 10% are thought to have familial aggregation, and most of them are not associated with germline mutations. A small proportion of the cases (1-3%) carry an inherited genetic predisposition [2]. The three major hereditary autosomal dominant syndromes affecting the stomach are: Hereditary Diffuse Gastric Cancer (HDGC) [4-6], Familial Intestinal Gastric Cancer (FIGC) [7-9] and Gastric Adenocarcinoma and Proximal Polyposis of the Stomach (GAPPS) [14].

1.1. Epidemiology, Incidence and Risk Factors

GC imposes a global health burden accounting a considerable incidence and mortality worldwide. In 2018, its occurrence was estimated in one million of individuals, of whom 723 100 died [18]. Although the incidence of GC is declining over time, the absolute number of patients with GC tends to increase as the global population and life-expectancy will also grow. In 20-years, the total number of new cases and deaths will possibly duplicate (1 758 810 new cases *vs.* 1 366 121 deaths).

Its epidemiology varies according to geography, ethnicity, pathogenic factors, life style factors and medication use [21].

Despite its wide geographical variation, approximately 60% of all GCs occurs in developing countries. The regions of highest incidence are Eastern Asia, Eastern Europe and South America, while North America and Western Europe represent the lowest incidence zones. Portugal has the highest incidence and mortality rates of Western Europe [23], accounting with 2885 newly diagnosed cases and 2275 deaths in 2018 [24].

Over the last decades, a gradually decline in incidence and mortality rates have been observed worldwide, more remarkably, in North America and Western Europe. This improvement is thought to reflect the higher standards of hygiene, improved food conservation, a high intake of fresh fruits and *Helicobacter pylori* eradication [26].

H. pylori infection is the strongest risk factor associated with several malignancies that arise from the stomach [27]. *H. pylori* infection rates tend to be higher in inferior socio-economic countries, where the GC incidence is also higher [28]. Among the *H. pylori* infected individuals, only a small portion develops GC, which may involve a well-synchronized interaction between the pathogen and the host, *i.e.* strain-specific bacterial factors and host genotypic traits [29, 30].

Some studies have shown the association between *H. pylori* eradication and a reduction of pre-malignant tumors, underlining that the bacteria itself is not carcinogenic but it influences the early stages of disease [31, 32]. Despite the variety of mechanisms whereby *H. pylori* infection may increase the risk of GC, its primary *modus operandi* is through inflammation, a process termed gastritis [33]. The persistence of inflammation (chronic gastritis) may cause atrophy of the gastric mucosa (atrophic gastritis) inducing the first histological changes and consequently stepping into the Correa's cascade [34]. The details on the Correa's cascade in the next subsection.

Recently, a 22-year follow up study assessed the impact of *H. pylori* treatment, vitamin and garlic supplementation in a group of individuals from a high risk rural province of China. Both *H. pylori* eradication and vitamin supplementation yielded a significant reduction in GC incidence and mortality, offering a potential preventive measure to reduce long-term GC risk [35].

Other key factors such as male gender (twice as high), obesity, smoking, atrophic gastritis, partial gastrectomy, Gastroesophageal Reflux Disease (GERD) and Ménétrier's disease has been widely associated with the increased risk of developing GC disease [36].

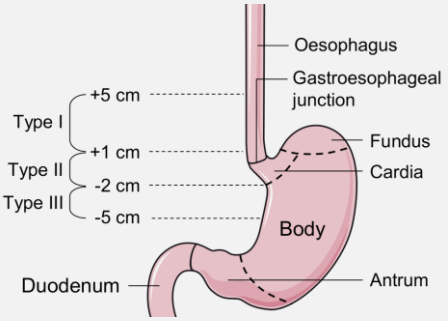
1.2. Histological and Molecular Classification of GC

Almost all GCs are adenocarcinomas (95%); less frequent variants include lymphoproliferative, mesenchymal and neuroendocrine tumors [37].

Although regularly regarded as a single entity, gastric adenocarcinomas can be primarily classified in two subtypes, based on their anatomic location. Cardia cancers arise in the upper cavity of the stomach adjoining the oesophagogastric junction (OGJ), and non-cardia cancers arise from the distal regions (antrum) of the stomach (**Box 1**) [38]. Obesity and GERD are risk factors more associated with cardia type, while *H.pylori* infection is linked with non-cardia type. As a result of *H.pylori* eradication, non-cardia cancers decreased in incidence, as opposed to cardia subtype that increased 7-fold, especially in the developed world [39].

Box 1. The Siewert-Stein Classification of Gastric Adenocarcinomas According to Tumor Location.

Concerning their adjacent location to the stomach, oesophagogastric adenocarcinomas were primarily classified based on the American Joint Committee on Cancer (AJCC) staging system for stomach cancer to promote diagnostic homogeneity [1]. However, many aspects regarding this tumor classification have attracted criticism. In 1980s, Siewert-Stein established a classification system that differentiates tumors arising at the distal esophagus (type I) from those arising at the OGJ (type II) or subcardial (type III) [12].



The latest American Joint Committee on Cancer (AJCC) edition specified that any cancer whose epicentre is > 2 cm distal from the OGJ (type III), even if it involves the junction, should be staged accordingly the TNM-GC system [13]. The Siewert type I/II cancers are staged as oesophageal cancers [19].

The classification of tumors into biologically rational entities is the basis for a proper clinical management of the disease [40]. Many classification systems have been proposed along the years (Laurén [41], World Health Organization (WHO) [21], Ming [42], Mulligan [43], Goseki [44], Carneiro [45]) (**Table 1**). Some are based on the clinical and endoscopic appearance of the tumors, others on histopathology. So far, and particularly when survival is concerned, histopathological classifications have limited significance. Notwithstanding, they are clinically valid and useful when integrated with clinical and epidemiological data [46].

Table 1. Comparison of multiple histopathological classification systems in GC**.** Adapted with permission from Gullo I. *et al.* [40].

Laurén 1965	Carneiro 1995	WHO 2010	Japanese classification 2011	Nakamura 1968	Goseki 1992	Solcia 2009
Intestinal	Glandular	Papillary Tubular	Papillary Tubular 1 (well differentiated) Tubular 2 (moderately differentiated)	Differentiated type	I. Good tubular differentiation, mucin poor	Cohesive, ordinary subtype Cohesive, tubular subtype Cohesive, ordinary subtype
		Mucinous	Mucinous	Undifferentiated type	I. Good tubular differentiation, mucin rich	Mucinous, muconodular Mucinous, infiltrative subtype
Diffuse	Isolated cell	Poorly cohesive, SRC phenotype	SRC carcinoma	Undifferentiated type	IV. Poor tubular differentiation, mucin rich	Diffuse, ordinary subtype
		Poorly cohesive, other cell types	Poorly differentiated, non solid type		III. Poor tubular differentiation, mucin poor	Diffuse, low-grade desmoplasic subtype
Mixed	Mixed	Mixed				
Indeterminate	Solid	Undifferentiated	Poorly differentiated, solid type	Undifferentiated type	III. Poor tubular differentiation, mucin poor	Anaplastic
	Rare variants	Rare variants				High lymphoid response

Although none of proposed systems have reached a desirable consensus, Laurén [41] and WHO [21] classification systems are widely used. The Laurén classification divides the tumors in two main types: intestinal-type GC (IGC), a well-to-moderately differentiated phenotype, and diffuse-type GC (DGC), a poorly differentiated. The third type – mixed - shows a dual pattern of differentiation involving both intestinal and diffuse components [47]. The others GCs that do not fit within these three major types are considered to be of an indeterminate subtype.

Intestinal and diffuse histotypes present distinct clinical-pathological profiles and epidemiologic settings and, more importantly, different molecular pathogenesis (**Table 2**) [48, 49].

Accordingly the Correa's model (1975) [50], IGC evolves in a step-wise manner through the sequential development of histopathological changes, under a chronic inflammatory background usually induced by *H.pylori* infection. Throughout the pathogenic process, multiple host and environmental factors lead to the transformation process from pre-malignant lesions (atrophic gastritis, intestinal metaplasia, dysplasia) to a malignant state, that harbours a number of acquired genetic and epigenetic alterations [51].

In contrast, diffuse-type GC does not follow a predictable order, arising instead from normal gastric mucosa without definitive pre-malignant features. They are characterized by a poorly cohesive and signet ring cells and often negative for *H. pylori* [52]. Therefore, diffuse histotype behaves more aggressively, rendering a worst prognosis than intestinal histotype. Although IGC is decreasing in incidence, diffuse histotype is quite stable or even increasing in incidence, mainly due to a limited diagnosis [53]. Not only the histological subtypes show different trends, but also the localization of the primary tumor has changed over time. There is an increase in the incidence of gastric cardia and OGJ cancer compared to distal GC [54].

Table 2. Differences between Lauren's intestinal and diffuse types. Adapted with permission from Assumpção PP *et al.* [55].

	Intestinal	Diffuse
Incidence trend	Declining	Stable or increasing
Prevalence in endemic areas	Higher	Lower
Environmental influence	Strong	Weak
Age of occurence	Elderly	Young
Genetic factors	Weak	Strong
Male/Female ratio	2:1	1:1
Location of tumor	Distal	Proximal
Macriscopic growth pattern	Growth into the lumen	Tumor spreads along the gastric wall
Carcinogenesis	Well known	Unknown
Prognosis	Better	Worse

A variety of molecular alterations contributes to the tumorigenesis of GC and are part of the complex biology of this disease. GC has several characteristic gene mutations, copy number alterations, epigenetic modification and transcriptional and translational alterations. For instance, E-cadherin deregulation, TP53 mutations and HER2 amplification/overexpression are often observed in GC [56-58]. In order to find GC molecular signatures with clinical significance, many novel GC classifications have been proposed combining different molecular alterations.

Recently, a seminal study by the Cancer Genome Atlas (TCGA) performed an integrative analysis of multiple genomic and proteomic data from GC tumors and established a novel molecular-based classification [59]. TCGA analysis revealed that GCs harbour a plethora of DNA alterations that result in a high number of neoantigens that might designate GCs amenable for novel therapies. The TCGA classification proposed four categories: Epstein-Barr virus (EBV) positive tumors (~9%), tumors with microsatellite instability (MSI) (~22%), tumors with chromosomal instability (CIN) (~50%) and genomically stable (GS) tumors (~20%). Another study from the Asian Cancer Research Group (ACRG) divided the GC into MSI and MSS types; the latter was further portioned into epithelial-mesenchymal transition (EMT), TP53+ and TP53− groups [60].

Despite their relevance, TCGA and ACRG molecular classifications are not yet consensual in the clinical context. In the near future, it would be interesting to bridge the current TNM classification system with the histopathological and molecular data, in order to improve the accuracy and effectiveness of the tailored precision medicine.

2. Current Management of GC Patients

Surgical resection with effective tumor margins clearance and adequate lymph node (LN) dissection is the universal curative treatment for loco-regional disease. Despite that, wide geographic variability in survival outcomes for GC exists worldwide.

Survival outcomes are largely dependent on whether the tumor is detected early on, whether it can be completely surgically removed, and whether it responds to chemotherapy (CT).

The silent and asymptomatic nature of early stages of GC contribute to diagnosis at an advanced stage of the disease, limiting the treatment options. Some of the symptoms experienced by the patients (early satiety (fullness), mild belly pain and fatigue) are similar to the manifestations of other

less serious and common conditions. However, other symptoms, such as, unexplained weight loss, dysphagia, dyspepsia, vomiting and/or iron deficiency anemia are more suggestive of GC and, when experienced for several weeks, should be surveilled by an expert [18].

Given the high incidence rates and considering the impact that early cancer detection in patients' outcome, Asian countries started implementing several changes in many aspects of the GC treatment, such as, screening programs, surgery approaches, chemotherapy, informed consent and clinical trials, that differ from the European methodologies [61].

Massive screening programs began in Japan in 1953 with the work of Toshio Kurokawa, who pioneered the development of many diagnostic tools and techniques used nowadays. Before 1955, less than 5% of GC cases were detected at an early stage, whereas this number sifted to over 50% in the 1980s, upon the screening program implementation [62]. Although the impact of massive screening programs in mortality rates is a question of debate, there is no doubt that this initiative detected a large number of cases that are curable [63]. Nowadays, the approval of a cut-off age for screening of GC has been discussed, similar to colorectal cancer screening. In Asian countries, the incidence of GC increases after 40-years, therefore screening is recommended between 40-45 years of age [64]. However, the appropriate cut-off age depends on the regional incidence of GC, therefore it might be inadequate to define a universal cut-off age for screening.

As such approach is missing in Western European countries with high GC incidence, GC patients are frequently diagnosed at an advanced and unresectable stage, rendering very poor survival rates [5]. As consequence, survival outcomes in high GC incidence European countries lag considerably behind the survival rates observed in East Asian countries [6, 7]. Survival disparities are complex and multifactorial. Some of the contributing factors are: race, ethnicity, sex, screening strategies, surgical management, multimodal therapies, molecular markers, and resource availability [65]. Some of these aspects will be discussed in the next subsections.

2.1. Staging and Risk Assessment

Current clinical guidelines recommend that, whenever GC diagnosis is suspected, patients undergo endoscopy to visually identify cancer lesions. Along with that, surgical biopsy is carried out to extract tissue for diagnosis, histological classification (accordingly WHO classification system) and study of biomarkers, visually reviewed by an experienced pathologist [18].

Of note, the histological criteria for evaluation of tumor biopsies also differs between Eastern and Western pathologists. Eastern pathologists have a lower threshold for diagnosis of primary tumor than Western pathologists, which is based on nuclear and structural criteria even when invasion is absent [66].

The prognosis of GC is closely related to the tumor stage, extent of stomach wall invasion or lesion depth by the tumor (T), spread to the lymph nodes (N) and the presence of distant organ metastases (M). Indeed, survival decreased in a stepwise fashion with increasing pT and pN [67]. The most commonly used staging system for GC was developed by the American Joint Committee on Cancer (AJCC), currently on its 8th edition [68]. This staging system has been extensively used

for clinical research and practice and represents the most relevant tumor-related prognostic factor [69]. The TNM staging system encompasses T, N and M parameters, which are combined to assign an overall stage (**Figure 1**).

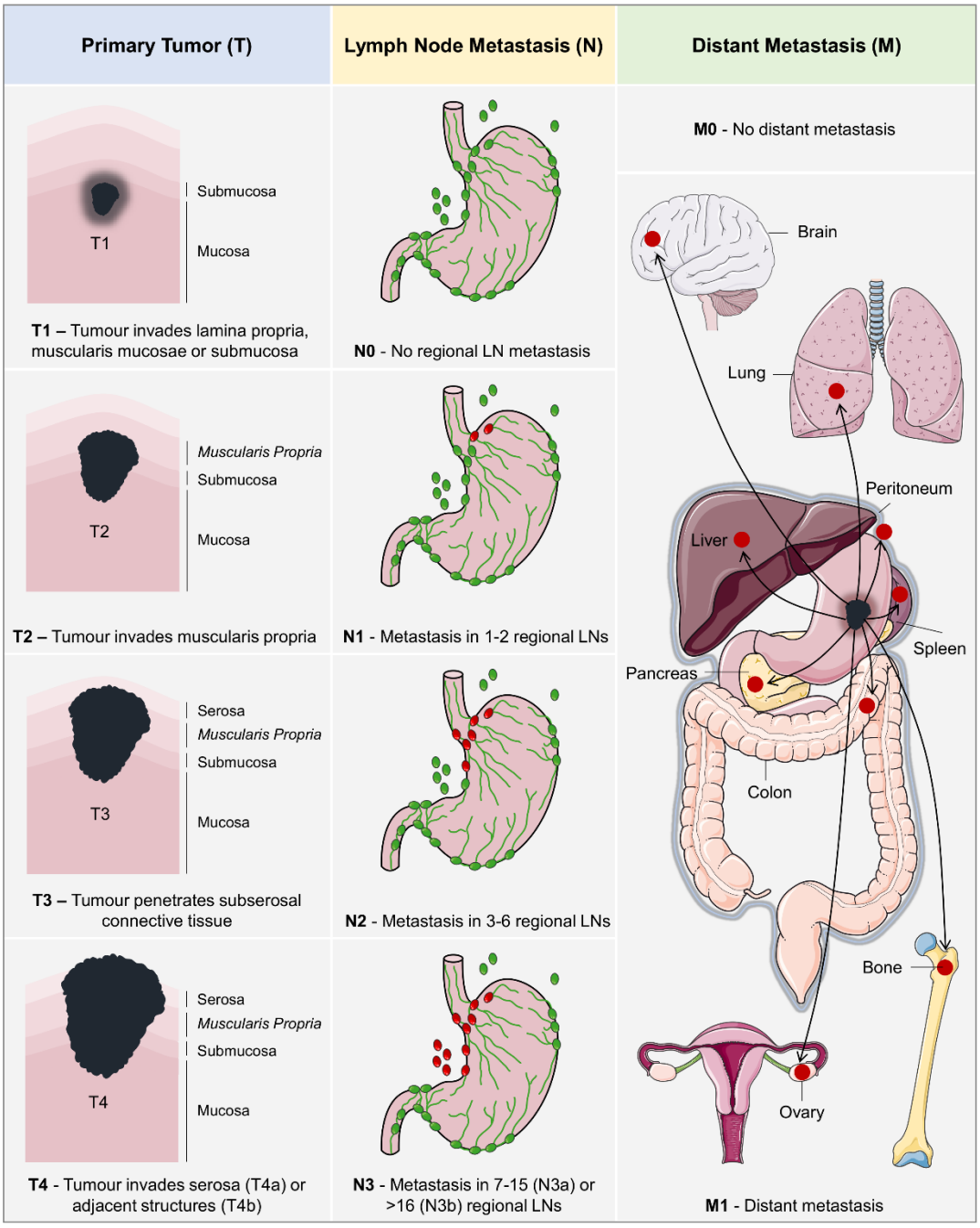
To ensure that diagnosed GC patients are appropriately selected for treatment interventions, tumor staging is essential. The staging of a patient may be defined at several time points of the disease, which represent different classifications - clinical, pathological, post-therapy, at recurrence or autopsy [70]. The clinical and pathological (or surgical) classifications are the most commonly used. The clinical staging is an initial estimative of the extent of cancer and is based on the physical examination, blood tests, imaging tests and tumor biopsies. If the surgery is an option, the pathological stage can be determined based on what was learned from the surgery. It is possible that the pathological stage is different from the clinical stage if the cancer has spread more than predicted. Although the pathological stage combines the clinical staging with the surgery results, the clinical stage might be the only consistent factor in all cancers, since it is independent of the treatment choice.

To determine the TNM stage, several minimal invasive imaging techniques can be employed depending on the clinical requirements [71]. Multidetector computed tomography is the most used technique that allows the identification of malignant lymph nodes based on certain features (short-axis diameter 6-8mm, round shape, central necrosis), however its sensitivity is variable and might leave some doubts [72]. Magnetic resonance imaging (MRI), despite its better soft tissue contrast and direct multiplanar imaging capability, is less preferred than CT due to prolonged scanning time and higher cost.

To the date, endoscopic ultrasound (EUS) is the most accurate and the preferred technique to determine the proximal and distal extent of the tumor. Positron emission topography (PET) is particularly useful to detect metastatic disease, but it has limited accuracy in the detection of peritoneal disease, therefore it may not be very informative in patients with mucinous or diffuse tumors [73]. Altogether, EUS, CT and PET may be best utilized as complementary tools for comprehensive workup of preoperative staging and post-treatment response of GC patients.

Classically, TNM staging system represents the most relevant tumor related prognostic factor to guide the prediction of oncologic outcomes and treatment recommendations.

However, GC disease is a very heterogeneous disease where even similar clinical and pathological settings lead to distinct patients' outcome. This fact underline that staging systems have reached their limit of usefulness and encouraged the assimilation of other clinico-pathological parameters or molecular classifications.



Pathological Staging groups (8th edition)

	N0	N1	N2	N3a	N3b	M1
T1	IA	IB	IIA	IIB	IIIB	IV
T2	IB	IIA	IIB	IIIA	IIIB	IV
T3	IIA	IIB	IIIA	IIIB	IIIC	IV
T4a	IIB	IIIA	IIIA	IIIB	IIIC	IV
T4b	IIIA	IIIB	IIIB	IIIC	IIIC	IV

Figure 1. TNM staging system of GC according to the 8th edition of AJCC. Adapted from [74].

2.2. Treatment Planning of Early Stage GC Patients

After examining the patients' fitness and tumor staging, a treatment planning is organized by a multidisciplinary team including surgeons, oncologists, radiologists and pathologists. The aim of any treatment approach is to accomplish a proper post-operative state, with quality of life and long-term oncological outcomes. [18]

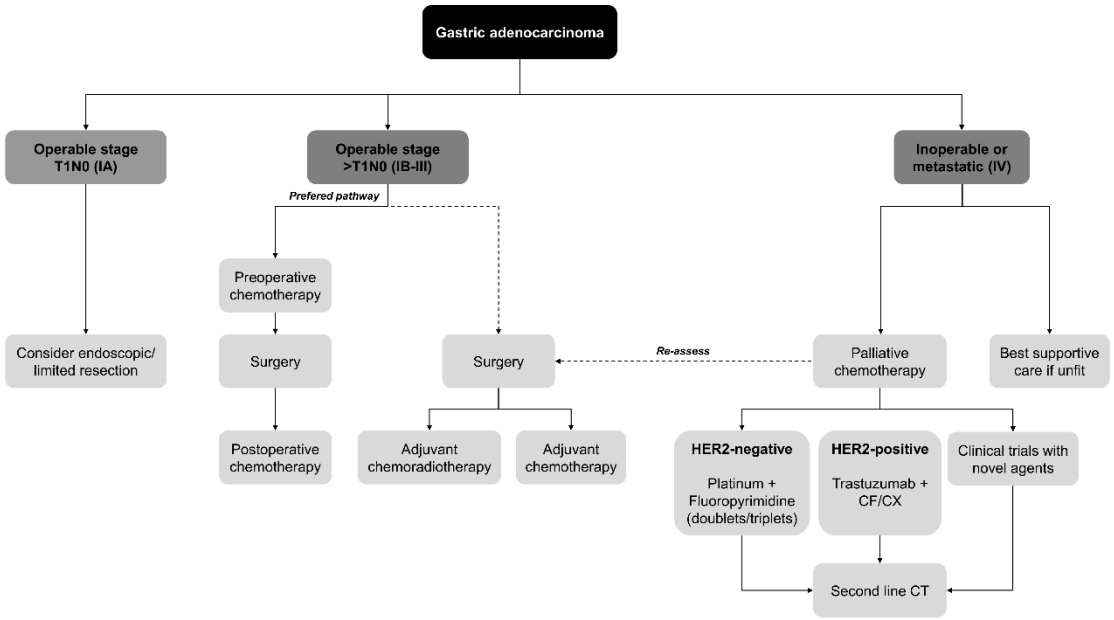


Figure 2. Algorithm of treatment of GC patients. Endoscopic resection is the preferential modality for very early tumors (IA). For stage IB-III tumors, perioperative therapy and surgery is the preferred pathway. For patients without upfront pre-operative due to stage underestimation, post-operative (or adjuvant) CT is recommended. For medically fit patients in advanced stage, doublets or triplets CT regimens are recommended. For medically unfit patients, supportive care is the appropriate alternative. HER2 stands for Human Epidermal Growth Factor 2, CF refers to cisplatin and 5-fluorouracil and CX refers to cisplatin and capacitabine. Adapted with permission from Smyth EC *et al*. [18].

When GC tumors are clearly confined to the mucosa and can be removed by endoscopic resection since the lymph nodes metastatic risk is virtually zero (**Figure 2**). Nowadays, two forms of endoscopic resection are in practice: endoscopic mucosal resection (EMR) and endoscopic submucosal dissection (ESD). Although it requires higher technical skills, ESD can remove larger tumors. Unsatisfied with the low cure rates and high recurrence rates, Asian countries prefer to adopt ESD [75]. Notwithstanding, both modalities are efficient and must be adopted according the tumor depth, tumor diameter, histological grade and ulcerative component [76].

2.2.1. Gastrectomy

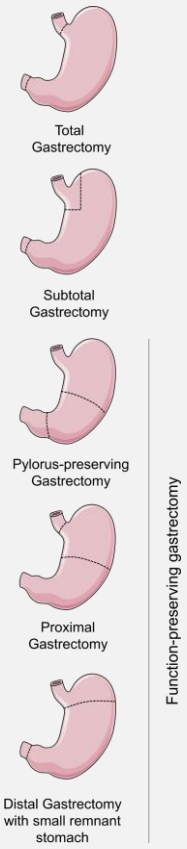
Some patients' tumors do not follow the requirements for endoscopic resection and therefore are treated with a combination of two modalities: surgery, which includes surgical resection

(gastrectomy) with nodal dissection (lymphadenectomy) adjuvant or perioperative chemotherapy (**Figure 2**).

The limits of gastrectomy depend on the size, location and ability of clear the surgical margins of the tumor (**Box 2**) [77]. In turn, the resection boundaries depend on the Laurén histotype. Considering the growth behavior, intestinal-type GC requires a proximal margin of 5 cm, measured in situ, while in the diffuse-type an 8 cm margin is advocated [18]. If it is not possible to accomplish this macroscopic margin, total gastrectomy is performed instead. Some authors claim that the implication of total or subtotal gastrectomy are similar in terms of overall survival and recurrence rates, however subtotal gastrectomy has demonstrated better quality of life [78].

Box 2. Gastrectomy

If the tumor is located enough away from the upper/lower portion of the stomach, **subtotal gastrectomy** can be performed, saving a part of the stomach. If the tumor has spread throughout the stomach, **total gastrectomy** is performed, in which all the stomach is removed, including a portion of the duodenum and oesophagus. There are several advantages in performing more conservative surgery, namely related with lower short-term morbidity and mortality rates, shorter hospital stay, higher calorie intake and better nutritional status. However, many surgeons advocate that total gastrectomy allows a better control of the disease [3]. Notwithstanding, the number of early stage GCs has been increasing and with them, the application of function-preserving GC surgery. Because early stage GC has excellent survival rates after surgical treatment, it has been discussed whether gastric resection could be reduced to preserve the stomach' function including food storage and breakdown, reducing the post-gastrectomy syndromes experienced after conventional radical gastrectomy. Function-preserving laparoscopic procedures include pylorus-preserving, proximal and distal gastrectomy with very small remnant stomach, among others. Although function preserving gastrectomy can improve the quality of life of early stage GC patients, it is crucial to confirm the oncological safety and standardize the techniques [16]. At last, it is important to mentioned that the extent of gastric resection is not a prognostic factor [3].



2.2.2. Lymph Node Dissection

The extension of LN dissection (i.e. lymphadenectomy) plays a decisive role in GC patients' outcome and has been a matter of extensive debated between the European and Asian surgical schools.

It is well documented that LN metastasis is one of the most important determinants of prognosis for both early and advanced GC patients [79, 80]. However, the determination of pN may be directly affected by the total number of harvested LNs. Indeed, several studies have shown that the number of metastatic LNs increases with the number of LNs removed [81-83]. To promote the accuracy of

the pN category, AJCC guidelines recommend the excision of 15 lymph nodes (minimum) for a reliable determination of N status [84].

Recently, another indicator of LN metastasis, namely, the LN ratio (number of metastatic LNs/number of harvested LNs), has been proposed for evaluating GC with curative resection. Indeed, LN ratio demonstrated to be an independent prognostic factor in node-positive GC patients, regardless the number of retrieved LNs [85, 86].

LN dissection can be classified as D1, D1-plus or D2 according to the extent of LNs removed (**Box 3**). D1 lymphadenectomy removes only the nodal groups strictly adjacent to the stomach. D1-plus lymphadenectomy removes the nodes along the coeliac axis (i.e. left gastric artery, the splenic artery and the hepatic artery), in addition to the perigastric LNs. When even more distant lymph nodes are removed (i.e., para-aortic lymph nodes) the LN dissection is named D2 lymphadenectomy [87]. For tumors invading the esophagus, D1-plus includes station 110 (paraesophageal LN in the lower thorax) and D2 includes stations 19 (infradiaphragmatic LN), 20 (LN in the esophageal hiatus of the diaphragm), 110 and 111 (supradiaphragmatic LN).

Asian surgeons believe that GC patients' outcome improves if lymphatic spread is surgically prevented. While D2 lymphadenectomy was routinely performed for early and advanced disease stages in Asian countries, D1 lymphadenectomy was the preferred approach in European countries [18]. The removal of more LNs guarantees not only a more accurate disease staging but also increases the likelihood of removing micrometastasis that may lead to disease recurrence [88].

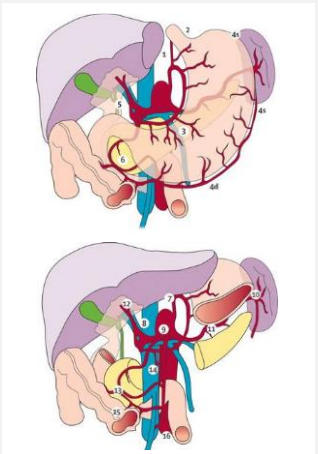
D2 resection was seen with little enthusiasm in the Western countries because many trials, such as Dutch Gastric Cancer Trial (DGCT), UK Medical Research Council (MDR) and Italian Research Council, failed to demonstrate the survival advantage of D2 resection and were associated with higher morbidity and mortality rates. A 15-years follow-up study from Dutch trial demonstrated that D2 resection was associated with lower loco-regional recurrence and GC-related death rates, but significantly higher postoperative mortality, morbidity, and reoperation rates [89].

Notwithstanding, the paradigm of extensive lymphadenectomies has shown signs of change. Western surgeons have increasingly accepted the importance of performing more than a D1 dissection, and Eastern surgeons are accepting that more than a D2 dissection does not always justify the terms of survival and morbidity. Although D1-plus approach is less frequently performed, it stands equidistant from D1 and D2 opposite approaches. For early stage GC not suitable for endoscopic resection or T1 tumors, D1-plus dissection is currently recommended [90].

Box 3. Lymph Node Stations Adjoining the Stomach

Sixteen different LNs stations were identified around the stomach [10]. The regional LNs of the stomach are the perigastric nodes along the lesser and greater curvatures, the nodes along the left gastric, common hepatic, splenic and coelic arteries and the hepaduodenal nodes (stations 1-12). The other LNs (retropancreatic, mesenteric and para-aortic, stations 13-16) may be also eligible in the lymphatic drainage, but they are accounted for distant metastasis (M1) [15].

Although it is hard to predict the pattern of LN metastases, there are LN stations metastases that are more frequently observed depending on the tumor location.



Metastasis is a systemic disease process, thought to be initiated during the early steps of GC progression while the disease is asymptomatic [91]. Once tumour cells reach the submucosa, they may intravasate into blood or lymphatic vessels and spread systemically. Dissemination through the lymphatic vasculature results in LN metastasis, while haematogenous dissemination typically leads to liver and lung metastasis. Alternatively, tumour cells that perform a complete invasion of the gastric wall, may directly invade the organs that are adjacent to the stomach, *i.e.*, spleen, pancreas, colon and liver, or disseminate through the peritoneum [92] (**Figure 1**). Of note, the skeleton is a common metastatic site for visceral carcinomas. Although GC rarely metastasize to the bone, there are some clinical reports describing these cases [93]. The same holds true for central nervous system (CNS) metastasis, which are extremely rare and often associated with metastasis to other organs [94]. Additionally, intraperitoneal dissemination may lead to ovary metastasis (Kruckenberg tumours), in women [95]. In sum, the patterns of metastasis differ notably

Recurrence after curative resection is another factor that negatively impacts patients' survival. The pattern of recurrence is very diverse and may include local-regional recurrence, peritoneum implanting and distant or hematogenous metastasis [96, 97]. In this matter, it is essential to guarantee that the extent of surgical resection is adequate - R0-resection - assuring the macro- and microscopic removal of tumor cells at all the margins [98, 99]. Some hospitals have implemented routine use of intra-operative pathology consultation (IOC) to guide immediate surgical decisions [100]. Frozen sections were considered accurate, sensitive and specific enough to identify a positive margin. On this basis, the surgeon has the opportunity to resect additional tissue to guarantee a R0 resection [101].

Nowadays, minimal invasive approaches, namely laparoscopic gastrectomy (LG), are largely accepted and enthusiastically used as an alternative to open gastrectomy (OG), especially in East Asia (China, Japan and Korea). Compared with OG, LG is associated with less blood loss, fewer surgical complications, less pain, shorter hospitalization periods, better recovery times and less post-operative morbidities. Although some concerns exist regarding the efficacy of LG in performing

lymphadenectomy, a recent meta-analysis suggested that the number of lymph nodes retrieved in LG from D2 lymphadenectomy did not differ from that of OG [102]. However, it remains to be elucidated whether LG can achieve the same results as OG, in patients requiring D2 lymphadenectomy. In any case, if the imaging techniques evolve and improve lymph node prediction, laparoscopy can be validated to operate negative lymph node patients, while positive lymph node patients are operated with OG.

Moreover, the authors observed that the distances of the proximal and distal resection margins did not differ between LG and OG. Even though these parameters envision good clinical prospects for LG-operated patients, the long-term outcomes of LG still need to be validated by two ongoing phase III trials [103, 104].

2.2.3. Chemotherapeutic Choices for Early-Stage GC Patients

Following tumor resection, the patients can still relapse, therefore, GC patients are usually treated with a combination of both modalities, surgery plus perioperative treatment (pre- and post-operative chemotherapy (CT)) (**Figure 3**).

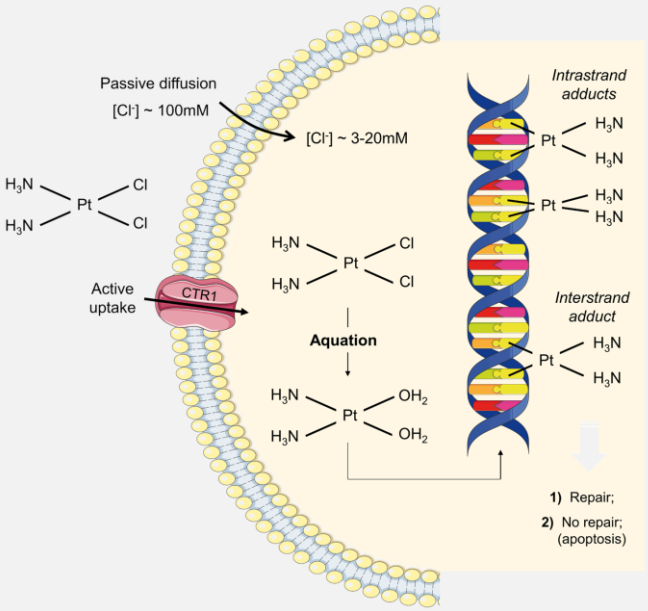
It has been shown that combination chemotherapy comprising a platinum compound and a fluoropyrimidine is, in principle, more efficacious than monotherapy [105].

Cisplatin has been an integral part of GC reference regimens globally (**Box 4**). Due to its side effects, including nephrotoxicity, ototoxicity, and propensity to cause nausea, vomiting or retching, other platinum agents have been studied. Carboplatin exhibited inferior activity in phase II studies and was therefore not studied any further in randomized controlled trials [106]. In contrast, oxiplatin was compared to cisplatin in REAL-2 study, who received epirubicin/cisplatin plus either 5-FU (ECF) or capecitabine (ECX) or epirubicin/oxaliplatin plus either 5-FU (EOF) or capecitabine (EOX). The oxaliplatin–cisplatin comparison proved that oxaliplatin was not inferior to cisplatin, and the former was indeed associated with lower incidences of neutropenia, alopecia, renal toxicity, and thromboembolism, but with slightly higher incidences of diarrhea and neuropathy. Remarkably, capecitabine and oxaliplatin were as effective as 5-FU and cisplatin, respectively.

The REAL2 trial and ML17032 trials tested capecitabine (X) as alternative to infused 5-FU (F) for untreated advanced oesophagogastric cancer The XP doublet regimen was considered as effective as FP regimen for progression-free survival (PFS) in the first-line treatment of AGC, with the clinical benefit of avoiding a catheter [107, 108].

Box 4. Cisplatin Mechanism of Action

Cisplatin is composed by one platinum atom linked to two amides and two chlorides. Cisplatin can enter the cells via passive diffusion or active uptake, mediated by the copper membrane transporter (CTR1) [11]. In the bloodstream, cisplatin is relatively stable and maintains its neutral state, due to the high concentration of chloride ions (~100 mM). In conditions of low chloride concentration, as in the cytosol (~3−20 mM), cisplatin undergoes a process known as aquation, whereby one or two chlorides are replaced by water molecules [17].



Hence, cisplatin becomes highly reactive and covalently binds to DNA on purine bases, forming DNA adducts. Intra-strand or inter-strand cross-linking between adjacent guanine residues is crucial to the cytotoxicity of cisplatin [20]. Also, the greater the number of DNA adducts of cisplatin, the greater the cytotoxic effects seen within the cell. The creation of platinum-DNA adducts interferes with DNA replication and transcription causing cell cycle arrest and activation of pro-apoptotic signals. DNA repair pathways, particularly nucleotide excision repair (NER), is capable of repairing DNA adducts of cisplatin by excising the damaged region and could allow for cell survival [22]. If the damage is to extensive to repair, apoptosis will likely occur [25].

Neoadjuvant CT for resectable GC was rapidly introduced in GC guidelines after the landmark UK MDR MAGIC and FNLCC/FFCD clinical trials demonstrated the benefit of perioperative CT based on platinum/fluoropyrimidine combination (epirubicin (E), cisplatin (C) and 5-FU (F) - ECF) over surgery alone [109]. Due to the criticism raised about the surgical quality, particularly the inadequate lymphadenectomy extension, neoadjuvant CT was not well accepted in EA countries. Factually, both WE and EA world regions continued to use a post-operative/adjuvant chemotherapeutic approach to treat early stage GC patients as the clinical strategies for neoadjuvant CT were only recently grounded.

Since capecitabine is not inferior to 5-FU in advanced disease and avoid the need to intravenous administration, capecitabine containing regimens (epirubicin, cisplatin and capecitabine (ECX)) can also be suggested in preference to ECF in the perioperative setting [108, 109].

Until now, no evidences supported the use of targeted-based therapies in the perioperative setting or in early-stage GC. Recently, NEOHX study is investigating the use a trastuzumab in the perioperative treatment [110], which had previously showed to improve the patient overall survival when added to the CT regimen of HER2-positive metastatic GC patients [111], but no remarkable correlation was achieved. The ongoing EORTC INNOVATION phase II trial is currently recruiting

patients to investigate the efficacy of trastuzumab or the combination of trastuzumab and pertuzumab with standard CT compared with CT alone, before and after surgery [112]. In the future, it is possible that the treatment of early-stage GC patients may include targeted-based strategies. Other details and advances in perioperative clinical trials have been recently and extensively reviewed by Khan and colleagues [113].

More often that desired, the staging of early diagnosed patients is underestimated and the patients are considered for surgery without upfront pre-operative CT. For these patients, post-operative chemoradiotherapy (CRT) or adjuvant CT is recommended.

Radiotherapy (RT) has traditionally been one of the most efficient treatment of cancer provoking the lethal damage of cellular DNA. The benefits of adjuvant CRT in the management of early-staged GC patients was clearly demonstrated in a landmark publication from Intergroup 0116 trial. Surgically resected patients were randomized to undergo merely clinical observation or adjuvant CRT with 5-FU/leucovorin plus fractioned radiation therapy (45Gy in 25 fractions) at the tumor bed and regional lymph nodes. Remarkably, the OS of GC patients who received adjuvant CRT had an improvement of 9 months, compared with patients who underwent surgery alone [114]. A study performed ten years after showed the benefit of adjuvant CRT was persistent [115].

Despite the positive findings, this treatment approach has not gained wide acceptance in Europe mainly due to the toxic effects of CRT approach and the inadequate lymphadenectomy, in which >50% of patients were treated with D0 resection although a more extensive lymphadenectomy was recommended. Given the low lymph node resection, many have argued that CRT was compensating a suboptimal surgery approach. To clarify this shortcoming, a Korean clinical trials, ARTIST, evaluated adjuvant CRT in patients who had undergone R0 resection, which aims resection to cure or complete remission, with D2 lymphadenectomy. Although the use of CRT apparently did not improve the OS and DFS, a posthoc analysis revealed a statistically significant DFS benefit in node-positive patients [116]. The results from a follow-up trial ARTIST2 were recently launched. In sum, adjuvant CT or CRT effectively prolong the DFS, however, no differences were found between both modalities [117]. To date, the addition of post-operative RT to patients who previously received pre-operative CT showed no additional benefit [118].

2.3. Treatment Planning of Advanced-Stage GC Patients

The majority of the GC patients are diagnosed in advanced/unresectable stage and the surgery is no longer a possibility. In these cases, tumor has invaded the adjacent structures to the stomach, possibly affecting more than 15 regional lymph nodes and possibly spreading to other parts of the body. In exceptional cases, the tumor might be resectable but the patient is not fit enough for such surgery [119]. These patients should be guided for palliative CT, which has proven to support better survival and quality of life compared with best supportive care alone (**Figure 5**) [120]. Although the resection is not recommended in the palliative setting, chemo-responsive patients may be deemed operable after careful re-assessed by imaging exams [18].

In late stage GC, the primary aim of CT is palliation, *i.e.* improvement of quality of life. Therefore, the treatment decisions in advanced GC should be concerned with an equilibrated CT benefit and minimization of toxic and side effects, since patients may be frail due to the extensive cancer burden [18].

In the first-line setting, advanced GC patients are treated with doublet or triplet regimens combining platinum (cisplatin or oxaliplatin) plus a fluoropyrimidine (5-FU or capecitabine, an oral fluoropyrimidine) drugs [121].

Although the use of triplet regimens is still debatable, the REAL2 trial demonstrated significant benefit from adding an anthracycline (epirubicin) to a platinum (cisplatin or oxiplatin) and fluoropyrimidine (5-FU or capecitabine) doublet. Moreover, the EOX regimen (epirubicin, oxiplatin, capecitabine) was associated with longer median overall survival (OS) compared with ECF (epirubicin, cisplatin, 5-FU), with the clinical benefit of fewer thromboembolic events and avoiding use of a catheter.

Triplet regimens including taxanes (docetaxel) are also a treatment of choice for first-line chemotherapy [122, 123]. The addition of docetaxel to 5-FU/cisplatin was associated with improved OS, but also added significant toxic effects [123].

For patients with good performance status, a second-line treatment is associated with improved survival rates and quality of life, compared with best supportive care alone. If not used before, the treatment option can include irinotecan, docetaxel or paclitaxel.

In second and third line, treatment options may be used sequentially but there is no clear evidence for a benefit beyond second line treatment [18].

2.3.1. Approved Targeted Therapies: Anti-HER2 and anti-VEGFR2

In addition to the standard regimen, selected patients may be eligible to two targeted therapies: the monoclonal antibodies trastuzumab (first line setting) and ramucinab (second line setting).

Trastuzumab was the first targeted therapy introduced in the treatment plan of advanced GC patients expressing the Human Epidermal Growth factor 2 (HER2), a well-studied tyrosine kinase receptor involved in many cellular functions of tumor development like cell proliferation, differentiation and invasion. HER2 activation is independent of ligand binding, thus, is either constitutively activated or dimerizes with itself or others receptors of the HER2 family. Trastuzumab binds to the extracellular domain of HER2, impairing receptor dimerization and its consequent activation.

Trastuzumab was firstly introduced in the clinical practice as a new targeted therapy for advanced breast cancer patients [124]. A few years later, the focus turned to GC given the high number of cases (10-15%) with HER2 overexpression, mainly cause by gene amplification. In 2010, the TOGA (Trastuzumab for Gastric Cancer) clinical trial randomly assigned GC patients with unresectable/advanced tumors into two groups of treatment: standard chemotherapy (cisplatin and 5-FU doublet) plus trastuzumab and chemotherapy alone. The authors demonstrated that the first group presented a significant increase of 2.7 months in terms of median OS [111]. This benefit was even more accentuated when the tumors were stratified in HER2 high versus HER2 low expression.

Indeed, HER2-positive subgroup improved the median OS in 4.2 months compared with HER2-low patients.

After TOGA trial, HER2 screening was implemented in GC patients with advanced/unresectable disease, by immunohistochemistry (IHC) for HER2 protein overexpression and by fluorescence *in situ* hybridization (FISH) for ERBB2/HER2 gene amplification. Tumors classified as HER2 3+ (strong; complete or basolateral membranous staining; ≥ 10% of the tumor cells) or presenting HER2 2+ (moderate/weak; complete or basolateral membranous staining; ≥ 10% of the cells) and FISH positivity are selected for treatment with trastuzumab-containing regimen, either the combination of capecitabine or 5-FU with cisplatin. [111]

Although not being a GC cancer-specific feature, angiogenesis is one of the main strengths that feeds tumor growth and therefore, one of the cancer hallmarks [125]. Angiogenesis stands as the formation of new blood vessels from pre-existing ones and VEGFR2 is one of its main regulators [126]. This receptor, like the others VEGFR family members, can be activated in two distinct manners: 1) canonical pathway, by VEGFR ligand binding, and 2) non-canonical pathway, by non-VEGF ligand or mechanical forces, such as shear stress [127].

Anti-angiogenic therapies were recently implemented in the GC treatment pipeline with the validation of ramucinab monoclonal antibody, which targets the extracellular domain of VEGFR2 and impairs the binding of ligands. The relevance of ramucinab treatments was proven in two clinical trials: REGARD and RAINBOW. In the REGARD clinical trial, GC patients with inoperable/ advanced after first-line platinum- or fluoripyridine containing therapy were randomly assigned to two treatment groups: ramucinab and placebo. The survival benefit with ramucirunab increased in 1.4 months compared with the placebo group [128].

Simultaneously, the RAINBOW trial tested the benefit of ramucinab in combination with paclitaxel observing a 2.2 months increment in the OS of advanced staged GC patients compared with paclitaxel alone [129]. The results from these two trials culminated in the approval of ramucinab as a second-line therapy combined with standard CT for the treatment of advanced/unresectable GC patients.

Considering the limited survival benefits of targeted therapies, other events may be operating at the tumor level, impairing the response to treatment. Thus, current strategies aim to analyze inter-patient and intra-tumoral heterogeneity, as well as, concomitant alterations in downstream or parallel pathways. Further patient stratification based on molecular heterogeneity may refine therapeutic outline and enhance their predictive value. Apart from that, it is essential to uniform the strategies of management for GC along the community, as much as possible. **Table 3** summarizes the most striking challenges and possible solutions to solve them.

Table 3. Summary of challenges and solutions in GC patients' management [130, 131].

Challenges	Possible solution
1. Control the locoregional relapse after resection;	**1.1.** Implement IOC to guide surgeons' decisions towards R0-resection; **1.2.** Induce R0 resection by performing pre-operative treatment; **1.3.** Improve the efficacy of chemotherapeutic regimens in the perioperative setting, possibly including targeted therapy [112]; **1.4.** Ongoing clinical trials to determine the benefit of combine CT and CRT in a perioperative setting [132, 133];
2. Cure rates following GC surgical resection in West stand behind Japan;	**2.1.** Standardization of D2 dissection for medically fit patients with resectable GC; **2.2.** Divulge and expand EURECCA platform [134], as a mean to spread the best clinical practices and monitor patients' outcomes; **2.3.** Audit the performance of hospitals, specialized equipment and medical teams.
3. Standardize the implementation of neo-adjuvant approaches for early stage disease;	**3.1.** Future implementation of neoadjuvant treatments depends on a more detailed pre-operative local staging;
4. Define the optimal combination of CT regimen (doublets, triplets or sequence);	**4.1.** Perform further clinical trials and define the benefit of third-line treatment;
5. Approval of targeted agents for first-line (trastuzumab) and second-line (ramucimab) treatment is poor;	**5.1.** Identify key mutations that might be useful as novel targeted therapies; **5.2.** Repeat biopsy examination to molecularly profile GC tumors growing after first-line treatment;
6. Individual genetic variation and tumor heterogeneity	**6.1.** Determine the roadmap of molecular alterations and develop appropriate methodological solutions suitable for routine diagnosis.

2.4. Timing Between Diagnosis and Treatment

The timespan between patients' diagnosis and treatment – waiting time – is considered an important indicator of cancer care systems, as it negatively influences patients' quality of life and is associated with unfavorable oncologic outcomes in various types of cancer. Considering the impact of early GC detection in patients' outcome, it is instinctive to ask: (1) what is the average waiting time

since GC patients are diagnosed until they are effectively treated? and (2) can longer waiting times lead to worse patients' survival rates?

In the GC setting, the first study on this topic was conducted in an Asian population and reported that longer waiting times did not adversely affect survival, however this study did not investigate the waiting time between the diagnosis and the beginning of treatment, but the impact of surgical treatment delay [135].

Another study recently conducted in Netherlands demonstrated the waiting time did not influence patients' survival rates, regardless of whether patients underwent primary gastrectomy or pre-operative chemotherapy following gastrectomy [136]. For reference, the maximum recommended waiting time from diagnosis to curative treatment is 5 weeks, in Netherlands. To date, there are no studies reporting the average waiting time in Portugal.

The waiting time between diagnosis and treatment is a measure that depends on other variables, such as, cancer site, time of manifestation and presence of screening programs. In addition, it is important to pay attention to the study methodology, *i.e.* if the exact time frame of the study is measured with reference to the same primary treatment (preoperative chemotherapy, surgery or chemotherapy only). Considering these aspects, the influence of waiting time on other cancer types should be critically and carefully interpreted. Nevertheless, factors such as patient delay and delay of diagnosis are possibly more determinant for oncologic outcomes than the time between diagnosis and treatment [137]. Notwithstanding, the quality of cancer care systems should keep the waiting times to a minimum to favor patients' quality of life.

3. Biomarkers in GC

Notwithstanding the significant progresses in diagnosis and treatment options, GC patients' present discouraging outcomes. Early detection is both beneficial and critical for the successful achievement of good outcomes, as in advanced stages surgery is no longer an option and chemotherapy performs with low efficacy.

Biomarker detection is a relatively non-invasive, convenient and economical method widely used in the clinics. By the words of WHO organization, "A biomarker is any substance, structure or process that can be measured in the body or its products and influence or predict the incidence of outcome or disease." [21]. In other words, biomarkers are objective measures of DNA, RNA or protein molecules that describe a normal biologic or pathogenic processes, or pharmacological response to a therapeutic intervention [138].

The search for cancer biomarkers is performed to identify a specific disease during the early stages of disease (diagnostic), assess on the likely course of the disease (prognostic), evaluate treatment response (predictive) and potentially, identify a target of cancer therapy (therapeutic) [139].

Despite the current progresses in biomarker research, only conventional biomarkers such as carcinoembryonic antigen (CAE), carbohydrate antigen (CA) (CA19-9), HER2 and VEGFA are used in the clinical setting [111, 129].

Notwithstanding, a larger number of single gene marker studies have been published. **Table 4 and 5** categorizes the most recently published up- and downregulated protein-coding genes according to their biological purpose and clinical applicability. Nevertheless, there are a plethora of other candidates described elsewhere (microRNAs, long non-coding RNAs and circular RNA, exossomes, circulating tumor cells and tumor DNA) whose presence/absence also correlates with disease status [140-143].

Table 4. Upregulated protein-coding genes as biomarkers in GC. IHC: Immunohistochemistry; WB: Western Blot; qPCR: Quantitative RT-PCR; TMA: tissue microarray

Gene	Biologic function	Clinical Purpose	Detection Method	Reference
AGR2	Required for MUC2 post-transcriptional synthesis and secretion.	Early detection	Tissue (IHC)	[144]
ANKRD49	May have a role in spermatogenesis	Prognosis	Tissue	[145]
ANOS1	Cell adhesion protein that is a component of the extracellular matrix	Prognostic	Tissue, circulating (IHC, qPCR)	[146]
APLN	Mitogenic factor for the endothelial cell	Early detection	Tissue, Circulating (IHC, ELISA)	[147]
ATM	DNA repair signaling and cell-cycle check point during DNA damage	Predictive	GC cell lines (IHC, qPCR)	[148]
CD44v6	Cell surface adhesion adhesion	Prognostic/Predictive	Tissue (IHC)	[149-151]
CISD2	Mediates mitochondrial integrity and lifespan	Early detection	Tissue (IHC, qPCR)	[152]
CYR6	Cell adhesion signaling	Prognostic	Tissue (IHC, TMA)	[153]
EGFR	Growth factor receptor	Prognosis	Tissue	[127, 154]
FGFR2	Growth factor receptor	Prognosis	Tissue	[129, 155]

GGT	Glutathione homeostasis	Predictive	Tissue, circulating (IHC, enzymatic)	[156]
HER2	Growth factor receptor	Prognosis/ Predictive/ Therapeutic	Tissue	[111]
MET (HGFR)	Growth factor receptor	Prognosis/Predictive	GC cell lines	[157-159]
MET4	Receptor-type tyrosine kinase	Prognostic	Tissue (IHC)	[160]
MMP16	Matrix metalloproteinase	Prognostic	Tissue (qPCR)	[161]
PCDHB9	Glycosylated transmembrane proteins that mediate cell-cell adhesion	Prognostic	Tissue (IHC; qPCR)	[162]
PI3K/Akt/ mTOR	Involved in processes of cellular physiology, such as glucose homeostasis, protein synthesis, cell proliferation, growth, metabolism, survival and angiogenesis	Prognosis/Therapeutic	Tissue	[157, 163]
PTTG3P	Related with cell cycle	Early detection	Tissue (qPCR)	[164]
STMN1	Regulate cellular microtubule dynamics	Predictive	Tissue (IHC, qPCR)	[165]
Trop2	Single-pass transmembrane protein	Early detection	Tissue (IHC, qPCR)	[166]
VEGFA	Growth factor ligand	Therapeutic	Tissue	[126]

Table 5. Downregulated protein-coding genes as biomarkers in GC. IHC: Immunohistochemistry; WB: Western Blot; qPCR: Quantitative RT-PCR; TMA: tissue microarray

Gene	Biologic function	Clinical Purpose	Detection Method	Reference
BAK	Apoptosis-related proteins	Predictive	Tissue (IHC)	[167]
BUB1	Cell proliferation	Prognostic	Tissue (IHC)	[168]
CDH1	Cell adhesion molecule	Diagnostic/Prognostic	Tissue	[169, 170]
FAM46C	Regulation of translation by acting as an mRNA stabilizing factor	Predictor of recurrence	Tissue (qPCR)	[171]
MGMT	A DNA repairer capable of removing methyl groups and larger adducts	Predictive	Tissue (IHC)	[172]
PRMT1	Tumor suppressor and promoter	Predictive	Tissue (IHC)	[173]
PROX1	Tumor suppressor gene or an oncogene	Prognosis	Tissue (IHC)	[174]
RBMS3	Regulator of DNA replication, gene transcription, cell cycle progression, and apoptosis	Early detection	Tissue (IHC, qPCR, WB,TMA)	[175]
S100A12	Regulation of calcium homeostasis	Early detection	Tissue (IHC)	[176]
SAMSN1	Mediator of B-cell function	Prognostic	Tissue (IHC, qPCR)	[177]
SASH1	Intracellular signal transduction	Prognostic	Tissue (IHC)	[178]

3.1. E-cadherin in GC – Old but Gold

Among the plethora of genetic mutations, epigenetic alterations and aberrant molecular signaling pathways known to be involved in GC development, E-cadherin is the most well established event involved in GC initiation and progression, involved in both sporadic and hereditary forms of the disease [169, 179].

E-cadherin, encoded by CDH1 gene, is a type I transmembrane glycoprotein mainly expressed in epithelial cells of mammalians that mediates cell-to-cell adhesion in a calcium-dependent manner, establishing homophilic interactions with adjoining E-cadherin molecules in the neighbor cell [180]. In addition, it functions as signaling molecule, blocking proliferative signaling by growth factors, thereby playing an important role in cellular growth, proliferation and differentiation and therefore gained the special title of growth and invasion suppressor [181].

E-cadherin mature proteins are structurally composed by five extracellular domains, a single transmembrane domain and a cytoplasmic domain (**Figure 3**). On the other side of the cell membrane, a myriad of molecules interacts with the intracellular domain to drive processes related with actin cytoskeleton, cell signaling and trafficking. The cytoplasmic tail of E-cadherin conserves specific binding sites for its partners, p120 or and β-catenin, which have independent anchorage sites. These two proteins, together with α-catenin that binds directly to β-catenin and links the actin cytoskeleton, form the cytoplasmic cell-adhesion complex (CCC) [182]. Binding of p120 is essential in normal epithelial cells, since this protein is responsible for the direct and indirect stabilization of E-cadherin at the cell membrane [183]. Moreover, p120 has a well-established role in regulating members of the Rho family of small GTPases, which are key mediators of cytoskeletal dynamics and cadherin-mediated cell-cell adhesion [184].

In addition to its functions in the CCC, both β- and α-catenin have roles in the cytoplasm and nucleus. The interaction between E-cadherin and β-catenin at the adherens junction provides one obvious mechanism by which CDH1 mutations could disrupt growth signaling and initiate tumorigenesis. β-catenin is a transcription co-factor in the Wnt signaling pathway. Because E-cadherin sequesters β-catenin into the adhesion complex, downregulation of E-cadherin is likely to result in increased levels of free β-catenin and transcription of the targets of Wnt signaling. In the absence of a Wnt signal, β-catenin is normally kept at very low levels within the cell by a destruction complex (GSK-3, APC, and axin), responsible by degrading β-catenin. When a Wnt ligand binds to a Wnt receptor on the cell surface, the destruction complex is dissociated and β-catenin is no longer degraded. [185] Instead, it builds up in the cytoplasm of the cell, binds to T cell factor (TCF), and translocates into the nucleus, initiating the transcription of genes that promote growth, including c-myc [186]. For instance, APC mutations are associated with the initiation of colorectal cancer, as the product of this gene normally targets b-catenin for degradation [187].

α-catenin is required for reinforcing cadherin-mediated cell–cell adhesion [188]. It comprises three domains (N-terminal, middle and C-terminal domain), each of which interacts with a unique set of proteins. For example, the N-terminal domain serves to bind to β-catenin, the middle domain binds vinculin, and the C-terminal domain binds fibrous actin (F-actin) [189].

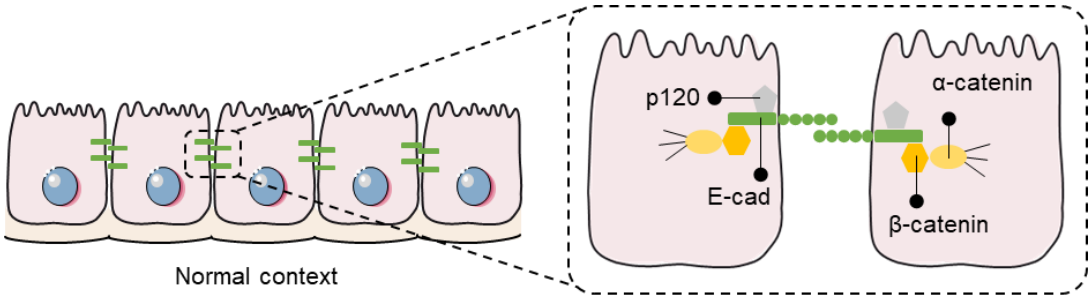


Figure 3. Schematic structure of E-cadherin and its binding to catenin partners. Reproduced with permission from Liu X. and Chu KM [13].

3.1.1. Clinical Significance of E-cadherin in GC

Given the aforementioned evidences, it is not surprising the abrogation of E-cadherin functions disturbs the intercellular adhesion and stromal interactions, altering the migratory and invasion abilities potentiating the oncogenic capabilities [190].

Intragenic mutations, large CDH1 locus deletions (loss of heterozygosity, LOH) and CDH1 promoter methylation represent mechanisms inactivating E-cadherin in hereditary [4–8] and sporadic GC with diffuse histology [9–13].

CDH1 mutations occur in a minor proportion of advanced carcinomas. Inherited germline mutations in CDH1 are a causative feature of HDGC, awarding E-cadherin as the culprit for the development of this cancer syndrome [191]. HDGC tumors appear when complete somatic CDH1 inactivation is acquired, leading to reduced or absent E-cadherin expression. This occurs through second hit mechanisms, pursuing the Knudson's model of tumor suppressor gene inactivation [192]. CDH1 promoter hypermethylation is the most frequent second hit inactivation mechanism in HDGC primary tumors, while a second mutation or deletion (loss of heterozygosity [LOH]/intragenic deletions) was less frequently identified [193].

Although these classical CDH1 inactivation mechanisms have been associated with hereditary and sporadic GCs with a diffuse component, we observed that the overall frequency of CDH1 alterations were equivalently present in both intestinal (~30%) and diffuse GC (~40%) [169]. In the same study, we also observed that a major subset of GC patients lacking CDH1 alteration display aberrant or absent E-cadherin protein expression reported in both intestinal and diffuse type. We investigated other mechanisms at the transcriptional/post-transcriptional level that could explain E-cadherin expression abnormalities [194]. Our group recently dissected an alternative non-classical pathway that causes E-cadherin dysfunction via loss of microRNA-101. This pathway occurs preferentially in IGCs retaining CDH1 WT allele(s) untargeted by classical CDH1 inactivation, but present great disturbance of E-cadherin expression and function [179].

Aberrant E-cadherin expression presents a strong correlation with tumor grade, depth of invasion and regional lymph node involvement [195, 196], being extensively correlated with poor outcome in GC [170, 197, 198]. Despite that, some studies do not support this evidence [199], bringing into question the usefulness of E-cadherin expression as an independent prognostic marker. One of the

factors contributing to this conflicting results may arise from the difficulty on quantifying E-cadherin imunohistochemical expression, which can assume a variety of patterns (null, dotted, cytoplasmic, incomplete membrane, complete membrane) [200], thus obscuring the analytical path and contributing to the inconsistencies that lead to erroneous prognostic indications. Given this, there is a huge interest in exploring the landscape of molecules triggered upon E-cadherin LoF in both diffuse and intestinal-type settings and explore their potential clinical utility.

3.2. CD44 – A Molecule of Many Faces

CD44 is a transmembrane glycoprotein encoded by a single gene comprised by 19 exons in human and 20 exons in mice [201]. This gene is constituted by a constant portion of seven extracellular domains (S1-S7), a transmembrane (TM) domain (S8) and a cytoplasmic (CP) domain, only encoded by exon S10 in humans and by S9 and S10 in mice [202]. This constant portion constitutes for itself the smallest CD44 isoform termed CD44 standard (CD44s), ubiquitously expressed at the cell membrane of most vertebrate cells [203]. Between domains S5-S6, up to ten variant exon products can be inserted by alternative splicing, resulting in many different protein transcripts that share the same TM and CP domain but slightly differ in the extracellular portion (**Figure 4A**) [204]. CD44 canonical protein presents a nominal molecular weight between 85-90 kDa, but it can increase substantially depending on the variant products inserted by alternative splicing and the post-translational modifications (**Figure 4B**) [205].

CD44 is the major receptor for hyaluronic acid (HA), a major component of the ECM, that binds to specific docking sites in the CD44 extracellular globular structure [206]. In addition, CD44 contains other binding sites for collagen [207], laminin, fibronectin [208], osteopontin [209] and other cell surface receptors, modulating a multitude of cell-matrix interactions. Post-translational modifications, specifically O-glycosylation, and the cytoplasmic tail of CD44 can directly affect the localization of CD44 at the membrane and influence the interaction with HA [210]. The cytoplasmic tail expands the range of functions mediated by CD44 by binding to cytoskeleton partners' proteins to ERM proteins (ezrin, radixin and moesin) and ankyrin [211, 212].

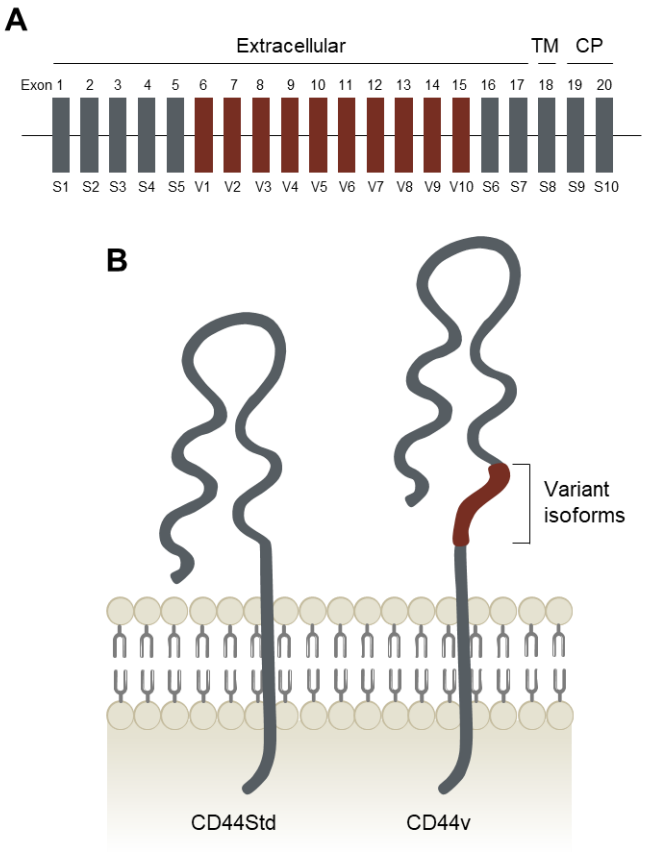


Figure 4. CD44 gene and protein structure. (A) In total, CD44 gene can comprise up to 20 exons, in which 10 belong to the constant portion (S1-S10, green) and constitute the shortest CD44 standard (CD44s) isoform. Up to 10 variant exons (V1-V10, red) can be inserted, in various combinations, in the proximal extracellular domain by alternative splicing to constitute CD44v isoform. **(B)** CD44 protein is composed by seven extracellular domains**,** within which the exon variant products (red) can be inserted, a transmembrane region (TM) and a cytoplasmic tail (CP). Post-translational modifications owing to N- and O-glycosylation are highlighted in brown and orange circles, respectively. Adapted with permission from Zoller M. [205].

In 1983, CD44 glycoprotein was primarily described as a lymphocyte homing receptor [213], however, since then, many other functions have been described. The major physiological role of CD44 is the maintenance of tissue integrity through cell-cell and cell-matrix interactions, however, it is also involved in other key cellular processes, such as lymphocyte activation, recirculation and homing [214, 215]. Nowadays, there is a strong evidence for the importance of CD44 in the progression of many cancer types, as well as, for its expression on cancer stem cells. At the present, no compelling evidence proved the role of CD44 in self-renewal and pluripotency, but is thought to contribute for the activation of stem cell regulatory genes [216].

Unlike CD44s, several variant isoforms were discovered to be only expressed in some epithelial cells of a few tumor types [203], such as lung [217], breast, ovarian, pancreatic, endometrium [218, 219], colon [220], and GC [221]. Among CD44v isoforms, CD44v6 is one of the most widely investigated due to the additional biding sites to HGF and vascular endothelial growth factor (VEGF) [222, 223].

In 1991, Gunthert and his colleagues reported, for the first time, the clinical relevance of CD44v6 describing that CD44v6 expression induced the metastatic phenotype of CD44 negative rat pancreatic cancer cells [224].

Few years later, Heider *et al.* was the first to explore CD44v6 in the context of GC, reporting the presence of CD44v6 expression in pre-cancerous tissue and suggesting its potential as a marker to distinguish intestinal and diffuse-type GC [225]. More recently, Cunha *et al.* postulated that CD44v6 isoform is *de novo* expressed from pre-malignant to malignant lesions in more than 60% of GC cases, either in the sporadic or the hereditary context [221]. Considering the potential role of CD44v6 in GC initiation and progression, new question arises on the relevance of CD44v6 expression in patients' outcome and treatment response.

3.2.1. Prognostic and Therapeutic Potential of CD44v6 in GC

Increased CD44v6 expression has been reported in LN metastasis of colorectal cancer and linked to a worst prognosis [226-229]. Under the scope of GC, a meta-analysis study showed that positive CD44 expression was significantly associated with lower OS and worse GC patients' prognosis [151]. Two subsequent studies showed a positive association between increased CD44v6 expression and worst OS in GC patients in higher TNM stages [149, 230]. This aspect raises some skepticism on whether CD44v6 is an independent factor of prognosis.

Considering the frequent and homogenous expression of CD44v6 in squamous carcinoma, some clinical trial have investigated the therapeutic potential of CD44v6 [202, 231, 232]. The first clinical attempt was conducted in patients with squamous cell carcinoma of the head and neck or esophagus. In this clinical trial, an antibody-drug conjugate of bivatuzumab, a humanized mAb against CD44v6 coupled with mertansine, the cytotoxic microtubule inhibitor, were tested [233]. However, some of the patients developed extensive skin reactions and, unfortunately, one the patients suffered from epidermal necrolysis and died. In fact, although the expression of CD44v6, and other CD44 variants, is mostly restricted to malignant cells, in the normal context, the one and only exception refers to epidermis and secretory glands in endometrium. Due to the observed skin toxicity, the development of this conjugate was discontinued, yet new CD44v6-targetted imaging diagnostic approaches are being considered [234].

All the previous evidences on CD44v6 landscape regard it as a prime candidate for targeted delivery or a treatment agent. In a collaborative work, our group developed poly(lactic-co-glycolic acid) (PLGA)-based nanoparticles (NPs) modified with polyethylene glycol (PEG) and conjugated with CD44v6 antibody [235]. CD44v6-conjugated NPs specifically bound to GC cell lines expressing CD44v6 and therefore may be eligible as a potential anti-cancer agent to use in $CD44v6^{+}$ stratified carcinomas.

4. Next-generation pathology

The analysis of protein expression in tumor sections by a pathologist is one of the main approaches to study the clinical implications of specific molecular alterations.

Pathologists are frequently tasked with the analysis of multiple slides taken consecutively from a block. Each slide is stained to highlight particular features of the tissue or to study a protein of interest. As each section is microscopically distanced (3-5 micrometers) from the neighboring section, the expression profiles of each protein can be compared directly. Cross-slide analysis is useful in many ways. A marker, of any type, conjugated with another, can highlight a specific regions or subset. This is particularly important, considering that the increasingly genetic diversity of tumors and consequent heterogeneity complicates the understanding and treatment of cancer. However, the visual examination performed manually by the pathologist can be subjective, slow and inaccurate, and difficult to assess in large image stacks, often resulting in significant inter- and intra-observer variability.

Capturing the image is just the beginning of a process that has become essential in the diagnostic routine of a pathologist [236]. Little by little, whole slide scanners are revolutionizing manual histopathology, rapidly converting tissue slides into digital images that contain the morphological and staining information necessary to auxiliary diagnosis, grading and therapeutic decisions [237]. This is the new domain of pathology, so-called digital pathology [238]. Whole slide images (WSI) denotes for digitalized images of an entire tissue histopathology slide or a region of interest (ROI) in the tissue slide [239]. The digitalization process involves three main phases: image acquisition, storage and an interface for image visualization and annotation [240]. The type of data generated by a WSI scanner requires drastic software and hardware changes with powerful storage and processing capabilities, as the size of these images can reach gigabytes [241].

The digital “realm” brings two immediate benefits over glass slides: portability and easy sharing [242]. Sending slides between institutes is no longer a problem, opening up new possibilities for collaboration and remote diagnosis or training, called telepathology. However, the cost of the equipment, the training of pathologists and technicians and the discrepancies between the diagnosis in a glass side or in a digital image may preclude its adoption by some institutions [243]. In the future, these limitations will be easily surpassed with the creation of more user-friendly visual interfaces and the reduction of the price of the equipment.

The applications of whole slide imaging go beyond the clinical pathology, education, research and industry. Once in the digital realm, tissue image analysis can generate precise and highly reproducible readouts, reducing inter-observer variability and driving faster diagnosis [244]. This is the purpose of computational pathology.

Computational pathology aims to integrate multiple data sets, such as, medical records, “omics” laboratory data and imaging (both radiology or pathology) to extract clinical and biological meaningful information [245]. It uses mathematical models to create inferences and predictions, providing decision support to stakeholders (from the medical team to the patients) [246]. Despite having different strengths and weaknesses, computational and human pathology can complement each other to improve the patients’ management. Human pathology is currently more versatile across different clinical and pathological settings compared with computer algorithms trained for specific applications. On the other hand, computational pathology objectively assesses, quantifies and relates a large number of features in a systematic and high-throughput way [247].

Advanced commercial and open-source software is starting to gain traction in the world of computational pathology and making its way into the clinical practice. One example is QuPath, designed for WSI analysis [248]. QuPath opens the most common image formats, offering a pallet of tools for high- throughput tissue analysis, from pixel classification, automated cell detection and counting and scripting customization. To the date, QuPath does not include registration tools, making it impossible to quantify co-expression. Another powerful software is Ilastik, an interactive machine learning tools for bioimage analysis [249]. Ilastik processes any number of images that can be annotated either manually or by loading labels. In other words, it workflows allows the interactive image classification, performing a real time visualization of the segmentation performance.

4.1. Computational Approaches to Study Protein Co-Localization

The analysis of multiple markers implies to computationally reconstruct the tridimensional structure in order to extract biologically relevant information. Image registration is the process of spatially align consecutive images. The basis of a registration is similarity, which means that the target image is transformed with reference to a fixed image to achieve the maximum accurate overlap [250].

Due to the diversity of images to be registered and the type of degradations that they present, it is impossible to design a universal method that fits all purposes. Nevertheless, the majority of the registration methods consists of the following four steps: feature detection, feature matching, transform model estimation and image resampling and transformation. Firstly, different kinds of features are detected and matched between the reference and the moving images. Depending on the application, two different types of geometric transformations can be used, linear and non-linear transformations (**Figure 5**). Linear transformations (linear, similarity or affine) are mainly used when the internal structure of an image does not have obvious distortions or deformations [251]. Local deformations in soft tissues or structures will not be solved by linear transformations. In these cases, non-linear transformations (elastic or deformable) are more accurate since they have the capacity of registering local deformations. In general, linear transformations are performed through the calculation of translational, rotational and scaling vectors, ensuring that the structures maintain unchanged, while non-linear transformations include extra parameters such as shear and rotation. [252]

Most of the times, linear transformations are the preferred because they involve a lower number of parameters. To determine whether the best parameters for the transformation were found, there are methods for measuring the similarity, feature-based and intensity-based approaches. When the number of features is not enough, intensity-based approaches are more appropriate [253]. This measurement strategy uses the pixel information from different images and design a correlation based on grayscales.

In sum, registration in an iterative process that results from the constant comparison of similarity until find the most accurate alignment. An iterative optimization algorithm aims to increase the sensitivity of the registration algorithm, reducing the search time. If the target image is very different

of the reference, independently of the amount of iterations, the result of the transformation will not be satisfactory. Therefore, the output of image registration must be always confirmed.

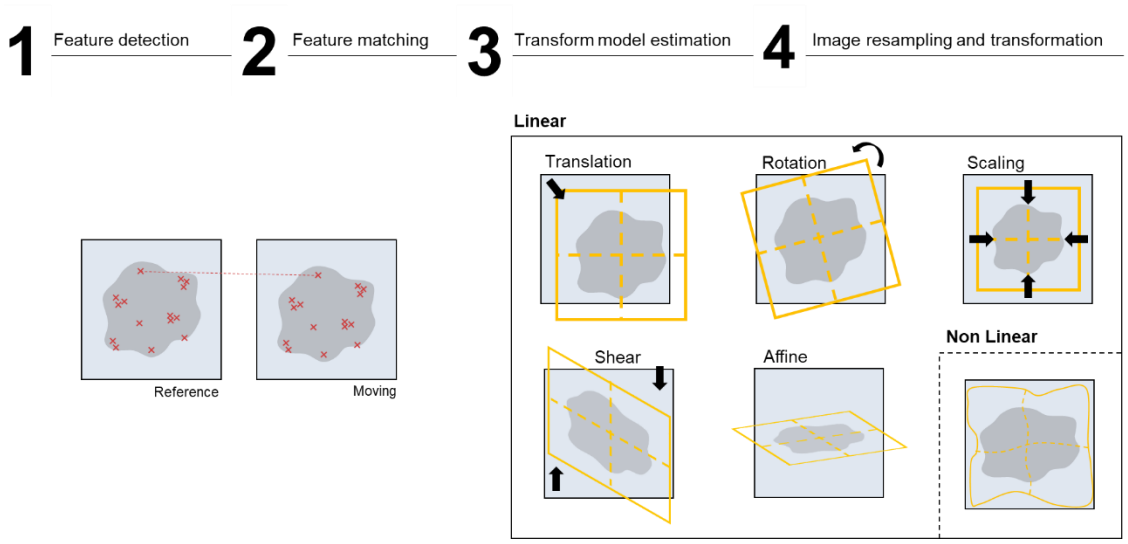


Figure 5. **Multistep process of image registration methods.** First, keypoints are detected at each image and then each keypoint is described as a feature vector, called local descriptor. Second, keypoint correspondence is established by using similarity between local descriptors. Finally, the geometric transformation function is estimated using the keypoint correspondence. Adapted with permission from [254].

Even when image registration is accurate, tumor sections are separated by a microscopic distance. Therefore, computational scientists struggle to find good computational methodologies to define co-localization. Frequently, collected data is analyzed under the user's subjective eye and based on color overlap. Visual evaluation is far of being quantitative, hence, old-fashion color merged images are being substituted by scatter plots, which correlate the intensity of two channels on a pixel-pixel basis [14].

In this matter, multiplex immunohistochemical (mIHC) staining regard as a compelling opportunity to assess multiple protein markers in the same tissue slide, in the same cells. The major advantage of mIHC is the amount of information that can be retrieved from a single slide, which is crucial when the amount of sample is limited [255]. However, multiple pre-analytical and analytical challenges arise from mIHC analysis, particularly the lack of standardization of fixation times, dehydration, paraffin impregnation and the cross-reactivity between multiple antibodies. Despite the huge efforts to standardize estrogen receptor, progesterone receptor and HER2 in breast cancer, they still demonstrate significant inter-laboratory and intra-laboratory variability [256]. In questions of cellular and spatial heterogeneity, it has been demonstrated that the analysis of small ROIs generates a huge variability and error of assessment in tumor and immune markers [257, 258]. Therefore, there is still a need to integrate the mIHC systems coupled with WSI scanners for digital analysis of high-resolution images.

Within the general workflow of digital and computational pathology, the pathologist has an important contribution ensuring the value and quality of the data generated.

In practical terms, pathologists add value in technical knowledge, namely, tissue processing and staining, as well as, biomarker expression patterns. Moreover, they can also participate in some tasks necessary to create a more accurate computational method, such as selection of regions of interest, manual image annotations to distinguish tumor from non-tumor regions and review the generated data. Also, many image analysis software requires the introduction or definition of parameters related with color definition, nuclear detection, expression thresholds or subcellular localization of the biomarker in question [259].

Given the amount of information and meaningful relationships that the computer finds, it is important that both pathologists and computational scientists cooperate to interpret the results within a specific biological and clinical context. To do so, the pathologist does not need to have an extensive knowledge in image analysis tools, although this knowledge can help to anticipate problems and find solutions.

5. References

1. Curtis, N.J., et al., *The relevance of the Siewert classification in the era of multimodal therapy for adenocarcinoma of the gastro-oesophageal junction.* Journal of surgical oncology, 2014. **109**(3): p. 202-207.
2. Oliveira, C., R. Seruca, and F. Carneiro, *Hereditary gastric cancer.* Best practice & research. Clinical gastroenterology, 2009. **23**(2): p. 147-157.
3. Santoro, R., G.M. Ettorre, and E. Santoro, *Subtotal gastrectomy for gastric cancer.* World journal of gastroenterology, 2014. **20**(38): p. 13667-13680.
4. Oliveira, C., R. Seruca, and F. Carneiro, *Genetics, pathology, and clinics of familial gastric cancer.* International journal of surgical pathology, 2006. **14**(1): p. 21-33.
5. Oliveira, C., et al., *Screening E-cadherin in gastric cancer families reveals germline mutations only in hereditary diffuse gastric cancer kindred.* Human mutation, 2002. **19**(5): p. 510-517.
6. Oliveira, C., et al., *Clinical utility gene card for: Hereditary diffuse gastric cancer (HDGC).* European Journal of Human Genetics, 2013. **21**(8): p. 891-891.
7. Caldas, C., et al., *Familial gastric cancer: overview and guidelines for management*.* Journal of Medical Genetics, 1999. **36**(12): p. 873.
8. Oliveira, C., et al., *Familial gastric cancer: genetic susceptibility, pathology, and implications for management.* The Lancet. Oncology, 2015. **16**(2): p. e60-e70.
9. Oliveira, C., et al., *Genetic Screening for Familial Gastric Cancer.* Hereditary Cancer in Clinical Practice, 2004. **2**(2): p. 51.
10. Japanese Gastric Cancer, A., *Japanese classification of gastric carcinoma: 3rd English edition.* Gastric Cancer, 2011. **14**(2): p. 101-112.

11. Holzer, A.K., G.H. Manorek, and S.B. Howell, *Contribution of the major copper influx transporter CTR1 to the cellular accumulation of cisplatin, carboplatin, and oxaliplatin.* Mol Pharmacol, 2006. **70**(4): p. 1390-4.
12. Siewert, J.R., et al., *[Cardia cancer: attempt at a therapeutically relevant classification].* Chirurg, 1987. **58**(1): p. 25-32.
13. Liu, K., et al., *Comparison between gastric and esophageal classification system among adenocarcinomas of esophagogastric junction according to AJCC 8th edition: a retrospective observational study from two high-volume institutions in China.* Gastric cancer : official journal of the International Gastric Cancer Association and the Japanese Gastric Cancer Association, 2019. **22**(3): p. 506-517.
14. Rudloff, U., *Gastric adenocarcinoma and proximal polyposis of the stomach: diagnosis and clinical perspectives.* Clinical and experimental gastroenterology, 2018. **11**: p. 447-459.
15. Nagtegaal, I.D., et al., *Digestive System WHO Classification of Tumours*. 5th edition ed, ed. W.C.o.T.E. Board. Vol. 1. 2019: World Health Organization.
16. Hiki, N., et al., *Function-Preserving Gastrectomy for Early Gastric Cancer.* Annals of Surgical Oncology, 2013. **20**(8): p. 2683-2692.
17. Jung, Y. and S.J. Lippard, *Direct Cellular Responses to Platinum-Induced DNA Damage.* Chemical Reviews, 2007. **107**(5): p. 1387-1407.
18. Smyth, E.C., et al., *Gastric cancer: ESMO Clinical Practice Guidelines for diagnosis, treatment and follow-up†.* Annals of Oncology, 2016. **27**: p. v38-v49.
19. Zhang, S., H. Orita, and T. Fukunaga, *Current surgical treatment of esophagogastric junction adenocarcinoma.* World journal of gastrointestinal oncology, 2019. **11**(8): p. 567-578.
20. Siddik, Z.H., *Cisplatin: mode of cytotoxic action and molecular basis of resistance.* Oncogene, 2003. **22**(47): p. 7265-7279.
21. *Digestive System Tumours - WHO Classification of Tumours*. 5th edition ed. Vol. Volume 1. 2019.
22. Browning, R.J., et al., *Drug Delivery Strategies for Platinum-Based Chemotherapy.* ACS Nano, 2017. **11**(9): p. 8560-8578.
23. Morais, S., et al., *Trends in gastric cancer mortality and in the prevalence of Helicobacter pylori infection in Portugal.* Eur J Cancer Prev, 2016. **25**(4): p. 275-81.
24. Bray, F., et al., *Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries.* CA Cancer J Clin, 2018. **68**(6): p. 394-424.
25. Rocha, C.R.R., et al., *DNA repair pathways and cisplatin resistance: an intimate relationship.* Clinics (Sao Paulo, Brazil), 2018. **73**(suppl 1): p. e478s-e478s.
26. Sitarz, R., et al., *Gastric cancer: epidemiology, prevention, classification, and treatment.* Cancer management and research, 2018. **10**: p. 239-248.
27. Khatoon, J., R.P. Rai, and K.N. Prasad, *Role of Helicobacter pylori in gastric cancer: Updates.* World J Gastrointest Oncol, 2016. **8**(2): p. 147-58.
28. Moayyedi, P., et al., *Relation of adult lifestyle and socioeconomic factors to the prevalence of Helicobacter pylori infection.* Int J Epidemiol, 2002. **31**(3): p. 624-31.

29. Figueiredo, C., et al., *Assessment of Helicobacter pylori vacA and cagA genotypes and host serological response.* Journal of clinical microbiology, 2001. **39**(4): p. 1339-1344.
30. Figueiredo, C., et al., *Helicobacter pylori genotypes are associated with clinical outcome in Portuguese patients and show a high prevalence of infections with multiple strains.* Scand J Gastroenterol, 2001. **36**(2): p. 128-35.
31. Mera, R., et al., *Long term follow up of patients treated for Helicobacter pylori infection.* Gut, 2005. **54**(11): p. 1536-40.
32. Wong, B.C., et al., *Helicobacter pylori eradication to prevent gastric cancer in a high-risk region of China: a randomized controlled trial.* Jama, 2004. **291**(2): p. 187-94.
33. Naumann, M. and J.E. Crabtree, *Helicobacter pylori-induced epithelial cell signalling in gastric carcinogenesis.* Trends Microbiol, 2004. **12**(1): p. 29-36.
34. Correa, P. and J. Houghton, *Carcinogenesis of Helicobacter pylori.* Gastroenterology, 2007. **133**(2): p. 659-72.
35. Li, W.Q., et al., *Effects of Helicobacter pylori treatment and vitamin and garlic supplementation on gastric cancer incidence and mortality: follow-up of a randomized intervention trial.* Bmj, 2019. **366**: p. l5016.
36. Forman, D. and V.J. Burley, *Gastric cancer: global pattern of the disease and an overview of environmental risk factors.* Best Pract Res Clin Gastroenterol, 2006. **20**(4): p. 633-49.
37. Ajani, J.A., et al., *Gastric adenocarcinoma.* Nature Reviews Disease Primers, 2017. **3**(1): p. 17036.
38. Rawla, P. and A. Barsouk, *Epidemiology of gastric cancer: global trends, risk factors and prevention.* Przeglad gastroenterologiczny, 2019. **14**(1): p. 26-38.
39. Mukaisho, K.-I., et al., *Two distinct etiologies of gastric cardia adenocarcinoma: interactions among pH, Helicobacter pylori, and bile acids.* Frontiers in microbiology, 2015. **6**: p. 412-412.
40. Gullo, I., et al., *Heterogeneity in Gastric Cancer: From Pure Morphology to Molecular Classifications.* Pathobiology, 2018. **85**(1-2): p. 50-63.
41. Lauren, P., *THE TWO HISTOLOGICAL MAIN TYPES OF GASTRIC CARCINOMA: DIFFUSE AND SO-CALLED INTESTINAL-TYPE CARCINOMA. AN ATTEMPT AT A HISTO-CLINICAL CLASSIFICATION.* Acta Pathol Microbiol Scand, 1965. **64**: p. 31-49.
42. Ming, S.C., *Gastric carcinoma. A pathobiological classification.* Cancer, 1977. **39**(6): p. 2475-2485.
43. Mulligan, R.M., *Histogenesis and biologic behavior of gastric carcinoma.* Pathology annual, 1972. **7**: p. 349-415.
44. Goseki, N., T. Takizawa, and M. Koike, *Differences in the mode of the extension of gastric cancer classified by histological type: new histological classification of gastric carcinoma.* Gut, 1992. **33**(5): p. 606-612.
45. Carneiro, F., M. Seixas, and M. Sobrinho-Simões, *New elements for an updated classification of the carcinomas of the stomach.* Pathology, research and practice, 1995. **191**(6): p. 571-584.
46. Berlth, F., et al., *Pathohistological classification systems in gastric cancer: diagnostic relevance and prognostic value.* World journal of gastroenterology, 2014. **20**(19): p. 5679-5684.

47. Carneiro, F., et al., *Mixed carcinoma of the stomach: a clinicopathological entity.* Histopathology, 2003. **43**(1): p. 94-95.
48. Adachi, Y., et al., *Pathology and prognosis of gastric carcinoma: well versus poorly differentiated type.* Cancer, 2000. **89**(7): p. 1418-24.
49. Tahara, E., *Genetic pathways of two types of gastric cancer.* IARC Sci Publ, 2004(157): p. 327-49.
50. Correa, P., et al., *A model for gastric cancer epidemiology.* Lancet (London, England), 1975. **2**(7924): p. 58-60.
51. Figueiredo, C., M.A. Garcia-Gonzalez, and J.C. Machado, *Molecular pathogenesis of gastric cancer.* Helicobacter, 2013. **18 Suppl 1**: p. 28-33.
52. Berger, H., et al., *Gastric cancer pathogenesis.* Helicobacter, 2016. **21 Suppl 1**: p. 34-38.
53. Correa, P., *Gastric cancer: overview.* Gastroenterology clinics of North America, 2013. **42**(2): p. 211-217.
54. Ma, J., et al., *Lauren classification and individualized chemotherapy in gastric cancer.* Oncology letters, 2016. **11**(5): p. 2959-2964.
55. Assumpção, P.P., et al., *The diffuse-type gastric cancer epidemiology enigma.* BMC Gastroenterology, 2020. **20**(1): p. 223.
56. Carneiro, P., et al., *E-cadherin dysfunction in gastric cancer--cellular consequences, clinical applications and open questions.* FEBS Lett, 2012. **586**(18): p. 2981-9.
57. Fenoglio-Preiser, C.M., et al., *TP53 and gastric carcinoma: a review.* Hum Mutat, 2003. **21**(3): p. 258-70.
58. Abrahao-Machado, L.F. and C. Scapulatempo-Neto, *HER2 testing in gastric cancer: An update.* World journal of gastroenterology, 2016. **22**(19): p. 4619-4625.
59. Cancer Genome Atlas Research, N., *Comprehensive molecular characterization of gastric adenocarcinoma.* Nature, 2014. **513**(7517): p. 202-209.
60. Cristescu, R., et al., *Molecular analysis of gastric cancer identifies subtypes associated with distinct clinical outcomes.* Nature Medicine, 2015. **21**(5): p. 449-456.
61. Sasako, M., *Progress in the treatment of gastric cancer in Japan over the last 50 years.* Annals of gastroenterological surgery, 2020. **4**(1): p. 21-29.
62. Johnstone, B., *How to detect a cancer before it kills. .* New Scientist, 1985: p. 33-35.
63. Irino, T., et al., *Gastric Cancer in Asia: Unique Features and Management.* American Society of Clinical Oncology educational book. American Society of Clinical Oncology. Annual Meeting, 2017. **37**: p. 279-291.
64. Leung, W.K., et al., *Screening for gastric cancer in Asia: current evidence and practice.* The Lancet Oncology, 2008. **9**(3): p. 279-287.
65. Irino, T., et al., *Gastric Cancer in Asia: Unique Features and Management.* American Society of Clinical Oncology Educational Book, 2017(37): p. 279-291.
66. Schlemper, R.J., et al., *Differences in diagnostic criteria for gastric carcinoma between Japanese and western pathologists.* Lancet, 1997. **349**(9067): p. 1725-9.

67. Sano, T., et al., *Proposal of a new stage grouping of gastric cancer for TNM classification: International Gastric Cancer Association staging project.* Gastric cancer : official journal of the International Gastric Cancer Association and the Japanese Gastric Cancer Association, 2017. **20**(2): p. 217-225.
68. Amin, M.B., et al., *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, 2017. **67**(2): p. 93-99.
69. Brierley, J., M. Gospodarowicz, and B. O'Sullivan, *The principles of cancer staging.* Ecancermedicalscience, 2016. **10**: p. ed61-ed61.
70. Greene, D.M.G.S.B.E.F.L., et al., *AJCC Cancer Staging Manual*. Eighth Edition ed. 2017.
71. Hallinan, J.T.P.D. and S.K. Venkatesh, *Gastric carcinoma: imaging diagnosis, staging and assessment of treatment response.* Cancer imaging : the official publication of the International Cancer Imaging Society, 2013. **13**(2): p. 212-227.
72. Barros, R.H.d.O., et al., *Multidetector computed tomography in the preoperative staging of gastric adenocarcinoma.* Radiologia brasileira, 2015. **48**(2): p. 74-80.
73. Wang, Z. and J.Q. Chen, *Imaging in assessing hepatic and peritoneal metastases of gastric cancer: a systematic review.* BMC Gastroenterol, 2011. **11**: p. 19.
74. *Stomach cancer - TNM staging*. 2020 1 Nov 2019 [cited Fev 2020; Available from: https://www.cancerresearchuk.org/about-cancer/stomach-cancer/stages/tnm-staging.
75. Cho, K.B., W.J. Jeon, and J.J. Kim, *Worldwide experiences of endoscopic submucosal dissection: not just Eastern acrobatics.* World journal of gastroenterology, 2011. **17**(21): p. 2611-2617.
76. Ko, W.J., et al., *Endoscopic resection of early gastric cancer: current status and new approaches.* Translational gastroenterology and hepatology, 2016. **1**: p. 24-24.
77. Dicken, B.J., et al., *Gastric adenocarcinoma: review and considerations for future directions.* Annals of surgery, 2005. **241**(1): p. 27-39.
78. Weledji, E.P., *The principles of the surgical management of gastric cancer.* International journal of surgery. Oncology, 2017. **2**(7): p. e11-e11.
79. Gotoda, T., et al., *Evaluation of the necessity for gastrectomy with lymph node dissection for patients with submucosal invasive gastric cancer.* Br J Surg, 2001. **88**(3): p. 444-9.
80. Kunisaki, C., et al., *Comparison of surgical results of D2 versus D3 gastrectomy (para-aortic lymph node dissection) for advanced gastric carcinoma: a multi-institutional study.* Ann Surg Oncol, 2006. **13**(5): p. 659-67.
81. Nitti, D., et al., *Ratio between metastatic and examined lymph nodes is an independent prognostic factor after D2 resection for gastric cancer: analysis of a large European monoinstitutional experience.* Ann Surg Oncol, 2003. **10**(9): p. 1077-85.
82. Bando, E., et al., *Outcome of ratio of lymph node metastasis in gastric carcinoma.* Ann Surg Oncol, 2002. **9**(8): p. 775-84.
83. Inoue, K., et al., *The superiority of ratio-based lymph node staging in gastric carcinoma.* Ann Surg Oncol, 2002. **9**(1): p. 27-34.

84. Marano, L., et al., *Comparison between 7th and 8th edition of AJCC TNM staging system for gastric cancer: old problems and new perspectives.* 2019, 2019. **4**.
85. Zhang, B.-y., et al., *Evaluation of the prognostic value of the metastatic lymph node ratio for gastric cancer.* The American Journal of Surgery, 2014. **207**(4): p. 555-565.
86. Kunisaki, C., et al., *Clinical Significance of the Metastatic Lymph-Node Ratio in Early Gastric Cancer.* Journal of Gastrointestinal Surgery, 2007. **12**(3): p. 542.
87. Mocellin, S., et al., *Extent of lymph node dissection for adenocarcinoma of the stomach.* The Cochrane database of systematic reviews, 2015. **2015**(8): p. CD001964-CD001964.
88. Sonoda, H. and T. Tani, *Clinical significance of molecular diagnosis for gastric cancer lymph node micrometastasis.* World J Gastroenterol, 2014. **20**(38): p. 13728-33.
89. Songun, I., et al., *Surgical treatment of gastric cancer: 15-year follow-up results of the randomised nationwide Dutch D1D2 trial.* Lancet Oncol, 2010. **11**(5): p. 439-49.
90. Lorenzon, L., et al., *D1-plus vs D2 nodal dissection in gastric cancer: a propensity score matched comparison and review of published literature.* BMC surgery, 2020. **20**(1): p. 126-126.
91. Sleeman, J.P., I. Nazarenko, and W. Thiele, *Do all roads lead to Rome? Routes to metastasis development.* Int J Cancer, 2011. **128**(11): p. 2511-26.
92. Namikawa, T. and K. Hanazaki, *Clinicopathological features and treatment outcomes of metastatic tumors in the stomach.* Surgery Today, 2014. **44**(8): p. 1392-1399.
93. Ameur, W.B., et al., *Bone metastasis as the first sign of gastric cancer.* The Pan African medical journal, 2017. **28**: p. 95-95.
94. Karamchandani, M.M., et al., *A Rare Occurrence of Isolated Brain Metastases from Gastric Cancer.* Case Reports in Medicine, 2019. **2019**: p. 8075421.
95. Muthukrishnan, S., et al., *Krukenberg tumours from gastrointestinal cancers-analysis from a tertiary care centre in India.* Journal of gastrointestinal oncology, 2018. **9**(6): p. 1164-1167.
96. Liu, D., et al., *The patterns and timing of recurrence after curative resection for gastric cancer in China.* World Journal of Surgical Oncology, 2016. **14**(1): p. 305.
97. Gunderson, L.L., *Gastric cancer--patterns of relapse after surgical resection.* Seminars in radiation oncology, 2002. **12**(2): p. 150-161.
98. Wang, S.-Y., et al., *Clinical impact of positive surgical margin status on gastric cancer patients undergoing gastrectomy.* Annals of surgical oncology, 2009. **16**(10): p. 2738-2743.
99. Chen, J.D., et al., *Prognostic improvement of reexcision for positive resection margins in patients with advanced gastric cancer.* European journal of surgical oncology : the journal of the European Society of Surgical Oncology and the British Association of Surgical Oncology, 2013. **39**(3): p. 229-234.
100. Celli, R., et al., *Optimal Intraoperative Assessment of Gastric Margins.* American Journal of Clinical Pathology, 2018. **150**(4): p. 353-363.
101. Mahe, E., et al., *Intraoperative pathology consultation: error, cause and impact.* Canadian journal of surgery. Journal canadien de chirurgie, 2013. **56**(3): p. E13-E18.
102. Quan, Y., et al., *Comparison of laparoscopic versus open gastrectomy for advanced gastric cancer: an updated meta-analysis.* Gastric Cancer, 2016. **19**(3): p. 939-950.

103. Kim, H.-H., et al., *Prospective randomized controlled trial (phase III) to comparing laparoscopic distal gastrectomy with open distal gastrectomy for gastric adenocarcinoma (KLASS 01).* Journal of the Korean Surgical Society, 2013. **84**(2): p. 123-130.
104. Nakamura, K., et al., *A phase III study of laparoscopy-assisted versus open distal gastrectomy with nodal dissection for clinical stage IA/IB gastric Cancer (JCOG0912).* Jpn J Clin Oncol, 2013. **43**(3): p. 324-7.
105. Koizumi, W., et al., *S-1 plus cisplatin versus S-1 alone for first-line treatment of advanced gastric cancer (SPIRITS trial): a phase III trial.* Lancet Oncol, 2008. **9**(3): p. 215-21.
106. Chang, H.M., et al., *Phase II Study of Paclitaxel and Carboplatin in Advanced Gastric Cancer Previously Treated with 5-Fluorouracil and Platinum.* Japanese Journal of Clinical Oncology, 2005. **35**(5): p. 251-255.
107. Okines, A.F.C., et al., *Meta-analysis of the REAL-2 and ML17032 trials: evaluating capecitabine-based combination chemotherapy and infused 5-fluorouracil-based combination chemotherapy for the treatment of advanced oesophago-gastric cancer.* Annals of oncology : official journal of the European Society for Medical Oncology, 2009. **20**(9): p. 1529-1534.
108. Kang, Y.K., et al., *Capecitabine/cisplatin versus 5-fluorouracil/cisplatin as first-line therapy in patients with advanced gastric cancer: a randomised phase III noninferiority trial.* Annals of oncology : official journal of the European Society for Medical Oncology, 2009. **20**(4): p. 666-673.
109. Cunningham, D., et al., *Perioperative chemotherapy versus surgery alone for resectable gastroesophageal cancer.* The New England journal of medicine, 2006. **355**(1): p. 11-20.
110. Rivera, F., et al., *NEOHX study: Perioperative treatment with trastuzumab in combination with capecitabine and oxaliplatin (XELOX-T) in patients with HER-2 resectable stomach or esophagogastric junction (EGJ) adenocarcinoma—18 m DFS analysis.* Journal of Clinical Oncology, 2015. **33**(3_suppl): p. 107-107.
111. Bang, Y.-J., et al., *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 (London, England), 2010. **376**(9742): p. 687-697.
112. Wagner, A.D., et al., *EORTC-1203-GITCG - the "INNOVATION"-trial: Effect of chemotherapy alone versus chemotherapy plus trastuzumab, versus chemotherapy plus trastuzumab plus pertuzumab, in the perioperative treatment of HER2 positive, gastric and gastroesophageal junction adenocarcinoma on pathologic response rate: a randomized phase II-intergroup trial of the EORTC-Gastrointestinal Tract Cancer Group, Korean Cancer Study Group and Dutch Upper GI-Cancer group.* BMC cancer, 2019. **19**(1): p. 494-494.
113. Khan, U. and M.A. Shah, *Optimizing Therapies in the Perioperative Management of Gastric Cancer.* Current Treatment Options in Oncology, 2019. **20**(7): p. 57.
114. Macdonald, J.S., et al., *Chemoradiotherapy after surgery compared with surgery alone for adenocarcinoma of the stomach or gastroesophageal junction.* The New England journal of medicine, 2001. **345**(10): p. 725-730.

115. Smalley, S.R., et al., *Updated analysis of SWOG-directed intergroup study 0116: a phase III trial of adjuvant radiochemotherapy versus observation after curative gastric cancer resection.* Journal of clinical oncology : official journal of the American Society of Clinical Oncology, 2012. **30**(19): p. 2327-2333.
116. Lee, J., et al., *Phase III trial comparing capecitabine plus cisplatin versus capecitabine plus cisplatin with concurrent capecitabine radiotherapy in completely resected gastric cancer with D2 lymph node dissection: the ARTIST trial.* Journal of clinical oncology : official journal of the American Society of Clinical Oncology, 2012. **30**(3): p. 268-273.
117. Park, S.H., et al., *ARTIST 2: Interim results of a phase III trial involving adjuvant chemotherapy and/or chemoradiotherapy after D2-gastrectomy in stage II/III gastric cancer (GC).* Journal of Clinical Oncology, 2019. **37**(15_suppl): p. 4001-4001.
118. Zhang, N., et al., *Progress of preoperative and postoperative radiotherapy in gastric cancer.* World Journal of Surgical Oncology, 2018. **16**(1): p. 187.
119. Wagner, A.D., et al., *Chemotherapy for advanced gastric cancer.* The Cochrane database of systematic reviews, 2010(3): p. CD004064-CD004064.
120. Glimelius, B., et al., *Randomized comparison between chemotherapy plus best supportive care with best supportive care in advanced gastric cancer.* Annals of oncology : official journal of the European Society for Medical Oncology, 1997. **8**(2): p. 163-168.
121. Cunningham, D., et al., *Capecitabine and oxaliplatin for advanced esophagogastric cancer.* The New England journal of medicine, 2008. **358**(1): p. 36-46.
122. Dank, M., et al., *Randomized phase III study comparing irinotecan combined with 5-fluorouracil and folinic acid to cisplatin combined with 5-fluorouracil in chemotherapy naive patients with advanced adenocarcinoma of the stomach or esophagogastric junction.* Ann Oncol, 2008. **19**(8): p. 1450-1457.
123. Van Cutsem, E., et al., *Phase III study of docetaxel and cisplatin plus fluorouracil compared with cisplatin and fluorouracil as first-line therapy for advanced gastric cancer: a report of the V325 Study Group.* J Clin Oncol, 2006. **24**(31): p. 4991-7.
124. Jeyakumar, A. and T. Younis, *Trastuzumab for HER2-Positive Metastatic Breast Cancer: Clinical and Economic Considerations.* Clinical Medicine Insights. Oncology, 2012. **6**: p. 179-187.
125. Hanahan, D. and R.A. Weinberg, *Hallmarks of cancer: the next generation.* Cell, 2011. **144**(5): p. 646-674.
126. Rajabi, M. and S.A. Mousa, *The Role of Angiogenesis in Cancer Treatment.* Biomedicines, 2017. **5**(2): p. 34.
127. Wiesmann, C., *Chapter 42 - The Mechanism of VEGFR Activation by VEGF*, in *Handbook of Cell Signaling (Second Edition)*, R.A. Bradshaw and E.A. Dennis, Editors. 2010, Academic Press: San Diego. p. 287-292.
128. Fuchs, C.S., et al., *Ramucirumab monotherapy for previously treated advanced gastric or gastro-oesophageal junction adenocarcinoma (REGARD): an international, randomised,*

multicentre, placebo-controlled, phase 3 trial. Lancet (London, England), 2014. **383**(9911): p. 31-39.

129. Wilke, H., et al., *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.* The Lancet. Oncology, 2014. **15**(11): p. 1224-1235.
130. Lordick, F., et al., *Unmet needs and challenges in gastric cancer: the way forward.* Cancer treatment reviews, 2014. **40**(6): p. 692-700.
131. Rawicz-Pruszyński, K., et al., *Current challenges in gastric cancer surgery: European perspective.* Surgical oncology, 2018. **27**(4): p. 650-656.
132. Leong, T., et al., *TOPGEAR: A Randomized, Phase III Trial of Perioperative ECF Chemotherapy with or Without Preoperative Chemoradiation for Resectable Gastric Cancer: Interim Results from an International, Intergroup Trial of the AGITG, TROG, EORTC and CCTG.* Annals of surgical oncology, 2017. **24**(8): p. 2252-2258.
133. Slagter, A.E., et al., *CRITICS-II: a multicentre randomised phase II trial of neo-adjuvant chemotherapy followed by surgery versus neo-adjuvant chemotherapy and subsequent chemoradiotherapy followed by surgery versus neo-adjuvant chemoradiotherapy followed by surgery in resectable gastric cancer.* BMC cancer, 2018. **18**(1): p. 877-877.
134. *Quality assurance in surgical oncology the EURECCA platform.* European journal of surgical oncology : the journal of the European Society of Surgical Oncology and the British Association of Surgical Oncology, 2014. **40**(11): p. 1387-1390.
135. Yun, Y.H., et al., *The influence of hospital volume and surgical treatment delay on long-term survival after cancer surgery.* Annals of oncology : official journal of the European Society for Medical Oncology, 2012. **23**(10): p. 2731-2737.
136. Brenkman, H.J.F., et al., *Association Between Waiting Time from Diagnosis to Treatment and Survival in Patients with Curable Gastric Cancer: A Population-Based Study in the Netherlands.* Annals of surgical oncology, 2017. **24**(7): p. 1761-1769.
137. Witzig, R., et al., *Delays in diagnosis and therapy of gastric cancer and esophageal adenocarcinoma.* Endoscopy, 2006. **38**(11): p. 1122-1126.
138. Strimbu, K. and J.A. Tavel, *What are biomarkers?* Current opinion in HIV and AIDS, 2010. **5**(6): p. 463-466.
139. Oldenhuis, C.N.A.M., et al., *Prognostic versus predictive value of biomarkers in oncology.* European journal of cancer (Oxford, England : 1990), 2008. **44**(7): p. 946-953.
140. Necula, L., et al., *Recent advances in gastric cancer early diagnosis.* World journal of gastroenterology, 2019. **25**(17): p. 2029-2044.
141. Matsuoka, T. and M. Yashiro, *Biomarkers of gastric cancer: Current topics and future perspective.* World journal of gastroenterology, 2018. **24**(26): p. 2818-2832.
142. Sawaki, K., M. Kanda, and Y. Kodera, *Review of recent efforts to discover biomarkers for early detection, monitoring, prognosis, and prediction of treatment responses of patients with gastric cancer.* Expert review of gastroenterology & hepatology, 2018. **12**(7): p. 657-670.

143. Baniak, N., et al., *Gastric biomarkers: a global review.* World journal of surgical oncology, 2016. **14**(1): p. 212-212.
144. Zhang, J., et al., *AGR2 is associated with gastric cancer progression and poor survival.* Oncol Lett, 2016. **11**(3): p. 2075-2083.
145. Liu, C.G., et al., *High expression of the ANKRD49 protein is associated with progression and poor prognosis of gastric cancer.* Cancer Biomark, 2018. **22**(4): p. 649-656.
146. Kanda, M., et al., *Function and diagnostic value of Anosmin-1 in gastric cancer progression.* International Journal of Cancer, 2016. **138**(3): p. 721-730.
147. Feng, M., et al., *Tumor apelin, not serum apelin, is associated with the clinical features and prognosis of gastric cancer.* BMC Cancer, 2016. **16**(1): p. 794.
148. Subhash, V.V., et al., *ATM Expression Predicts Veliparib and Irinotecan Sensitivity in Gastric Cancer by Mediating P53-Independent Regulation of Cell Cycle and Apoptosis.* Mol Cancer Ther, 2016. **15**(12): p. 3087-3096.
149. Xu, Y.-Y., et al., *Regulation of CD44v6 expression in gastric carcinoma by the IL-6/STAT3 signaling pathway and its clinical significance.* Oncotarget, 2017. **8**(28): p. 45848-45861.
150. Fang, M., et al., *CD44 and CD44v6 are Correlated with Gastric Cancer Progression and Poor Patient Prognosis: Evidence from 42 Studies.* Cellular Physiology and Biochemistry, 2016. **40**(3-4): p. 567-578.
151. Chen, Y., et al., *The prognostic value of CD44 expression in gastric cancer: a meta-analysis.* Biomedicine & pharmacotherapy = Biomedecine & pharmacotherapie, 2014. **68**(6): p. 693-697.
152. Wang, L., et al., *Overexpressed CISD2 has prognostic value in human gastric cancer and promotes gastric cancer cell proliferation and tumorigenesis via AKT signaling pathway.* Oncotarget, 2016. **7**(4): p. 3791-805.
153. Wei, J., et al., *CYR61 (CCN1) is a metastatic biomarker of gastric cardia adenocarcinoma.* Oncotarget, 2016. **7**(21): p. 31067-31078.
154. Wang, D., et al., *High expression of EGFR predicts poor survival in patients with resected T3 stage gastric adenocarcinoma and promotes cancer cell survival.* Oncology letters, 2017. **13**(5): p. 3003-3013.
155. Betts, G., et al., *FGFR2, HER2 and cMet in gastric adenocarcinoma: detection, prognostic significance and assessment of downstream pathway activation.* Virchows Arch, 2014. **464**(2): p. 145-56.
156. Wang, Q., et al., *Tumor and serum gamma-glutamyl transpeptidase, new prognostic and molecular interpretation of an old biomarker in gastric cancer.* Oncotarget, 2017. **8**(22): p. 36171-36184.
157. Yashiro, M., et al., *A c-Met inhibitor increases the chemosensitivity of cancer stem cells to the irinotecan in gastric carcinoma.* British Journal of Cancer, 2013. **109**(10): p. 2619-2628.
158. Noguchi, E., et al., *Clinical significance of hepatocyte growth factor/c-Met expression in the assessment of gastric cancer progression.* Mol Med Rep, 2015. **11**(5): p. 3423-31.
159. Huang, X., et al., *Clinical value of CagA, c-Met, PI3K and Beclin-1 expressed in gastric cancer and their association with prognosis.* Oncol Lett, 2018. **15**(1): p. 947-955.

160. Sakamoto, N., et al., *MET4 expression predicts poor prognosis of gastric cancers with Helicobacter pylori infection.* Cancer Sci, 2017. **108**(3): p. 322-330.
161. Cao, L., et al., *MMP16 is a marker of poor prognosis in gastric cancer promoting proliferation and invasion.* Oncotarget, 2016. **7**(32): p. 51865-51874.
162. Mukai, S., et al., *Overexpression of PCDHB9 promotes peritoneal metastasis and correlates with poor prognosis in patients with gastric cancer.* J Pathol, 2017. **243**(1): p. 100-110.
163. Xu, D.-z., et al., *Activated mammalian target of rapamycin is a potential therapeutic target in gastric cancer.* BMC Cancer, 2010. **10**(1): p. 536.
164. Weng, W., et al., *PTTG3P promotes gastric tumour cell proliferation and invasion and is an indicator of poor prognosis.* Journal of cellular and molecular medicine, 2017. **21**(12): p. 3360-3371.
165. Bai, T., et al., *High STMN1 level is associated with chemo-resistance and poor prognosis in gastric cancer patients.* Br J Cancer, 2017. **116**(9): p. 1177-1185.
166. Zhao, W., et al., *Trop2 is overexpressed in gastric cancer and predicts poor prognosis.* Oncotarget, 2016. **7**(5): p. 6136-45.
167. Kubo, T., et al., *BAK is a predictive and prognostic biomarker for the therapeutic effect of docetaxel treatment in patients with advanced gastric cancer.* Gastric Cancer, 2016. **19**(3): p. 827-38.
168. Stahl, D., et al., *Low BUB1 expression is an adverse prognostic marker in gastric adenocarcinoma.* Oncotarget, 2017. **8**(44): p. 76329-76339.
169. Corso, G., et al., *Somatic mutations and deletions of the E-cadherin gene predict poor survival of patients with gastric cancer.* J Clin Oncol, 2013. **31**(7): p. 868-75.
170. Xing, X., et al., *The prognostic value of E-cadherin in gastric cancer: a meta-analysis.* Int J Cancer, 2013. **132**(11): p. 2589-96.
171. Tanaka, H., et al., *FAM46C Serves as a Predictor of Hepatic Recurrence in Patients with Resectable Gastric Cancer.* Ann Surg Oncol, 2017. **24**(11): p. 3438-3445.
172. Cao, Y., et al., *Association of O6-Methylguanine-DNA Methyltransferase Protein Expression With Postoperative Prognosis and Adjuvant Chemotherapeutic Benefits Among Patients With Stage II or III Gastric Cancer.* JAMA Surg, 2017. **152**(11): p. e173120.
173. Altan, B., et al., *Nuclear PRMT1 expression is associated with poor prognosis and chemosensitivity in gastric cancer patients.* Gastric Cancer, 2016. **19**(3): p. 789-97.
174. Ueta, K., et al., *PROX1 Is Associated with Cancer Progression and Prognosis in Gastric Cancer.* Anticancer Res, 2018. **38**(11): p. 6139-6145.
175. Wu, Y., et al., *Down regulation of RNA binding motif, single-stranded interacting protein 3, along with up regulation of nuclear HIF1A correlates with poor prognosis in patients with gastric cancer.* Oncotarget, 2017. **8**(1): p. 1262-1277.
176. Li, D., et al., *Expression and clinical implication of S100A12 in gastric carcinoma.* Tumour Biol, 2016. **37**(5): p. 6551-9.
177. Kanda, M., et al., *Prognostic relevance of SAMSN1 expression in gastric cancer.* Oncol Lett, 2016. **12**(6): p. 4708-4716.

178. Zhou, N., et al., *Downregulated SASH1 expression indicates poor clinical prognosis in gastric cancer.* Hum Pathol, 2018. **74**: p. 83-91.
179. Carvalho, J., et al., *Lack of microRNA-101 causes E-cadherin functional deregulation through EZH2 up-regulation in intestinal gastric cancer.* J Pathol, 2012. **228**(1): p. 31-44.
180. Harris, T.J. and U. Tepass, *Adherens junctions: from molecules to morphogenesis.* Nat Rev Mol Cell Biol, 2010. **11**(7): p. 502-14.
181. Berx, G. and F. van Roy, *Involvement of members of the cadherin superfamily in cancer.* Cold Spring Harbor perspectives in biology, 2009. **1**(6): p. a003129-a003129.
182. van Roy, F., *Beyond E-cadherin: roles of other cadherin superfamily members in cancer.* Nature Reviews Cancer, 2014. **14**(2): p. 121-134.
183. Kourtidis, A., S.P. Ngok, and P.Z. Anastasiadis, *p120 catenin: an essential regulator of cadherin stability, adhesion-induced signaling, and cancer progression.* Progress in molecular biology and translational science, 2013. **116**: p. 409-432.
184. Anastasiadis, P.Z., *p120-ctn: A nexus for contextual signaling via Rho GTPases.* Biochim Biophys Acta, 2007. **1773**(1): p. 34-46.
185. Guilford, P., *E-cadherin downregulation in cancer: fuel on the fire?* Mol Med Today, 1999. **5**(4): p. 172-7.
186. He, T.C., et al., *Identification of c-MYC as a target of the APC pathway.* Science, 1998. **281**(5382): p. 1509-12.
187. Ilyas, M. and I.P. Tomlinson, *The interactions of APC, E-cadherin and beta-catenin in tumour development and progression.* J Pathol, 1997. **182**(2): p. 128-37.
188. Takeichi, M., *Dynamic contacts: rearranging adherens junctions to drive epithelial remodelling.* Nature Reviews Molecular Cell Biology, 2014. **15**(6): p. 397-410.
189. Takeichi, M., *Multiple functions of α-catenin beyond cell adhesion regulation.* Curr Opin Cell Biol, 2018. **54**: p. 24-29.
190. Rodriguez, F.J., L.J. Lewis-Tuffin, and P.Z. Anastasiadis, *E-cadherin's dark side: possible role in tumor progression.* Biochimica et biophysica acta, 2012. **1826**(1): p. 23-31.
191. Oliveira, C., et al., *Germline CDH1 deletions in hereditary diffuse gastric cancer families.* Human molecular genetics, 2009. **18**(9): p. 1545-1555.
192. Knudson, A.G., Jr., *Mutation and cancer: statistical study of retinoblastoma.* Proceedings of the National Academy of Sciences of the United States of America, 1971. **68**(4): p. 820-823.
193. Oliveira, C., et al., *Quantification of epigenetic and genetic 2nd hits in CDH1 during hereditary diffuse gastric cancer syndrome progression.* Gastroenterology, 2009. **136**(7): p. 2137-48.
194. Karayiannakis, A.J., et al., *E-cadherin expression as a differentiation marker in gastric cancer.* Hepatogastroenterology, 1998. **45**(24): p. 2437-42.
195. Torabizadeh, Z., et al., *Evaluation of E-cadherin Expression in Gastric Cancer and Its Correlation with Clinicopathologic Parameters.* International journal of hematology-oncology and stem cell research, 2017. **11**(2): p. 158-164.

196. Anbiaee, R., et al., *Abnormal expression of e-cadherin in gastric adenocarcinoma, and its correlation with tumor histopathology and helicobacter pylori infection.* Iran Red Crescent Med J, 2013. **15**(3): p. 218-22.
197. Gabbert, H.E., et al., *Prognostic value of E-cadherin expression in 413 gastric carcinomas.* Int J Cancer, 1996. **69**(3): p. 184-9.
198. Lazăr, D., et al., *The immunohistochemical expression of E-cadherin in gastric cancer; correlations with clinicopathological factors and patients' survival.* Rom J Morphol Embryol, 2008. **49**(4): p. 459-67.
199. Schizas, D., et al., *E-cadherin in gastric carcinomas: Relations with histological parameters and its prognostic value.* J buon, 2017. **22**(2): p. 383-389.
200. Oka, H., et al., *[Immunoreactive expression of E-cadherin in human gastric cancer: preliminary report].* Nihon Geka Gakkai zasshi, 1990. **91**(12): p. 1814.
201. Goodison, S., V. Urquidi, and D. Tarin, *CD44 cell adhesion molecules.* Mol Pathol, 1999. **52**(4): p. 189-96.
202. Orian-Rousseau, V., *CD44, a therapeutic target for metastasising tumours.* Eur J Cancer, 2010. **46**(7): p. 1271-7.
203. Naor, D., et al., *Involvement of CD44, a molecule with a thousand faces, in cancer dissemination.* Semin Cancer Biol, 2008. **18**(4): p. 260-7.
204. Screaton, G.R., et al., *Genomic structure of DNA encoding the lymphocyte homing receptor CD44 reveals at least 12 alternatively spliced exons.* Proc Natl Acad Sci U S A, 1992. **89**(24): p. 12160-4.
205. Zoller, M., *CD44: can a cancer-initiating cell profit from an abundantly expressed molecule?* Nat Rev Cancer, 2011. **11**(4): p. 254-67.
206. Aruffo, A., et al., *CD44 is the principal cell surface receptor for hyaluronate.* Cell, 1990. **61**(7): p. 1303-13.
207. Knutson, J.R., et al., *CD44/chondroitin sulfate proteoglycan and alpha 2 beta 1 integrin mediate human melanoma cell migration on type IV collagen and invasion of basement membranes.* Mol Biol Cell, 1996. **7**(3): p. 383-96.
208. Jalkanen, M., K. Elenius, and M. Salmivirta, *Syndecan — A Cell Surface Proteoglycan that Selectively Binds Extracellular Effector Molecules*, in *Heparin and Related Polysaccharides*, D.A. Lane, I. Björk, and U. Lindahl, Editors. 1992, Springer US: Boston, MA. p. 79-85.
209. Weber, G.F., et al., *Receptor-ligand interaction between CD44 and osteopontin (Eta-1).* Science, 1996. **271**(5248): p. 509-12.
210. Isacke, C.M., *The role of the cytoplasmic domain in regulating CD44 function.* J Cell Sci, 1994. **107 (Pt 9)**: p. 2353-9.
211. Lokeshwar, V.B., N. Fregien, and L.Y. Bourguignon, *Ankyrin-binding domain of CD44(GP85) is required for the expression of hyaluronic acid-mediated adhesion function.* J Cell Biol, 1994. **126**(4): p. 1099-109.
212. Fehon, R.G., A.I. McClatchey, and A. Bretscher, *Organizing the cell cortex: the role of ERM proteins.* Nature Reviews Molecular Cell Biology, 2010. **11**(4): p. 276-287.

213. Gallatin, W.M., I.L. Weissman, and E.C. Butcher, *A cell-surface molecule involved in organ-specific homing of lymphocytes.* Nature, 1983. **304**(5921): p. 30-34.
214. Ponta, H., D. Wainwright, and P. Herrlich, *The CD44 protein family.* Int J Biochem Cell Biol, 1998. **30**(3): p. 299-305.
215. Ponta, H., L. Sherman, and P.A. Herrlich, *CD44: from adhesion molecules to signalling regulators.* Nat Rev Mol Cell Biol, 2003. **4**(1): p. 33-45.
216. Yan, Y., X. Zuo, and D. Wei, *Concise Review: Emerging Role of CD44 in Cancer Stem Cells: A Promising Biomarker and Therapeutic Target.* Stem Cells Transl Med, 2015. **4**(9): p. 1033-43.
217. Ochiai, S., et al., *Expression of CD44 standard and CD44 variant 6 in human lung cancer.* Nihon Kyobu Shikkan Gakkai Zasshi, 1997. **35**(11): p. 1179-85.
218. Ayhan, A., et al., *Overexpression of CD44 variant 6 in human endometrial cancer and its prognostic significance.* Gynecol Oncol, 2001. **80**(3): p. 355-8.
219. Katsura, M., et al., *Overexpression of CD44 variants 6 and 7 in human endometrial cancer.* Gynecol Oncol, 1998. **71**(2): p. 185-9.
220. Ishida, T., *Immunohistochemical Expression of the CD44 Variant 6 in Colorectal Adenocarcinoma.* Surgery Today, 2000. **30**(1): p. 28-32.
221. da Cunha, C.B., et al., *De novo expression of CD44 variants in sporadic and hereditary gastric cancer.* Lab Invest, 2010. **90**(11): p. 1604-14.
222. Volz, Y., et al., *Direct binding of hepatocyte growth factor and vascular endothelial growth factor to CD44v6.* Bioscience reports, 2015. **35**(4): p. e00236.
223. Tremmel, M., et al., *A CD44v6 peptide reveals a role of CD44 in VEGFR-2 signaling and angiogenesis.* Blood, 2009. **114**(25): p. 5236-44.
224. Gunthert, U., et al., *A new variant of glycoprotein CD44 confers metastatic potential to rat carcinoma cells.* Cell, 1991. **65**(1): p. 13-24.
225. Heider, K.H., et al., *Differential expression of CD44 splice variants in intestinal- and diffuse-type human gastric carcinomas and normal gastric mucosa.* Cancer Res, 1993. **53**(18): p. 4197-203.
226. Mulder, J.W., et al., *Colorectal cancer prognosis and expression of exon-v6-containing CD44 proteins.* Lancet (London, England), 1994. **344**(8935): p. 1470-1472.
227. Parker, D., *Colorectal cancer prognosis and expression of exon-v6-containing CD44 proteins.* Lancet (London, England), 1995. **345**(8949): p. 583-584.
228. Peng, J., et al., *Prediction of treatment outcome by CD44v6 after total mesorectal excision in locally advanced rectal cancer.* Cancer journal (Sudbury, Mass.), 2008. **14**(1): p. 54-61.
229. Köbel, M., et al., *Epithelial hyaluronic acid and CD44v6 are mutually involved in invasion of colorectal adenocarcinomas and linked to patient prognosis.* Virchows Archiv : an international journal of pathology, 2004. **445**(5): p. 456-464.
230. Xie, J.-W., et al., *Evaluation of the prognostic value and functional roles of CD44v6 in gastric cancer.* Journal of cancer research and clinical oncology, 2015. **141**(10): p. 1809-1817.
231. Heider, K.H., et al., *CD44v6: a target for antibody-based cancer therapy.* Cancer Immunol Immunother, 2004. **53**(7): p. 567-79.

232. Orian-Rousseau, V. and H. Ponta, *Perspectives of CD44 targeting therapies.* Arch Toxicol, 2015. **89**(1): p. 3-14.
233. Tijink, B.M., et al., *A phase I dose escalation study with anti-CD44v6 bivatuzumab mertansine in patients with incurable squamous cell carcinoma of the head and neck or esophagus.* Clin Cancer Res, 2006. **12**(20 Pt 1): p. 6064-72.
234. Spiegelberg, D. and J. Nilvebrant, *CD44v6-Targeted Imaging of Head and Neck Squamous Cell Carcinoma: Antibody-Based Approaches.* Contrast Media Mol Imaging, 2017. **2017**: p. 2709547.
235. Kennedy, P.J., et al., *Fab-conjugated PLGA nanoparticles effectively target cancer cells expressing human CD44v6.* Acta Biomaterialia, 2018. **81**: p. 208-218.
236. Hanna, M.G., et al., *Whole slide imaging equivalency and efficiency study: experience at a large academic center.* Modern Pathology, 2019. **32**(7): p. 916-928.
237. Jukić, D.M., et al., *Clinical examination and validation of primary diagnosis in anatomic pathology using whole slide digital images.* Archives of pathology & laboratory medicine, 2011. **135**(3): p. 372-378.
238. Frei, A.L., et al., *Future Medicine: Digital Pathology.* Therapeutische Umschau. Revue therapeutique, 2019. **76**(7): p. 404-408.
239. Pantanowitz, L., et al., *Validating whole slide imaging for diagnostic purposes in pathology: guideline from the College of American Pathologists Pathology and Laboratory Quality Center.* Archives of pathology & laboratory medicine, 2013. **137**(12): p. 1710-1722.
240. Pantanowitz, L., *Digital images and the future of digital pathology.* Journal of pathology informatics, 2010. **1**: p. 15.
241. Brachtel, E. and Y. Yagi, *Digital imaging in pathology--current applications and challenges.* Journal of biophotonics, 2012. **5**(4): p. 327-335.
242. Pantanowitz, L., et al., *Review of the current state of whole slide imaging in pathology.* Journal of pathology informatics, 2011. **2**: p. 36-36.
243. Krishnamurthy, S., et al., *Multi-institutional comparison of whole slide digital imaging and optical microscopy for interpretation of hematoxylin-eosin-stained breast tissue sections.* Archives of pathology & laboratory medicine, 2013. **137**(12): p. 1733-1739.
244. Indu, M., R. Rathy, and M.P. Binu, *"Slide less pathology": Fairy tale or reality?* Journal of oral and maxillofacial pathology : JOMFP, 2016. **20**(2): p. 284-288.
245. Louis, D.N., et al., *Computational Pathology: A Path Ahead.* Archives of pathology & laboratory medicine, 2016. **140**(1): p. 41-50.
246. Louis, D.N., et al., *Computational pathology: an emerging definition.* Archives of pathology & laboratory medicine, 2014. **138**(9): p. 1133-1138.
247. Bui, M.M., et al., *Digital and Computational Pathology: Bring the Future into Focus.* Journal of Pathology Informatics, 2019. **10**: p. 10.
248. Bankhead, P., et al., *QuPath: Open source software for digital pathology image analysis.* Scientific Reports, 2017. **7**(1): p. 16878.

249. Berg, S., et al., *ilastik: interactive machine learning for (bio)image analysis.* Nature Methods, 2019. **16**(12): p. 1226-1232.
250. Moles Lopez, X., et al., *Registration of whole immunohistochemical slide images: an efficient way to characterize biomarker colocalization.* Journal of the American Medical Informatics Association : JAMIA, 2015. **22**(1): p. 86-99.
251. Zhao, Z., et al. *Multi-Modality Medical Image registration Using Support Vector Machines*. in *2005 IEEE Engineering in Medicine and Biology 27th Annual Conference*. 2005.
252. Song, G., et al., *A Review on Medical Image Registration as an Optimization Problem.* Current medical imaging reviews, 2017. **13**(3): p. 274-283.
253. Cattin, P.C., et al., *Retina mosaicing using local features.* Medical image computing and computer-assisted intervention : MICCAI ... International Conference on Medical Image Computing and Computer-Assisted Intervention, 2006. **9**(Pt 2): p. 185-192.
254. Uchida, S., *Image processing and recognition for biological images.* Dev Growth Differ, 2013. **55**(4): p. 523-49.
255. Decalf, J., M.L. Albert, and J. Ziai, *New tools for pathology: a user's review of a highly multiplexed method for in situ analysis of protein and RNA expression in tissue.* The Journal of pathology, 2019. **247**(5): p. 650-661.
256. Lin, C.-Y., et al., *Regional Variability in Percentage of Breast Cancers Reported as Positive for HER2 in California: Implications of Patient Demographics on Laboratory Benchmarks.* American Journal of Clinical Pathology, 2017. **148**(3): p. 199-207.
257. Barnes, M., et al., *Whole tumor section quantitative image analysis maximizes between-pathologists' reproducibility for clinical immunohistochemistry-based biomarkers.* Laboratory investigation; a journal of technical methods and pathology, 2017. **97**(12): p. 1508-1515.
258. Christgen, M., et al., *The region-of-interest size impacts on Ki67 quantification by computer-assisted image analysis in breast cancer.* Human pathology, 2015. **46**(9): p. 1341-1349.
259. Webster, J.D. and R.W. Dunstan, *Whole-slide imaging and automated image analysis: considerations and opportunities in the practice of pathology.* Veterinary pathology, 2014. **51**(1): p. 211-223.

Chapter 2

Rational and Aims

Although it is steadily declining in incidence, gastric cancer (GC) imposes a global health burden accounting a considerable morbidity and mortality worldwide. GC patients tend to experience symptoms late in the development of the disease, delaying the diagnosis at an initial stage and narrowing their treatment options. Therefore, it is crucial to explore the key molecular alterations underlying GC pathogenesis and investigate their clinical usefulness to improve GC patients' outcome.

E-cadherin loss of function (LoF), leading to defective protein expression, is the most well-established alteration in GC initiation and progression, being the most frequent inherited cause of a GC-associated syndrome and deeply involved in GC invasion and metastases. Overexpression of CD44v6, besides promoting cancer cell self-renewal, has been associated with increased invasion, desmoplasia and tumor dissemination. Our previous observations showed that aggressive inherited *CDH1*-deficient cancers expressed *de novo* CD44v6 in regions without E-cadherin expression. Although the role of E-cadherin LoF is vastly studied in GC, its relationship with the *de novo* expression of CD44v6 is poorly explored.

One of the most broadly established approaches to study the clinical implications of specific molecular alterations is based on the visual examination of protein expression in tumor sections by a pathologist. Often, there is disagreement between pathologist evaluations due to some degree of subjectivity, complexing the conclusion of the analytical path. This gets even more complicated if cross-slide image analysis of multiple proteins is requested, and if strict sections of the slide need to be defined. Another layer of complexity is the recording of quantitative analysis by pathologists. Computer-aided methodologies emerge as a great ally to pathologists and research teams, as it objectively assesses, quantifies and records a large number of features in a systematic and high-throughput way, thus reducing diagnostic turn-around time and increased biomarker scoring reproducibility.

Most sporadic GCs lose E-cadherin expression and gain expression of CD44v6. Each of these events individually provide GC with particular aggressive features, but the simultaneous impact of these alterations in tumor aggressiveness and in patients' outcome remains to be explored. Also, the geographic variability in GC patients' outcomes worldwide related to clinical management of this disease remains as an unexplored layer. We believe that this asymmetry may shed light into the role of GC biomarkers, such as the two herein studied.

De novo CD44v6 expression is observed throughout the process of malignant transformation of gastric mucosa, from precursor lesions to advanced GCs. CD44v6-containing isoforms overall have been widely associated to drug resistance in several cancers, including GC. However, the cellular processes and the means by which CD44v6-containing isoforms or the CD44v6-exon itself condition response to therapy remains unknown.

Given the challenge in the analysis of defective patterns of E-cadherin protein expression in GC clinical samples, both by pathologists and computational tools, we set out to identify overexpressed proteins that may constitute surrogate markers for E-cadherin LoF in GC. Isogenic cellular models representing specific GC histopathological subtypes (diffuse and intestinal) and different modes of E-cadherin inactivation have not been so far reported.

Considering the aforementioned knows and unknowns, this thesis was dedicated to explore the role and clinical implications related to two molecular alterations deeply involved in GC pathogenesis (E-cadherin LoF and CD44v6 overexpression, together or individually), with emphasis in tumor aggressiveness features and response to therapy.

In this work, we have addressed the following specific aims:

- To establish an image analysis pipeline to automatically align immunohistochemical slides and quantify E-cadherin and CD44v6 co-expression.
  Original Paper 1, Chapter 3

- To unveil whether co-occurrence of E-cadherin LoF and CD44v6 overexpression in tumors, conditions disease aggressiveness and patient´s outcome, by comparing two contemporary GC patients' cohorts, from Western Europe and East Asia, submitted to distinct clinical approaches.
  Original Paper 2, Chapter 3

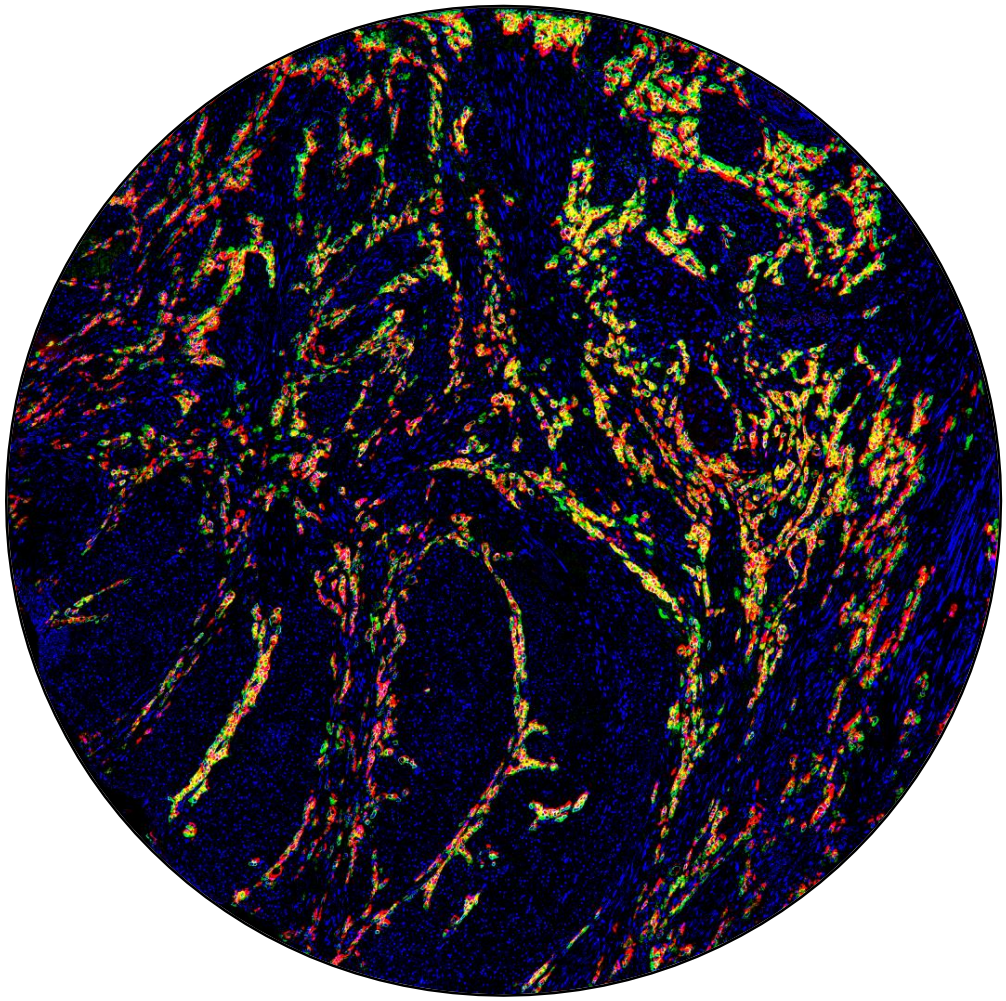
- To explore whether *de novo* CD44v6 expression has a central role in the transformative process and/or progression of GC, if it conditions response to therapy, or if it represents a bystander marker in any of these events.
  Original Paper 3, Chapter 4

- To investigate whether the presence of exon-v6 itself is the key responsible for conditioning the chemotherapeutic response in tumour cells overexpressing CD44v6-containing isoforms.
  Original Paper 4, Chapter 4

- To identify novel surrogate markers for E-cadherin LoF, through an integrative analysis of the proteomic content of E-cadherin depleted *in vitro* isogenic cell models.
  Original Paper 5, Chapter 5

Chapter 3

The Co-Operative Role of E-cadherin and CD44v6 in GC



E-cadherin **CD44v6** **Hematoxylin**

Original Paper 1

Towards automatic protein co-expression quantification in immunohistochemical TMA slides

Leslie Solorzano, **Carla Pereira**, Fátima Carneiro, Raquel Almeida, Gabriela M. Almeida, Carla Oliveira, Carolina Wälhby

IEEE Journal of Biomedical and Health Informatics (2020), DOI: 10.1109/JBHI.2020.3008821

ABSTRACT

Immunohistochemical (IHC) analysis of tissue biopsies is currently used for clinical screening of solid cancers to assess protein expression. The large amount of image data produced from these tissue samples requires specialized computational pathology methods to perform integrative analysis. Even though proteins are traditionally studied independently, the study of protein co-expression may offer new insights towards patients' clinical and therapeutic decisions. To explore protein co-expression, we constructed a modular image analysis pipeline to spatially align tissue microarray (TMA) image slides, evaluate alignment quality, define tumor regions, and ultimately quantify protein expression, before and after tumor segmentation. The pipeline was built with open-source tools that can manage gigapixel slides. To evaluate the consensus between pathologist and computer, we characterized a cohort of 142 gastric cancer (GC) cases regarding the extent of E-cadherin and CD44v6 expression. We performed IHC analysis in consecutive TMA slides and compared the automated quantification with the pathologists' manual assessment. Our results show that automated quantification within tumor regions improves agreement with the pathologists' classification. A co-expression map was created to identify the cores co-expressing both proteins. The proposed pipeline provides not only computational tools forwarding current pathology practices to explore co-expression, but also a framework for merging data and transferring information in learning-based approaches to pathology.

Co-expression, Computational pathology, Gastric cancer, Image analysis, Immunohistochemistry, Protein, Registration

INTRODUCTION

Gastric cancer (GC) ranks as the fifth most common cancer worldwide and the third most frequent cause of cancer-related mortality [1]. GC patients rarely experience symptoms at early stages of the disease, and more than 80% are diagnosed at an unresectable stage, where the 5-year survival rate is only 20% [2]. A better understanding of tumor heterogeneity and local molecular signatures is crucial to guide better clinical and therapeutic decisions, a fact that is true for many types of cancer [3] where the knowledge of underlying mechanisms is limited.

Immunohistochemical (IHC) analysis of tissue samples is the mainstream approach for diagnosis and therapeutic decision in solid cancers. Pathologists are frequently tasked with the cross-slide analysis of consecutive tissue sections stained to highlight particular features of the tissue or to study a new protein of interest. Despite robust guidelines, IHC is often limited by the subjectivity associated with qualitative visual interpretation of expression levels. In this matter, computational pathology can be an ally, as it objectively assesses, quantifies and relates a large number of features in a systematic and high throughput way.

Advanced commercial and open-source software is starting to gain traction in the world of computational pathology and making its way into the clinical practice. One example is QuPath [4], designed for whole slide image analysis. With the help of BioFormats [5], QuPath is capable of opening and reading the most common slide formats and allows for flexible and efficient tissue analysis, including, e.g., annotation and automated cell detection and counting by pathologists. QuPath can also be programmed via custom macros which can be useful to extract data for further processing. However, QuPath does not include registration tools, making it impossible to quantify co-expression in data coming from different slides and spaces.

Python's scientific computing ecosystem composed of core numeric libraries such as Numpy and Scipy, and domain specific libraries such as scikit-image and scikit-learn provide interfaces to perform image processing and make available several machine learning models that can be used for pixel classification i.e. segmentation.

In order to quantify protein co-expression, we developed and present here an image analysis pipeline to perform image registration on tissue microarray (TMA) slides, define tumor regions and quantify protein co-expression, built on such libraries. As proof of concept, we explored two proteins frequently altered in GC, namely E-cadherin [6] and CD44v6 [7]. Each protein was characterized by IHC, in separate but consecutive TMA sections, distanced a few micrometers apart. Therefore, tissue morphology is expected to be very similar between both tissue sections.

Recent developments in multispectral imaging and multiplex immunohistochemical staining makes it possible to study co-occurrence of several markers in the same tissue core [8]. These methodologies present multiple pre-analytical and analytical challenges such as the lack of standardization of fixation times, dehydration, paraffin impregnation and the cross-reactivity between multiple antibodies. Additionally, the availability of these systems is still very limited in terms of scalability and high throughput. The presented study is focused on well-established protocols commonly used in research and clinical practice.

IHC usually combines two stainings: haematoxilyn (H) and diaminobenzidine (DAB). H stains the general scaffold of the tissue in blue/purple, mainly highlighting nuclei morphology. DAB produces a brown precipitate, revealing the expression status of a protein targeted by a specific antibody (i.e. subcellular location, intensity and expression pattern). The morphological features revealed by H can be used as input for automated spatial alignment of multiple parallel tissue sections, including alignment quality assessment, as previously shown in [9]. This motivates the alignment of the cores. To do so, we used a recent registration framework [10] that takes advantage of the spatial information carried by H and also the intensity information of the unmixed stains.

However, quantifying co-expression patterns from TMA slides presents additional challenges in the management of gigapixel images. TMA images are stored in resolution pyramidal formats that require specialized programming libraries. We therefore developed custom tools for slide visualization, registration, and analysis within the scope of the open source TissUUmaps project [11]. TissUUmaps is built on top of OpenSeadragon [12], meaning that an arbitrary number of slides can be visualized and annotated in parallel, on a browser, which enables easy collaboration and sharing of information. The pipeline is modular and built with open-source tools, and modules can be changed to suit the tools used in any already existing pipelines.

Overall, using our image analysis pipeline, we were able to quantify protein co-expression in TMAs that can hopefully unveil novel insights in the clinical and molecular impact of multiple protein markers.

PREVIOUS WORK

The workflow presented in this paper involves several image analysis methods developed over the years. These are introduced in the following section while the methods section provides the details of the final workflow.

1. Color unmixing

Bright-field tissue slide scanning outputs an RGB image, where each channel represents the absorption of each of the three colors (red, green and blue). In consequence, the resulting TMA slide has to be processed to find, separate and quantify the protein of interest, i.e., provide a way to find how much contribution in an RGB channel is given to the colors of the desired stains. The relationship between the concentration of a stain and its absorbance or optical density depends on the chemical and physical properties of the stains. There is an ongoing debate on the interpretation of a pixel intensity and saturation with respect to the amount of stain [13]. This means that while intensities provide information about the presence of stain, they might not be linearly related with the amount of stain it represents. This is important to remember when considering a pixel as positive or negative for a certain stain. There are specially hard cases when the DAB stain is very faint and there can be a high discrepancy between decisions.

Color unmixing (also referred to as color deconvolution) [14], is the process of finding a matrix of optical densities or absorbances from samples of the image. Several ways exist to do color unmixing [15, 16]. The goal is always to find a way to transform from the RGB color representation to a space spanned by the stains in the tissue.

2. Image registration

Image registration is the process of spatially aligning two different images. We need to align the core images so that their corresponding features are aligned in space, enabling analysis of protein co-expression. After alignment, structures composed by cells in the tissue will overlap and the images will be considered spatially aligned. The first image is referred to as the reference, it is in a reference space. The second image is referred to as the floating image. Once registered, all the main features, the tissue scaffold, will be aligned.

There are different types of transformations, features, measures and evaluation methods for different registration methods and they are always application dependent; there is no one size-fits-all. There are linear and non-linear transformations. Linear transformations have a smaller number of parameters and a smaller search space with fewer local optima and are usually easier to interpret physically [17]. Non-linear transformations, also called elastic or deformable transformations, have a very high number of parameters that need regularization terms to achieve any kind of physical or anatomical interpretability.

Different types of registration have been used in high resolution tissue images, and registration examples with feature extraction include [18-20], while [21-23] exemplify non-linear approaches. A detailed study of registration of consecutive tissue sections can be found in [24] where linear transformations are performed on images of the same slide (washed and re-stained). The same study shows that much of the alignment in the pathology field is done manually which is arduous and inaccurate.

There are several automated registration methods based on Mutual Information, both as a measure to guide the registration and as an evaluation metric [25, 26]. However, these methods do not make full use of the spatial information. A recent registration framework uses a measure that combines intensity and spatial information in the same regularization term [10], typically exhibiting a larger region of convergence, increasing the chance of finding the correct transformation parameters under larger rotations, translations and scaling.

More formally, the spatial alignment of two images is modelled as a minimization process where the distance between the overlapped images is the main criterion. This distance d is measured iteratively as several transformations from a valid set Ω are applied to an image A to match its counterpart B. It is formulated as:

$$\hat{T} = \underset{T \in \Omega}{\arg min}\, d\,(T\,(A), B))$$

For affine transformations, Ω includes rotations, translation, scaling, shearing, and reflections. The distance d combines both intensity and spatial information, with the benefit of few local optima as compared to intensity based approaches, and no need for feature matching as compared to feature based methods.

3. Tumor segmentation

TMA slides represent pieces of a transversal section in a tumor sample. Tumor samples are three dimensional structures composed of tumor cells but also stromal non-tumor components. Therefore, regions highly enriched with tumor cells within TMA cores must be found. As manual labeling is labor intensive, automatic segmentation with the aid of fewer to no labels is preferred. Different tumors exhibit different morphological features depending on the histological type. With daily practice, pathologists learn to identify tumor-specific features and interpret them under a specific clinical context. In image analysis, tumor segmentation becomes a problem of pixel classification where each pixel is assigned to a class e.g. tumor, non-tumor, and background.

To distinguish tumor vs. non-tumor cells, the pathologist often evaluates the H staining that highlights cell nuclei features. More often H&E staining is used to locate tumor. Visually speaking, cancer cell nuclei are recognized by a decreased staining intensity, while normal cells display an homogeneous purple color, which translates to textures in image analysis. A pathologist can then manually label a few pixels representing different cells types and this information can be used in any of a great number of pixel classifiers. Not all the DAB staining is biologically meaningful, even when disregarding technical artifacts. In specific, normal epithelial cells of the stomach lack CD44v6 expression, but it is expressed in premalignant and malignant lesions of gastric mucosa. On the other hand, E-cadherin is expressed at the membrane of normal epithelial cells. The loss of its function, either complete or aberrant protein expression, reflects a pathological condition in the gastric mucosa. Therefore, it is crucial to perform tumor segmentation to ensure a reliable automated estimation of protein expression in tumor situ.

There are both classical and modern learning-based methods that allow for pixel classification, from graphical models, thresholding, K-means, naive Bayes, support vector machines, decision trees, random forest, to deep learning, as reviewed in [27]. One example of pixel classification for TMAs can be found in [28], where tumors in lung tissue are detected using Markov Random fields.
Random forests (RF) are a very powerful ensemble method that creates several decision trees (hence the forests) and these trees make a decision together. All trees try to find the best features that represent the data, which makes them an ideal choice for pixel classification. RF are the default choice in the interactive Ilastik software [29].

4. Protein co-expression

The spatial overlap of two pixels (protein 1 and protein 2) allows to infer the co-expression of both proteins. Consecutive TMA slides in close proximity are expected to represent the same structures. This proximity allows the quantification of colocalization to be interpreted as co-expression within a

level of certainty. Co-expression can then be assessed by finding the distances between the nearest pixels that best represent the protein. But even if proteins are spatially distributed over a similar region, some co-expression could be the result of random overlap, specially from pixels with lower intensities.

The work in [30] has been crucial for establishing the mathematical framework for colocalization quantification borrowing many elements from statistics. Originally used to quantify colocalization in fluorescence microscopy [31], colocalization is quantified using the Pearson Correlation Coefficient.

METHODS AND DATA

This section presents the proposed image analysis pipeline and an overview on the data analyzed. The presented pipeline is based on open source-source code, it is modular, and each module can be implemented according to the tools available to any laboratory searching to incorporate computational pathology pipelines. Detailed video tutorials and an implementation of the pipeline are available at github.com/wahlby-lab/TMAstudies.

1. Protein IHC image data from a GC patient cohort

Tumor samples from a cohort of gastric adenocarcinoma patients surgically treated between January 2008 and December 2014 at Centro Hospitalar Universitário de São João (CHUSJ, Porto, Portugal). A TMA was assembled from paraffin embedded tumor material of these patients. Representative areas of the tumors were selected based on H&E stained tissue specimens and one tissue core with 2 mm diameter was obtained from each selected specimen.

Nine TMA blocks were constructed and in each block a normal gastric mucosa core and a GC cell line core were included as controls, as well as a core of a non-gastric tissue sample for orientation purposes. IHC staining of E-cadherin and CD44v6 was performed in 3-micrometer TMA sections, in close proximity. The assay was carried out on the automated Ventana BenchMark ULTRAStaining System, using the OptiView DAB IHC Detection Kit (Roche/Ventana Medical Systems, Tucson, AZ, USA). H staining is used to reveal all tissue. Positive and negative staining controls were performed in parallel with paraffin sections. To monitor Ecadherin normal expression, either normal gastric mucosa, intestinal metaplasia or cell lines included at the TMA, while normal skin was used to control CD44v6 normal expression. Each TMA slide was scanned using NanoZoomer 2.0 HT (Hamamatsu) whole slide imaging scanner. TMA image sizes range between 150,000 pixels in width and 100,000 pixels in height, resulting in sizes of 5000 pixels^2 for each core with a average resolution of 0.226 micrometers/pixel. Overall, our image data set sums up cores from 18 TMA slides (9 for each protein) and the respective pathologist grade per TMA core, per protein. After excluding patients with no clinical-pathological or treatment data available, patients lost to follow-up and without available paraffin-embedded material for analysis, in total, 261 cases were eligible for analysis.

One expert pathologist graded all cores for one of the proteins, while another expert pathologist independently graded all cores stained for the other protein. In both cases one of four classes (from

0 to 3+) were assigned based on the estimated tumor area and the expression pattern. Due to the distinct biological context of each protein, each class represents different criteria. Specifically, CD44v6 is not expressed in normal tissue but becomes expressed in pre-malignant and malignant cells, and exerts its function at the membrane (Table I). On the other hand, E-cadherin is expressed at the membrane of normal cells but, in a tumor context, is often abnormally expressed (cytoplasmic or aberrant membrane, i.e. dotted or incomplete membrane) or absent (Table II). In this dataset, a fraction of cases (118/261 (45%)) preserved membranous E-cadherin expression. As we are interested in exploring E-cadherin and CD44v6 dysfunction, cases preserving complete membranous E-cadherin expression were excluded and the remaining counterpart (142/261 (55%)) proceeded for analysis.

Table I. Pathologist classification scheme for CD44v6.

Score	Estimated tumor area coverage (%)	Expression pattern
0	0%	Negative or cytoplasmic
1+	<10%	Membrane
2+	10-50%	Membrane
3+	>50%	Membrane

Table II. Pathologist classification scheme for E-cadherin.

Score	Estimated tumor area coverage (%)	Expression pattern
0	0%	Negative
1+	<25%	Cytoplasmic or aberrant membrane
2+	25-50%	Cytoplasmic or aberrant membrane
3+	>50%	Cytoplasmic or aberrant membrane

2. Methods in Image Analysis

The pipeline modules of our work are shown in Figure 1, serving as a visual guideline to the outcomes of every module. Our methods begin by separating the core images into its main components and all the necessary information will then be extracted from these separated stains. The documentation for the API of the scientific python libraries used is available in their respective pages online. We present our methodology and the parameters used.

2.1. Color unmixing

Figure 1 shows a TMA slide containing multiple IHC-stained cores where the blue/violet color represents the *H* staining and the brown color represents the *DAB* staining of E-cadherin and CD44v6. Color unmixing was done separately for E-cadherin and CD44v6.

First, representative patches were manually selected and RGB values were grouped into 255 clusters by k-means clustering. These 255 colors (representative of all RGB values expected from a given IHC staining) were thereafter sorted by their value (in HSV space), and colors of very high value were removed due to their closeness to the white and very light background colors. Starting from the color samples found in this reduced RGB space, an absorbance model was extracted as explained in [14]. Solving a least squares problem between the strongest absorbance and the remaining ones, creates the orthogonal basis M whose pseudo inverse is used to create gray scale images representing *H* and *DAB*. The result is two grayscale images, H_1 and DAB_1 for protein 1 (E-cadherin) and correspondingly two images H_2 and DAB_2 for protein 2 (CD44v6), as shown in Figure 1, where intensity corresponds to presence of the respective stains. M was created once for each of the two stains, and thereafter used in the full experiment.

2.2. Image registration

The H images from the cores carry information about the common structure of the core; a common scaffold, and are therefore used for image registration (while the DAB images may differ as they represent different proteins that may have different spatial distribution within the tissue). Using the recent registration framework proposed by [10], the main transformation between cores and its parameters is obtained by using H_1 and H_2 in equation 1:

$$\hat{T} = \arg\min_{T \in \Omega} d\,(T\,(H_1), H_2))$$

Where the distance *d* is Alpha-AMD, a distance that incorporates both intensities and spatial information to guide the registration.

The transformation *T* is chosen to be affine, meaning that Ω only includes scaling, rotation and translation. Since the aligned scaffolds represent very similar structures but not exactly the same, any elastic deformation could potentially create undesirable false structures. The transformation *T* is stored and applied to DAB_2 in order to bring it to the same space as DAB_1 where they can be effectively compared.

The Alpha-AMD method requires a few parameters to work. For the level of pyramids we use (32, 16, 8, 4) which means that the first computations for the registration will be done on an image 32 times smaller all the way until 4. We used gaussian blur with sigmas of (15.0, 8.0, 4.0, 2.0) which means that each of the previously mention levels will be blurred applying a gaussian filter with the selected sigmas. The method allows for the input of masks from where to sample so we input binary masks that cover the core. We use 300 itrerations for the registration.

We used CD44v6 as the reference image meaning that all the cores from E-cadherin were transformed to fit in the same space of the CD44v6 cores.
In general, we observed that the *H* images contain enough information to guide the transformation, particularly because the registration is robust enough to use the information coming from nucleus and cytoplasm in *H* despite them not being exactly equal in both images. The transformation *T* is stored and kept to be applied to any information that has to be sent to the CD44v6 space.

2.3. Registration confidence map

In order to trust quantification of co-expression from registered images, we must first be sure that similar tissue actually exists in all parts of the sample. Images I_1 and I_2 come from different slides which can present artifacts in different locations such as rips, folds, dust, noise, etc. When bringing I_2 to the space of I_1 all the artifacts have to be masked out. This means that information present in I_1 might be missing from I_2 and vice versa. If information is missing in one of the images, no attempt of quantifying co-expression should be done in the area.

For this purpose, a Registration Confidence Map (RCM) is created by finding the colocalization between max (H_1, DAB_1) and max (H_2, DAB_2) with artifacts masked out [9]. This maximum intensity projection (MIP) of both *H* and *DAB* from the same core indicates the presence of tissue. The RCM will then indicate the pixels that are trusted to provide information on expression and co-expression of proteins. In Figure 1, *H* and *DAB* are shown and the colocalization of the MIPs becomes the RCM, shown in the box with the same name. White regions represent regions with high confidence.

2.4. Tumor segmentation

An RF classifier was used to segment each core into tumor, non-tumor and background. Due to the lack of H&E staining we performed tumor segmentation on the CD44v6 image, and as the E-cadherin images are registered with CD44v6, we later applied the same tumor mask when quantifying both proteins. Sparse labeling was performed per category (background, tumor, non-tumor) manually drawn by the pathologist in 25 cores to train the RF classifier. This results in hundreds of thousands of features per class. Figure 1 shows an illustration of this process.

Using classical image processing, features are extracted from the pixel information under the labels. The features extracted are intensities, edges and texture information. In our case, the most meaningful features were Gaussian smoothing GS (10.0), Laplace of Gaussian LoG(3.5), Gaussian gradient magnitude GGM(10.0), Hessian Matrix Eigenvalues HME(1.6, 3.5, 10.0). GS features remove high frequencies and blur the image. GS is the input for LoG and GGM which find edges. HME serves as corner detector. The features are selected by their gini importance, or mean decrease in impurity (MDI) meaning how much they contribute to the purity of the tree. This is computed automatically along with the RF and the chosen features are those that have a MDI over the mean.

To train the RF we used the Scikit-Learn implementation version 0.22.1 from the “ensemble” module. From the RandomForestClassifier class we set the n estimators to 100 which is the number

of trees in the forest. We also set parameter warm start which allows for a continued training if needed, even after the script has finished.

Using the module “io” the images I and their corresponding labels are loaded. For each channel and each pixel from I that lies under a label, the modules “filters” and “features” compute GS, LoG, GGM and HME which are then organized in an array of size (n samples, n features) along with an array of size (n samples), which is need as input for the fit method from the classifier.

To perform tumor segmentation, features are extracted from of unlabeled pixels and classes (tumor, non-tumor, background) are predicted.

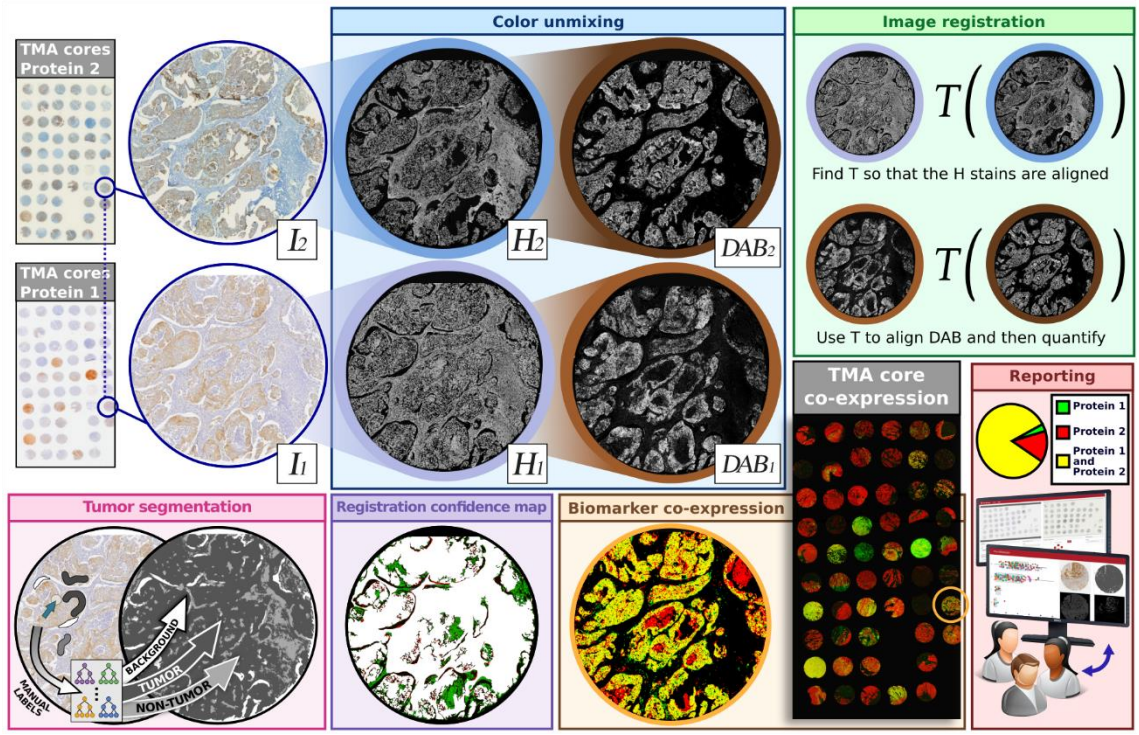


Figure 1. Overview of the stages to quantify co-expression in TMA cores. Two individual proteins stainings of the core images (I_1, I_2) are unmixed to reveal tissue scaffolding (H_1.H_2) and protein expression (DAB_1, DAB_2). Each pair of TMA cores expected to come from the a very near spatial location are registered using the *H* image to find the transformation *T*, which is applied to the *DAB* image. Using image *I* and manual labels, tumor segmentation is found in each core. A registration confidence map (RCM) ensures *H* stained tissue is present and correlates across the two cores (white = high correlation, green/red = tissue missing in H_1 or H_2). Only white regions are kept for analysis of co-expression. DAB_1 and transformed DAB_2 stains are superimposed and co-expression quantified using the tumor segmentation and registration confidence maps as masks.

2.5. Protein expression quantification

The developed image analysis methods to quantify protein expression are inspired by the pathologist’s method for visual assessment.

From a computational perspective, once stains are separated and tumor segmented, the images DAB_1 and DAB_2 contain the information of the protein expression. In a first approach, the expression

of *H* and *DAB* images was quantified by direct thresholding of the images. However, in [13] it is suggested that caution should be taken when quantifying protein expression directly from the unmixed images; they must undergo additional processing before quantification correlating with visual assessment can be done.

DAB stains appears in membranes as well as some other tissue structures, and while the strength of the staining in the membranes is variable, yet lower intensities count as much as strong ones at visual scoring, if the stain is located in a membrane. Inspired by the visual assessment method, we apply the Robert's cross edge filters on the *DAB* images and combine this information with the original *DAB* image. We let *DAB* be the input image and *E* be the Robert's cross edge information and construct the membrane image *B*:

$$B = \begin{cases} 2 \times DAB \times E & if\ E \leq 0.5 \\ 1 - 2 \times (1 - DAB) \times (1 - E) & if\ E > 0.5 \end{cases}$$

This means that weak edges, i.e. $E \leq 0.5$, take the information from reduced *DAB* intensities so that the *B* image will be darker (as $2 \times E$ is always less than 1.0). For pixels with strong edges, i.e. $E > 0.5$, the intensity is enhanced so that the B image will be close to 1 if *E* or *DAB* have strong values.

Next, we removed background noise by thresholding *B* at a fixed value (0.03) applied across both stains in all images. A fixed value, rather than automated thresholding, was applied to avoid enhancement of background in absence of signal. The resulting binary image represents the final presence/absence of the protein. The final percentages of tumor area coverage were calculated by finding the fraction of positive pixels in the part of each core fulfilling the RCM criteria and either including or excluding the tissue regions outside the tumor segmentation mask.

The protein co-expression was quantified for each pair of aligned cores fulfilling the RCM criteria and excluding the tissue regions outside the tumor segmentation mask. All overlapping pixels in the aligned DAB_1 and $T(DAB_2)$ images were aggregated and their total number over the number of pixels in the tumor returns the percentage of co-expression. An example a co-expression map can be observed in Figure 3.

RESULTS

In this study, we developed an image analysis pipeline to align and quantify co-expression of two proteins in GC TMA cores. Prior to this work, experienced pathologists visually inspected the expression profile (both extension and pattern) of E-cadherin and CD44v6 and grouped cores in four categories (from 0 to 3+), called scores.

To evaluate the reliability of our co-expression quantification, we first evaluated CD44v6 and E-cadherin expression individually. Our image analysis pipeline is inspired by the pathologists classification methodology, where focus is on expression within tumor areas.

The box plots in Figure 2 compare the manual and automated classification of two proteins in 142 cores, prior to and after tumor segmentation. Within each class, we can observe that the automated measurements are widely distributed for both proteins. To relate manual and automated

classifications, we determine the quartiles from the data and calculated thresholds based on the weighted medians between the groups. These thresholds allow the categorization of the predicted percentages and compare classes with the pathologists' classification as shown in the confusion matrices. To evaluate the quantification of CD44v6 and E-cadherin expression, we present confusion matrices to compare the number of cases assigned to each class according to manual versus automated classification. Classes are imbalanced so we colored each cell to reflect the percentage of cores within the predicted classes with respect to the pathologist class.

Before tumor segmentation, the predicted amount of CD44v6 and E-cadherin expression was underestimated for all the scores which makes the discrimination of classes difficult.

After tumor segmentation, the discrimination between classes improved for both proteins. We observed that our method performed better quantifying *de novo* CD44v6 expression, particularly when CD44v6 membranous expression was present in the majority (> 50%) of tumor cells (class 3+, 41 out of 64 ~ 64%). In contrast, predicting E-cadherin aberrant expression was not a trivial task for the computer, nor was it for the pathologists, due to the difficulty of estimating multiple of patterns. Nevertheless, our method performed well in classifying E-cadherin complete loss of expression (class 0, 9 out of 12 ~ 75%), which are often regarded the most aggressive GC tumors.

The comparison between the pathologists' classification and the computer prediction gives a Cohen's Kappa of 0.362 for CD44v6 and 0.241 for E-cadherin within tumor areas. The values are reduced when tumor segmentation is not used, resulting in 0.355 for CD44v6 and 0.236 for E-cadherin. While the interpretation of Cohen's Kappa is somewhat ambiguous, numbers higher than 0.2 indicate some measure of agreement. We can argue that unbalanced distribution across the classes usually results in lower Kappa statistics.

Considering that ambiguous cases can be challenging for either pathologist or computer, we developed a core explorer visualization tool for IHC data. We performed a depth analysis of four cases (A and B for CD44v6 and C and D for E-cadherin) and identified a few sources of discrepancies between pathologists and computers assessment.

For case A, given the absence of CD44v6 staining, the pathologist scored 0. However, the computer scored it as a 3+ because, during color unmixing step, very light browns had to be included (which can be confused with light purples/blues) resulting in the overestimation of the amount of DAB staining. In case B, our method scored a lower percentage of CD44v6 compared to the pathologist classification of 2+. Even though CD44v6 is present at the cell membrane, its expression is focal and faint which results in an underestimation of DAB staining.

Case C was classified by the pathologist as 0, but our method reports a significant amount of E-cadherin. Even though DAB staining is visibly present at the membrane, it belongs to a non-tumor region which was incorrectly considered as tumor. Improving tumor segmentation would clear this problem. In case D, E-cadherin protein expression was aberrant in more than 50% of the cells and thus the pathologist classified it as 3+. Despite its vast extention, the pattern of E-cadherin aberrant expression is dotted and incomplete membranous, which resulted in the underestimation of DAB staining predicted by the computer, scored 2+.

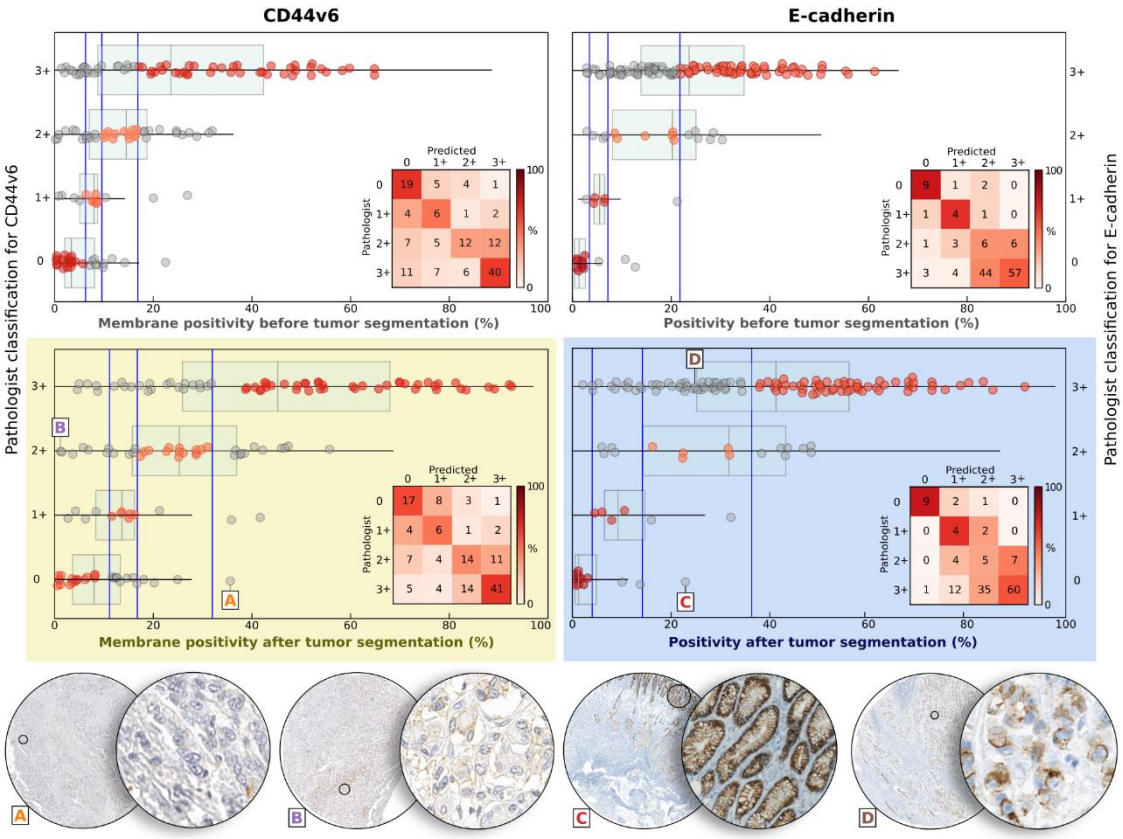


Figure 2. Comparison of manual and automated classification of E-cadherin and CD44v6 per core. Confusion matrices show manual classification score vs. automated classification score based on the weighted median of the continuous automated classification results. Colors in the matrices reflect the percentage of cases in the predicted classes with respect to the pathologist class. In the bottom area, the yellow background highlights the classification of CD44v6 within tumor and the blue background highlights E-cadherin classification within tumor. Four cases (A, B, C and D) illustrate some examples of cores that have been misclassified due to complex expression patterns, less accurate tumor segmentation, or poor color unmixing.

All the TMA slides were processed to create co-expression maps for visual assessment of the relationship between CD44v6 and E-cadherin expression. In the majority of the cases, core alignment was successfully achieved and the morphological structures matched. Despite thin TMA slices, protein co-localization with a single cell resolution is not achievable since subcellular details are not shared between slides. Nonetheless, the overlap of both proteins brings important insights regarding the heterogeneity degree of a tumor core. In Figure 3 we highlight three cores which illustrate the level of heterogeneity that a single core may harbor. For instance, the first core shows a unique and homogeneous population that co-expresses both proteins, while the other cores harbor multiple populations, the majority expressing one of both proteins and others that co-express both.

At a glance, this type of map helps the quick identification of cores co-expressing both proteins and the assessment of the level of tumor heterogeneity that was previously challenging to do by visual assessment. For example, to evaluate protein co-expression, the pathologist starts by visually identifying the tumor areas in the core and then evaluates the pattern for each individual protein in the tumor area. Only after that, the pathologist displays two separate images side by side, mentally

divides the image in quarters and starts his/her estimation of expression level of both proteins simultaneously. On average, the manual assessment of concomitant protein expression takes between 5-8 minutes per step. Even with the support of another operator/researcher, it is unlikely that all the information mentioned by the pathologist during his/her assessment is properly documented. In contrast, our pipeline processes one slide containing 60 cores in 30 minutes (30 seconds per core), saving all the information processed in each step of the analysis. Therefore, full images, tumor segmentation, protein intensity and patterns are documented and can be revisited and re-analyzed at any time. Note that our pipeline is not an end-to-end software product; each module of the pipeline should be executed independently which gives the opportunity to verify the result of each step. The clinical value of the information has to be evaluated and verified by an expert, and further studies have to be carried out to relate the measurements to relevant biological and clinical information.

Research is moving away from categorical scores to continuous scores such as the results of our pipeline. Either way, we believe that the contribution of pathologist and computer should not be independent of each other but complementary.

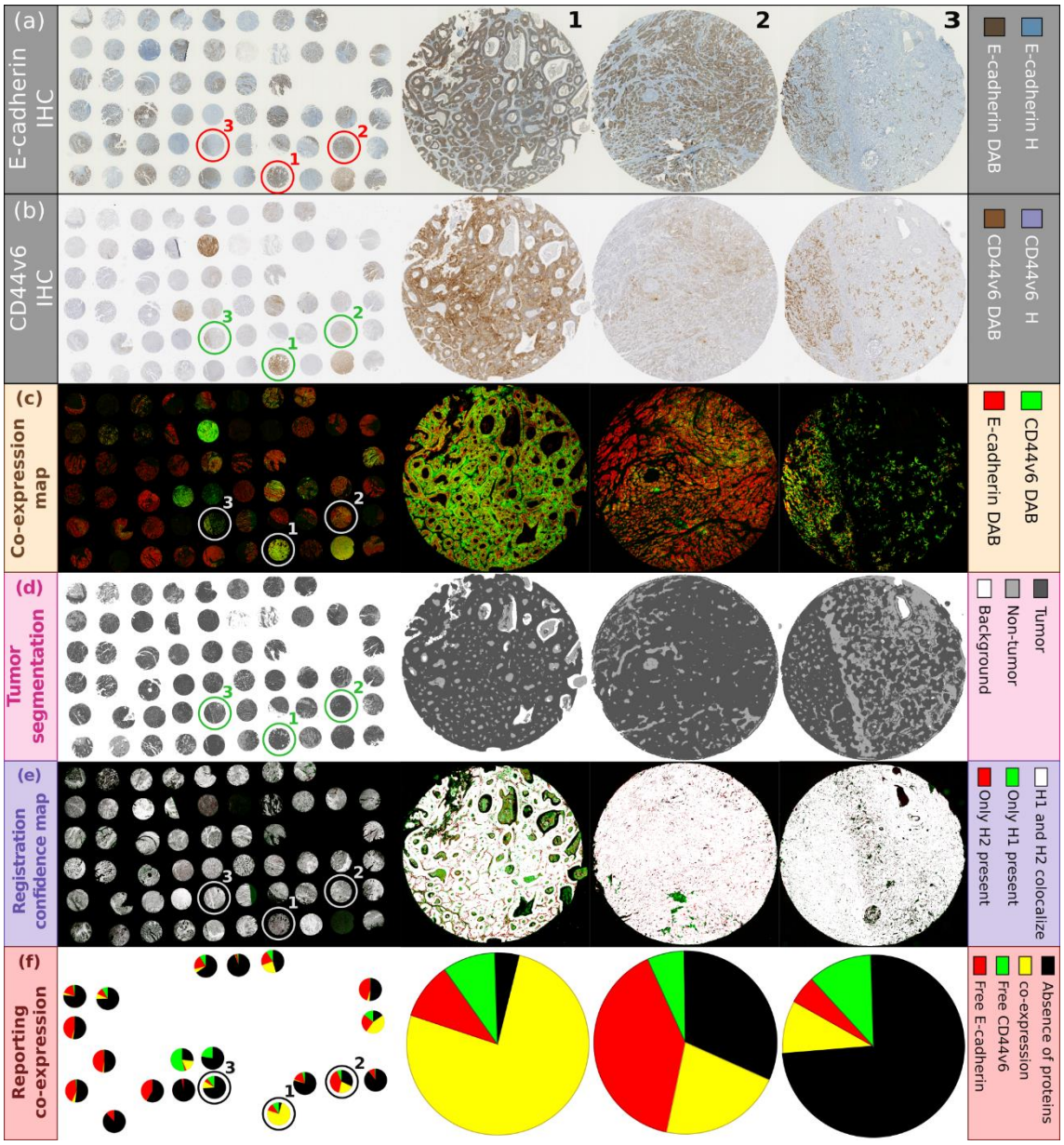


Figure 3. Overview of the output from each of the steps to quantify and visualize E-cadherin, CD44v6, and their co-expression. (a, b) Original core array for E-cadherin and CD44v6. (c) Co-expression map showing E-cadherin in red and CD44v6 in green. (d) Tumor segmentation, darkest gray shows tumor. (e) The registration confidence map (RCM) highlights tissue colocalization zones in white. (f) Overview of co-expression (yellow), free proteins (free E-cadherin in red, free CD44v6 in green), and absence of protein (black) within tumor and RCM. The cores preserving E-cadherin membranous expression or without available paraffin-embedded material were excluded from the co-expression quantification analysis.

DISCUSSION AND FUTURE WORK

Classifying tissue automatically is never a trivial task. We have presented here an initial approach with a methodology that may be improved by using more advanced components and a continuous feedback from pathologist to researcher. Also, an evaluation that includes intra and inter pathologist variation would give a better recognition of difficult-to-grade cases and more fairly define faults of the computational approach.

Faint and misty protein stainings are one of the most frequent sources of equivocal information that lead to mistakes in the subsequent steps of the automated method. In the future, there is room to improve the performance of our method by refining the distinction of multiple aberrant patterns that E-cadherin can assume in the tumor nest and explore the intensity patterns of a protein.

The computational quantification may be improved by better tumor segmentation, and taking into account the expression patterns of both proteins. Our method reports better results for CD44v6 compared to E-cadherin, because its patterns are less complex and simpler to quantify.

Data analysis is tightly coupled with good data visualization where efficient visualization methods are able to summarize multiple facets of data and can bring new insights. Our process and interface with the pathologists was improved by developing a core explorer visualization tool for IHC data to explore cores according to their pathologist given grade and the percentage calculated with our method. The information generated throughout the computer quantification process (as color unmixing and tumor segmentation) is displayed in the interface. At a glance, researchers and pathologists can quickly observe the distribution of the cores, inspecting the outliers or any technical mistake or bias in a specific slide. Inspecting the cores was crucial to understand the difficulties of the pipeline and even help with the correction in a few scores.

Beyond the scope of this paper, we can ask new questions such as: can the co-expression maps be developed further for the study of multiple markers? Can the present method be used to reveal co-localization at the subcellular level? Can any particular co-expression patterns within the tumor be used as reference for analysis of other markers? If so, they can become labels for other pieces of tissue aligned to the same context.

We believe that the proposed pipeline for registration and quality control for TMA data could be highly valuable in learning-based approaches to explore the clinical and molecular impact of multiple protein markers and improve our knowledge in intra- and inter-tumor/patient heterogeneity. Given the variability associated with manual assessment and reporting of IHC results, the key strength of our method is to extract, register and easily document the analysis of multiple protein markers. Using our pipeline, researchers are able to run preliminary analysis in a high throughput manner, on the direction of their biological or clinical question. This strategy is expected to save pathologists time, while serving as the basis for faster verification, validation, adjustment and documentation of data. In future work, combining rare markers with common markers can provide invaluable and unbiased information for network training.

Acknowledgments

The authors thank Dina Leitão and Gilza Gonçalves for the technical support with TMA slides processing, Mário Seixas for technical support with TMA slides digitalization and Johan Öfverstedt for his help with the registration framework.

Funding

This work was financed by ERC CoG 682810 to C.W., FCT: SFRH/BD/113031/2015 to C.P., FEDER/COMPETE 2020/PORTUGAL2020/POCI/FCT: Ref. POCI01-0145-FEDER-007274, 030164, and 016390 to C.O., and Project NORTE-01-0145-FEDER-000029 to C.O.

REFERENCES

1. Bray, F.;Ferlay, J.;Soerjomataram, I.;Siegel, R.L.;Torre, L.A.;Jemal, A. Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA Cancer J Clin.* **2018**, 68, 394-424, DOI: 10.3322/caac.21492
2. Smyth, E.C.;Verheij, M.;Allum, W.;Cunningham, D.;Cervantes, A.;Arnold, D. Gastric cancer: ESMO Clinical Practice Guidelines for diagnosis, treatment and follow-up†. *Annals of Oncology.* **2016**, 27, v38-v49, DOI: https://doi.org/10.1093/annonc/mdw350
3. Taylor, C.R., *Introduction to Predictive Biomarkers: Definitions and Characteristics*, in *Predictive Biomarkers in Oncology: Applications in Precision Medicine*, S. Badve and G.L. Kumar, Editors. 2019, Springer International Publishing: Cham. p. 3-18.
4. Bankhead, P.;Loughrey, M.B.;Fernández, J.A.;Dombrowski, Y.;McArt, D.G.;Dunne, P.D.;McQuaid, S.;Gray, R.T.;Murray, L.J.;Coleman, H.G., et al. QuPath: Open source software for digital pathology image analysis. *Scientific Reports.* **2017**, 7, 16878, DOI: 10.1038/s41598-017-17204-5
5. Linkert, M.;Rueden, C.T.;Allan, C.;Burel, J.-M.;Moore, W.;Patterson, A.;Loranger, B.;Moore, J.;Neves, C.;MacDonald, D., et al. Metadata matters: access to image data in the real world. *Journal of Cell Biology.* **2010**, 189, 777-782, DOI: 10.1083/jcb.201004104
6. Carneiro, P.;Fernandes, M.S.;Figueiredo, J.;Caldeira, J.;Carvalho, J.;Pinheiro, H.;Leite, M.;Melo, S.;Oliveira, P.;Simões-Correia, J., et al. E-cadherin dysfunction in gastric cancer--cellular consequences, clinical applications and open questions. *FEBS Lett.* **2012**, 586, 2981-2989, DOI: 10.1016/j.febslet.2012.07.045
7. da Cunha, C.B.;Oliveira, C.;Wen, X.;Gomes, B.;Sousa, S.;Suriano, G.;Grellier, M.;Huntsman, D.G.;Carneiro, F.;Granja, P.L., et al. De novo expression of CD44 variants in sporadic and hereditary gastric cancer. *Lab Invest.* **2010**, 90, 1604-1614, DOI: 10.1038/labinvest.2010.155
8. Mezheyeuski, A.;Bergsland, C.H.;Backman, M.;Djureinovic, D.;Sjöblom, T.;Bruun, J.;Micke, P. Multispectral imaging for quantitative and compartment-specific immune infiltrates reveals

distinct immune profiles that classify lung cancer patients. *J Pathol.* **2018**, 244, 421-431, DOI: 10.1002/path.5026

9. Solorzano, L.;Almeida, G.M.;Mesquita, B.;Martins, D.;Oliveira, C.;Wählby, C. *Whole Slide Image Registration for the Study of Tumor Heterogeneity*. 2018. Cham: Springer International Publishing.
10. Öfverstedt, J.;Lindblad, J.;Sladoje, N. Fast and Robust Symmetric Image Registration Based on Distances Combining Intensity and Spatial Information. *IEEE Transactions on Image Processing.* **2019**, 28, 3584-3597, DOI: 10.1109/TIP.2019.2899947
11. Solorzano, L.;Partel, G.;Wählby, C. TissUUmaps: interactive visualization of large-scale spatial gene expression and tissue morphology data. *Bioinformatics.* **2020**, 36, 4363-4365, DOI: 10.1093/bioinformatics/btaa541
12. OpenSeadragon: An open-source, web-based viewer for high-resolution zoomable images, implemented in pure JavaScript Available online: https://openseadragon.github.io. (acessed on
13. van der Loos, C.M. Multiple immunoenzyme staining: methods and visualizations for the observation with spectral imaging. *J Histochem Cytochem.* **2008**, 56, 313-328, DOI: 10.1369/jhc.2007.950170
14. Ruifrok, A.C.;Johnston, D.A. Quantification of histochemical staining by color deconvolution. *Anal Quant Cytol Histol.* **2001**, 23, 291-299, DOI:
15. Wemmert, C.;Krüger, J.M.;Forestier, G.;Sternberger, L.;Feuerhake, F.;Gançarski, P. *Stain unmixing in brightfield multiplexed immunohistochemistry*. in *2013 IEEE International Conference on Image Processing*. 2013.
16. van Der Laak, J.A.;Pahlplatz, M.M.;Hanselaar, A.G.;de Wilde, P.C. Hue-saturation-density (HSD) model for stain recognition in digital images from transmitted light microscopy. *Cytometry.* **2000**, 39, 275-284, DOI:
17. Zitová, B.;Flusser, J. Image registration methods: a survey. *Image and Vision Computing.* **2003**, 21, 977-1000, DOI: https://doi.org/10.1016/S0262-8856(03)00137-9
18. Trahearn, N.;Epstein, D.;Snead, D.;Cree, I.;Rajpoot, N., *A fast method for approximate registration of whole-slide images of serial sections using local curvature*. SPIE Medical Imaging. Vol. 9041. 2014: SPIE.
19. Sarkar, A.;Yuan, Q.;Srinivas, C. *A robust method for inter-marker whole slide registration of digital pathology images using lines based features*. in *2014 IEEE 11th International Symposium on Biomedical Imaging (ISBI)*. 2014.
20. Cooper, L.;Sertel, O.;Kong, J.;Lozanski, G.;Huang, K.;Gurcan, M. Feature-based registration of histopathology images with different stains: an application for computerized follicular lymphoma prognosis. *Computer methods and programs in biomedicine.* **2009**, 96, 182-192, DOI: 10.1016/j.cmpb.2009.04.012
21. Mueller, D.;Vossen, D.;Hulsken, B. Real-time deformable registration of multi-modal whole slides for digital pathology. *Comput Med Imaging Graph.* **2011**, 35, 542-556, DOI: 10.1016/j.compmedimag.2011.06.006

22. Pichat, J.;Iglesias, J.E.;Yousry, T.;Ourselin, S.;Modat, M. A Survey of Methods for 3D Histology Reconstruction. *Medical Image Analysis.* **2018**, 46, 73-105, DOI: https://doi.org/10.1016/j.media.2018.02.004
23. Malandain, G.;Bardinet, E.;Nelissen, K.;Vanduffel, W. Fusion of autoradiographs with an MR volume using 2-D and 3-D linear transformations. *Neuroimage.* **2004**, 23, 111-127, DOI: 10.1016/j.neuroimage.2004.04.038
24. Moles Lopez, X.;Barbot, P.;Van Eycke, Y.-R.;Verset, L.;Trépant, A.-L.;Larbanoix, L.;Salmon, I.;Decaestecker, C. Registration of whole immunohistochemical slide images: an efficient way to characterize biomarker colocalization. *Journal of the American Medical Informatics Association : JAMIA.* **2015**, 22, 86-99, DOI: 10.1136/amiajnl-2014-002710
25. Pluim, J.P.W.;Maintz, J.B.A.;Viergever, M.A. Mutual-information-based registration of medical images: a survey. *IEEE Transactions on Medical Imaging.* **2003**, 22, 986-1004, DOI: 10.1109/TMI.2003.815867
26. Pluim, J.P.W.;Maintz, J.B.A.;Viergever, M.A. *Image Registration by Maximization of Combined Mutual Information and Gradient Information*. in *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2000*. 2000. Berlin, Heidelberg: Springer Berlin Heidelberg.
27. Fernández-Delgado, M.;Cernadas, E.;Barro, S.;Amorim, D. Do we need hundreds of classifiers to solve real world classification problems? *J. Mach. Learn. Res.* **2014**, 15, 3133–3181, DOI:
28. Wang, C.-W. Robust Automated Tumour Segmentation on Histological and Immunohistochemical Tissue Images. *PLOS ONE.* **2011**, 6, e15818, DOI: 10.1371/journal.pone.0015818
29. Berg, S.;Kutra, D.;Kroeger, T.;Straehle, C.N.;Kausler, B.X.;Haubold, C.;Schiegg, M.;Ales, J.;Beier, T.;Rudy, M., et al. ilastik: interactive machine learning for (bio)image analysis. *Nature Methods.* **2019**, 16, 1226-1232, DOI: 10.1038/s41592-019-0582-9
30. Costes, S.V.;Daelemans, D.;Cho, E.H.;Dobbin, Z.;Pavlakis, G.;Lockett, S. Automatic and quantitative measurement of protein-protein colocalization in live cells. *Biophys J.* **2004**, 86, 3993-4003, DOI: 10.1529/biophysj.103.038422
31. Comeau, J.W.D.;Costantino, S.;Wiseman, P.W. A guide to accurate fluorescence microscopy colocalization measurements. *Biophysical journal.* **2006**, 91, 4611-4622, DOI: 10.1529/biophysj.106.089441

Original Paper 2

Comparison of East Asian and Western European cohorts explains disparities in survival outcomes and highlights predictive biomarkers of early-stage gastric cancer aggressiveness

Carla Pereira, Ji-Hyeon Park, Sofia Campelos, Irene Gullo, Carolina Lemos, Leslie Solorzano, Diana Martins, Gilza Gonçalves, Dina Leitão, Ana André, Carolina Wählby, Raquel Almeida, Woo Ho Kim, Fátima Carneiro, Han-Kwang Yang, Gabriela M. Almeida, Carla Oliveira

Ready for submission

ABSTRACT

Surgical resection with lymphadenectomy is the universal mainstay for curative treatment of gastric cancer (GC) patients with loco-regional disease. However, survival rates between Western- and Eastern-world regions remain asymmetric. Herein, we used this survival asymmetry as an opportunity to identify molecular biomarkers and clinical features of disease aggressiveness. Particularly, we explored E-cadherin loss-of-function (LoF) and *de novo* CD44v6 expression as frequent molecular alterations in GC. Clinicopathological and survival data was systematically collected for two geographically distinct surgical cohorts of chemo-naïve patients, 170 from Portugal (Western Europe) and 367 from South Korea (East Asia). E-cadherin and CD44v6 immunohistochemistry was performed in consecutive tissue microarrays sections and their expression profiles were correlated with patient survival and clinicopathological features. Survival analysis revealed a subset of 12.4% (21/170) of Western European GC patients displaying extremely poor overall survival, even at early disease stages, whose tumors presented concomitantly abnormal E-cadherin and very high CD44v6 expression. Early stage tumors from this patient subset were particularly aggressive compared to all tumors, invading deeper into the gastric wall (p=0.032) and more often permeating the vasculature (p=0.018) and nerves (p=0.009). A similar molecular profile was found in 11.9% (44/367) of East Asian patients, but the survival was significantly better. These early stage tumors permeated the lymphatic vessels (p=0.003), promoting lymph node (LN) metastasis (p=0.019). Better survival in the East Asian cohort could be explained by earlier cancer detection (8-years on average) and a more extensive LN dissection, which counteracted the consequences of nodal invasion. This comparative study supports concomitant E-cadherin LoF and CD44v6 overexpression as predictors of tumor aggressiveness in early stage GC, which may lead to poor outcomes if extensive surgery/lymphadenectomy is not performed.

E-cadherin, CD44v6, Lymphadenectomy; Immunohistochemistry; Tissue microarray

INTRODUCTION

Gastric cancer (GC) imposes a global health burden, ranking as the fifth commonest and third deadliest cancer worldwide [1]. Surgical resection with effective tumor margins clearance and adequate lymph node (LN) dissection is the universal curative treatment for loco-regional disease. Despite that, wide geographic variability in survival outcomes for GC exists worldwide [2].

Survival outcomes are largely dependent on whether the tumor is detected early on, whether it can be completely surgically removed, and whether it responds to chemotherapy (CT). In Asian countries, GC is diagnosed at earlier stages compared with non-Asian countries. Indeed, in some East Asian (EA) countries like Japan and South Korea, which present high GC incidence rates, routine screening strategies have been implemented to improve early detection of GC [3]. Such approach is missing in Western European (WE) countries with high GC incidence, and therefore most GC patients are diagnosed at advanced/unresectable GC, rendering considerably poorer survival rates [4, 5]. The extension of LN dissection (i.e. lymphadenectomy) also plays a decisive role in GC patients' outcome. This is more extensive in EA than in WE countries and has been a matter of extensive debate between European and Asian medical societies [6, 7].

Neoadjuvant CT for resectable GC was rapidly introduced in GC guidelines after the landmark MAGIC and FNLCC/FFCD clinical trials demonstrated the benefit of perioperative CT based on platinum/fluoropyrimidine combination over surgery alone [8, 9]. Due to a lack of consensus regarding the surgical approaches, particularly related to the lymphadenectomy extension, neoadjuvant CT was not accepted in EA countries as an alternative to their extensive surgical approach. Factually, and although neoadjuvant CT was proven to be a reasonable option, both WE and EA world regions continued using mainly post-operative/adjuvant chemotherapy to treat early stage GC patients. Only recently, neoadjuvant CT started to be gradually used as the standard strategy for GC treatment. As chemo-naïve GC patients are slowly disappearing, the opportunity to determine therapy-independent factors that modulate GC aggressiveness is also fading away. In particular, the analysis of retrospective cohorts of chemo-naïve GC patients creates the unique opportunity to study and further explore the pros and cons of EA and WE surgical strategies and their impact in patient survival.

Herein, we hypothesize that differences in survival outcomes in WE and EA GC patients lie beyond the multifactorial causes of GC in these world regions. To address this hypothesis, we studied chemo-naïve GC patients' cohorts from these world regions and looked for tumors sharing a specific molecular profile related to disease aggressiveness, but presenting dissimilar survival outcomes. Then we focused on identifying the clinicopathological factors that could explain those differences.

Classically, clinical staging is the cornerstone for determining treatment options and clinical outcomes [10]. However, GC is a heterogeneous and molecularly complex disease, challenging the usefulness of staging systems for this purpose, and the one-fits-all approach is generally used. Therefore, the use of molecular biomarkers for stratifying patients for treatment and outcome prediction is particularly relevant in GC. E-cadherin and CD44v6 are adhesion molecules, which loss-of-function (LoF) or *de novo* expression, respectively, recurrently appear associated to disease aggressiveness in several cancers, but particularly in GC. E-cadherin LoF, leading to defective

protein expression, is one of the most well-established alterations in GC initiation and progression [11]. While germline E-cadherin LoF triggers the development of Hereditary Diffuse Gastric Cancer syndrome, most sporadic GCs lose E-cadherin expression, at the somatic level, triggering loss of intercellular adhesion, enhanced migration, invasion and tumor dissemination [12-14]. Overexpression of CD44v6, beyond promoting cancer cell self-renewal, has been associated with increased invasion, desmoplasia and tumor dissemination [15, 16], Further, CD44v6 is *de novo* expressed in approximately 70% of sporadic GC cases [17]. Our previous observations showed that aggressive inherited *CDH1*-deficient cancers expressed *de novo* CD44v6 in regions without E-cadherin expression [17]. Despite the unequivocal role of these two molecules in GC aggressiveness, their prognostic value remains controversial in worldwide GC cohorts [18, 19], therefore we considered them adequate markers to address our hypothesis.

MATERIALS AND METHODS

1. Patient sample/data collection and tissue microarray (TMA) assembly

A cohort of GC patients surgically treated between January 2008 and December 2014 at Centro Hospitalar Universitário de São João (CHUSJ, Porto, Portugal) and part of the Tumor biobank of CHUSJ/Ipatimup was assembled [20]. Briefly, a TMA was prepared from paraffin-embedded tumor material of these patients, as follows. Representative areas of the tumors were selected on hematoxylin (H) and eosin stained sections and one tissue core with 2 mm diameter was obtained from each selected specimen and deposited into a recipient paraffin block using an Arraymold Kit A (IHC World, Woodstock, USA). All samples were from the “Tumor and Tissue Biobank” of the Pathology Department of CHUSJ, for which informed consent was obtained from patients or their legal representatives. Clinical-pathological, treatment and follow-up data was obtained from the Pathology and Surgical Departments of the CHUSJ. Overall survival (OS) was calculated from the time of surgery. This study was approved by the Institutional Ethics Committee of CHSJ (Ethics Committee references CES 122/15 and CES 117/18), and informed patient consent obtained.

An independent cohort of GC patients was collected between January 2010 and December 2011 at Seoul National University Hospital (South Korea). Representative tumor tissues (2 mm diameter) were taken from individual paraffin-embedded gastric tumors (donor blocks) and arranged in new recipient paraffin blocks (TMA blocks) using a trephine apparatus. Data on patient age, sex, clinical stage, and pathological parameters were retrieved from their medical records. The survival information was request from the Ministry of Security and Public Administration (Korea) according to institutional regulations. Local ethics approval was obtained (IRB: H1706-105-860 - Seoul National University Hospital).

Data on clinicopathological features were collected prospectively and analyzed retrospectively in both cohorts. In this study, GC staging was reported according to the 7th edition of the American Joint Committee on Cancer (AJCC) classification system.

2. E-cadherin and CD44v6 immunohistochemistry (IHC) of GC samples

Immunostaining targeted for E-cadherin (clone 24E10, 1:400 dilution for 32 min; Cell Signaling, Beverly, MA, USA) and CD44v6 (clone MA54, 1:400 dilution for 32 min; Invitrogen, Carlsbad, California, USA), was performed in consecutive TMA tissue sections, distanced a few micrometers apart (3-5µm). The staining was carried out on the automated Ventana BenchMark ULTRAStaining System, using OptiView DAB IHC Detection Kit (Roche/Ventana Medical Systems, Tucson, AZ, USA), according to the manufacturers' instructions. H staining was used to reveal all tissue. Positive staining controls were placed next to the TMA paraffin sections. Normal gastric mucosa and normal skin were used as positive controls for E-cadherin and CD44v6 expression, respectively. E-cadherin and CD44v6 IHC analysis was performed independently for each cohort, but using the same technical conditions.

3. E-cadherin and CD44v6 IHC analysis

E-cadherin and CD44v6 were individually assessed, each by two expert pathologists in a blind manner.

The extent and pattern of E-cadherin staining in tumor cells (absent; aberrant - cytoplasmic or dotted; incomplete membrane; complete membrane) was visually estimated in the tumor regions and registered according to their proportion in TMA cores. GC cases were subsequently grouped in two categories: 1) Normal, presence of complete membranous staining in >75% of the tumor cells; 2) Abnormal, presence of complete membranous staining in <75% of the tumor cells, or absent/aberrant/incomplete membranous staining in > 25% of the tumor cells (figure S1A, appendix pp 4).

The percentage of tumor cells displaying *de novo* membranous CD44v6 expression (either complete or focal) was visually assessed and classified in three categories: 1) Absent/Low, no staining at the cell membrane or membranous staining in < 50% of the tumor cells; 2) High, membranous staining in 50-75% of the tumor cells; 3) Very High, membranous staining in >75% of the tumor cells (figure S1B).

Taking advantage of the close proximity between tissue sections, an experienced pathologist characterized the multiplicity of populations existing in the tumor nest of the GC patient subset from WE-C presenting abnormal E-cadherin and very high membranous CD44v6 expression (n=21) which had a particular dismal survival. To evaluate concomitant protein expression, the pathologist displayed two separate images side by side, mentally divided the image in quarters and started his/her estimations of both proteins simultaneously in the same tumour areas. The contribution of each of four possible populations was visually estimated: 1) both proteins concomitantly expressed; 2 or 3) E-cadherin is expressed and CD44v6 is not, or vice-versa; and 4) none of the proteins is expressed.

4. Automated image registration of TMA cores

Each TMA slide was scanned using NanoZoomer 2.0HT (Hamamatsu Photonics, Japan) whole slide imaging scanner. Digital TMA images were converted in a multi-resolution pyramidal format and visualized in TissUUmaps interactive tool [21], where individual regions were drawn around the cores of interest. E-cadherin or CD44v6 DAB (3,3'-diaminobenzidine, brown) and H (blue/violet) double stained images were decomposed in single stainings based on the palette of colors characteristic of each TMA image. After color unmixing, two monochromatic images were generated for each protein, DAB_1 and H_1 for E-cadherin and DAB_2 and H_2 for CD44v6. Each pair of TMA cores was registered using the registration framework proposed by Öfverstedt and colleagues [22]. Herein, the *H* component was used to find an affine transformation matrix *T,* which is automatically determined by an iterative process (*i*=300). This matrix includes parameters to rotate, translate and scale a target image (H_1) with respect to a reference image (H_2) so that they overlap perfectly. Considering that the protein of study might not be expressed and thus there is no DAB staining, we applied the transformation *T* found using H_1 to the DAB_1 image. All the cores were stitched together and a new TMA map was generated with the transformed cores for the target protein. Transformed DAB_1 for E-cadherin (represented in red) and DAB_2 for CD44v6 (represented in green) TMA maps were merged using ImageJ software, where co-expression was illustrated in yellow, whenever both proteins overlap [23].

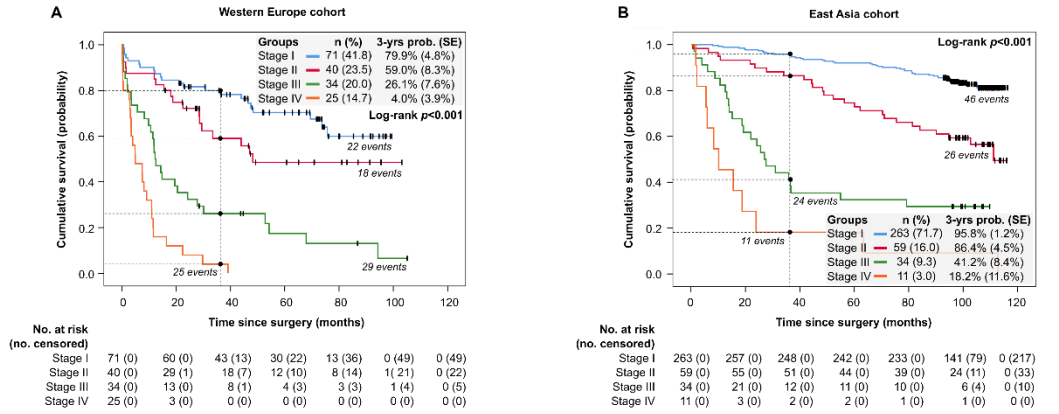
5. Statistical analysis

Categorical variables were described as absolute and relative frequencies, and continuous variables were described using mean values. Means for continuous variables were compared using Student's t-test or One-way ANOVA, with Tukey's Post Hoc test when appropriate. Categorical variables were compared using the Chi-square ($\chi 2$) test or Fisher exact test, as appropriate. Variables total no. of LNs and metastatic no. of LNs were analyzed using a non-parametric Mann-Whitney *U* comparison test. Kaplan-Meier estimates of overall survival (OS) were obtained between groups. Statistical significance was considered as 2-sided α of less than 0.05. Statistical analyses were performed using SPSS (Statistical Package for the Social Sciences, IBM Inc) version 26 for Windows or GraphPad Prism version 8.2.1 for Windows.

RESULTS

1. *East Asian GC patients exhibit superior survival outcomes in all disease stages, as compared to Western European GC patients*

We compared the clinical behavior and OS of chemo-naïve GC patients from two countries with high incidence of GC, Portugal representing WE (n=170) and South Korea representing EA (n=367) (table S1). As expected, OS significantly worsens as stage increases in both cohorts (figure 1A and B, table S2 and S3). We observed that the proportion of GC patients surviving 3-years post-diagnosis is higher in the EA cohort (EA-C), for all stages (figure 1B). Accordingly, the number of patients still alive 3-years post-diagnosis in stage I and II was significantly higher in the EA-C compared to the WE cohort (WE-C) (figure 1C). In stage I, this may be explained by the intensive screening initiatives implemented in EA countries that not only detect more cases at stage I (263/367 (71%) EA-C *vs.* 71/170 (42%) WE-C), but also at an earlier age of onset (61.0±11.1 EA-C *vs.* 69.0±10.9 WE-C, $p<0.001$), likely increasing patients' survival odds. These data demonstrate the benefit of implementing GC population-based screening to increase detection rate of early stages in high-risk countries. However, it does not explain the poorer 3-years survival probabilities observed for stage II patients from the WE-C (59.0±8.3% stage II) as compared to the EA-C (86.4±4.5% stage II).



C

	Stage I			Stage II			Stage III			Stage IV		
	Western Europe	East Asia	*p* value	Western Europe	East Asia	*p* value	Western Europe	East Asia	*p* value	Western Europe	East Asia	*p* value
Patient status (3-yrs post-diagnosis)			<0·001 *1			0.008 *1			>0.05 *1			>0.05 *1
Alive	57 (80.3%)	251 (95.8%)		25 (62.5%)	51 (86.4%)		9 (26.4%)	14 (41.2%)		1 (4.0%)	2 (18.2%)	
Deceased	14 (19.7%)	11 (4.2%)		15 (37.5%)	8 (13.6%)		25 (73.5%)	20 (58.8%)		24 (96.0%)	9 (81.8%)	
Age of onset (years) Mean ± SD	69.0 ± 10.9	61.0 ± 11.1	<0·001	70.5 ± 11.1	69.4 ± 11.6	>0.05	77.3 ± 9.7	71·2 ± 9·9	0.013	75.1 ± 8.5	71.8 ± 5.1	>0.05

*1 Exact test of Fisher (2-sided)

Figure 1. Comparative survival analysis and age of diagnosis of gastric cancer (GC) patients per stage. **(A-B)** Kaplan-Meier curves displaying the cumulative survival of chemo-naïve GC patients by TNM stage (I, II, III and IV) in WE-C **(A)** and EA-C **(B)**. Pooled log-rank test indicates a significant difference between the survival curves, in both cohorts ($p<0.001$). Dashed lines trace the estimated 3-years survival probability and standard error (SE). **(C)** Patient status 3-years after diagnosis and age of onset. *P* values below 0.05 indicate a significant difference between patients' subsets.

2. *Patients stratification based on combined E-cadherin and CD44v6 expression highlights subgroups with distinct clinical outcomes*

To determine which early stage WE-C patients will present poor outcome, we tested the usefulness of two molecular markers formerly associated with tumor aggressiveness, E-cadherin LoF and *de novo* CD44v6 expression. A retrospective series of chemo-naïve GC cases were categorized according to E-cadherin expression (normal or abnormal). Then, to each category, the status of CD44v6 expression (absent/low, high or very high) was added.

In the WE-C, half of tumors (86/170, 50.6%) presented normal E-cadherin expression. For these, any combination with CD44v6 expression failed to demonstrate differences in OS (figure 2A). Among patients whose tumors presented abnormal E-cadherin expression (84/170, 49.4%), those displaying very high CD44v6 expression (21/84, 25.0%) had significantly worst OS compared to patients presenting absent/low expression (5 *vs.* 48 months, $p<0.001$) (figure 2B and table S4). Patients with abnormal E-cadherin had considerably lower OS than those with normal E-cadherin expression (24 *vs.* 74 months, $p=0.012$) (figure S2A). To clarify if the poor outcome of patients whose tumors harbor concomitant E-cadherin LoF and *de novo* CD44v6 expression was a consequence of diagnosis at advanced stage, we separated patients' subsets according to early or advanced (stages I/II *vs.* III/IV) pathological stages (figure 2C and D). This analysis highlighted a subset of early stage GC patients harboring abnormal E-cadherin and very high CD44v6 expression (9/46, 19.6%) presenting a particularly poor outcome (OS of only 28 months) (figure 2C).

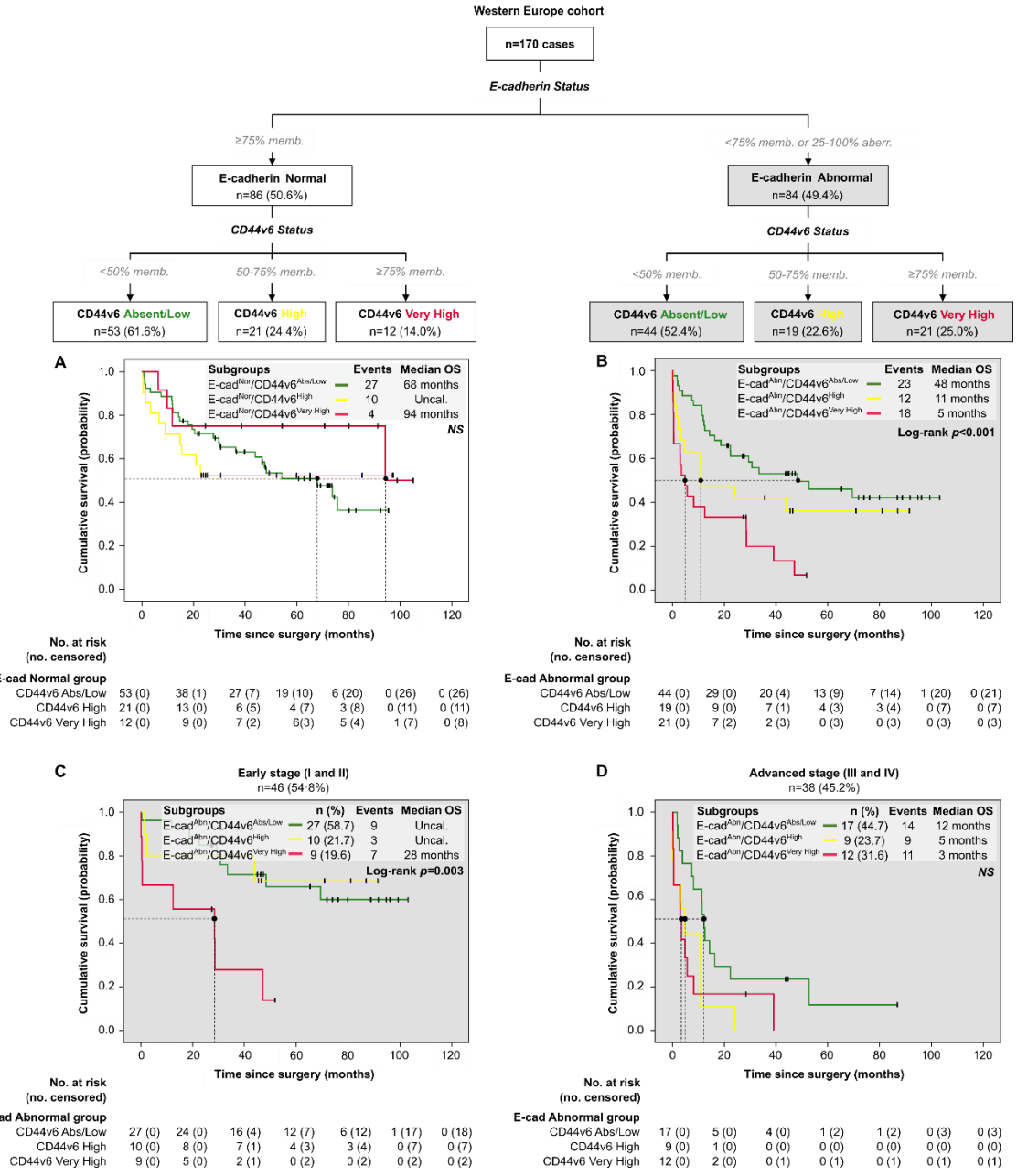


Figure 2. Survival analysis of Western European patients stratified according to E-cadherin and CD44v6 expression. **(A-B)** Kaplan-Meier estimates showing OS of GC patients harboring normal or abnormal E-cadherin expression according to CD44v6 expression status (Absent/Low *vs.* High *vs.* Very High). **(C-D)** Details on the survival rates of GC patients harboring abnormal E-cadherin and CD44v6 *de novo* expression (Absent/Low vs. High vs. Very High) according to early or advanced stages. Dashed lines indicate the median OS. Polled log-rank test indicates a significant difference between the survival curves of GC patients expressing abnormal E-cadherin and very high CD44v6 expression (p<0.01).

To understand the level of co-localization of E-cadherin LoF and *de novo* CD44v6 expression, we used an image analysis pipeline to automatically align immunohistochemical images. We generated a co-existence map by performing a cross-slide analysis of both E-cadherin and CD44v6 immunostainings' that highlighted the coincident expression of both markers (figure 3A). Our analysis further demonstrated that aberrant E-cadherin and membranous CD44v6 expression represent the predominant population within these tumors (figure 3B). Given tissues' microscopic similarity, it is likely that both molecular events occur in the same cancer cells' population, cooperating to potentiate GC aggressiveness.

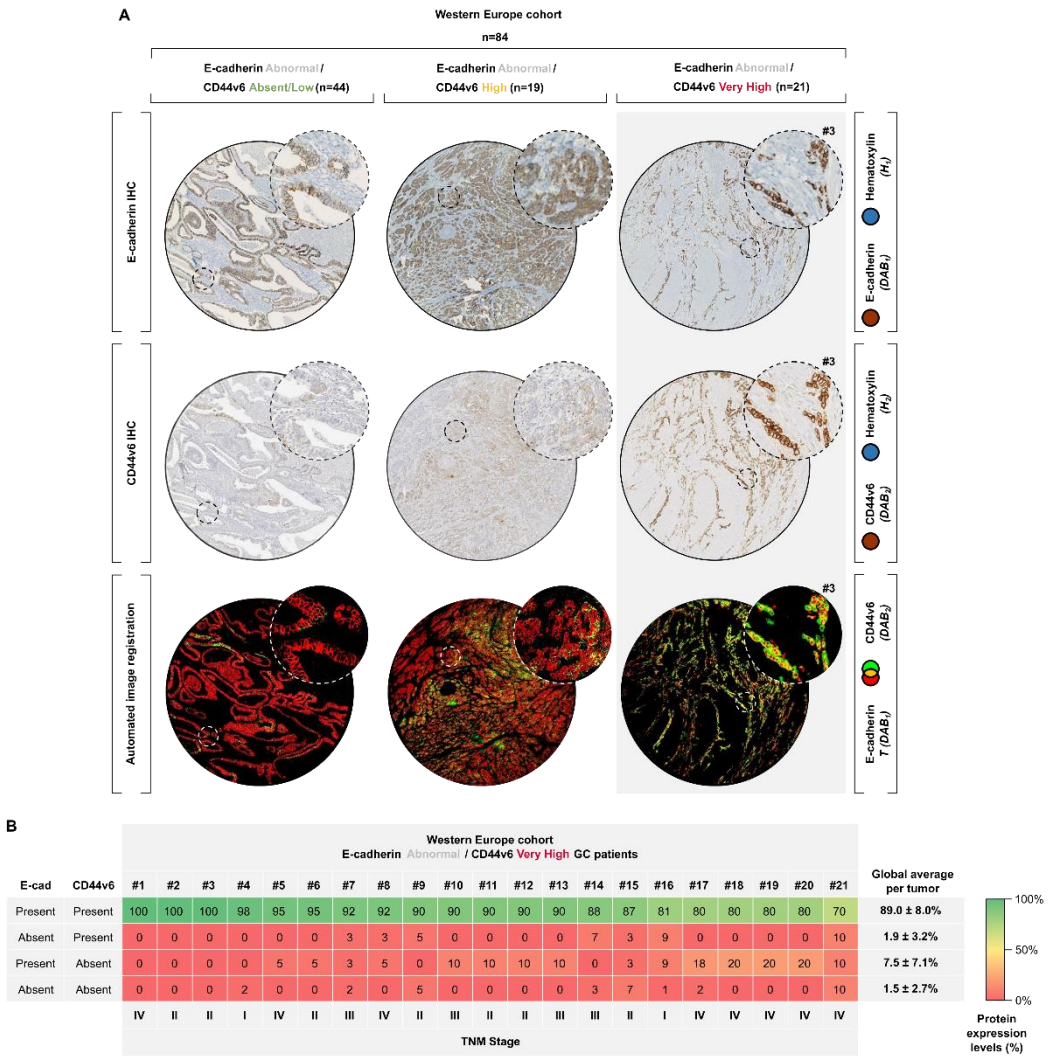


Figure 3. Analysis of E-cadherin and CD44v6 concomitant expression in the WE-C. **(A)** Representative single staining immunohistochemistry images and composite illustrating the automated overlap (yellow) of consecutive TMA slides of CD44v6 (green) and E-cadherin (red). Images are representative of GC patients harboring abnormal E-cadherin expression and CD44v6 *de novo* expression (Absent/Low, High and Very High). **(B)** Estimation of cancer cell populations' comprising GC tumors expressing overlapping abnormal E-cadherin and very high CD44v6 expression. A colored gradient scale was used to express the protein expression levels in the heatmap.

We next analyzed a similar dataset from the EA-C. In this cohort, less than 10% of cases (33/367) presented normal E-cadherin expression. As observed in the WE-C, any combination with CD44v6 expression did not demonstrate differences in OS (figure 4A). More than 90% (334/367) of EA-C patients presented abnormal E-cadherin expression, which was not correlated with poor outcome (figure S2B). From these, 13.2% (44/334) also presented very high CD44v6 expression. We also confirmed that these two molecular events occur in the same cancer populations, as observed in the WE-C (figure S3). However, in contrast to the WE-C, patients from EA-C concomitantly harboring abnormal E-cadherin and very high CD44v6 expression did not present a significantly worse OS (figure 4B and table S5), specially at early stages (figure 4C and D). Indeed, EA-C patients exhibited exceptionally good survival rates compared to the WE-C patients. Indeed, more than 75% of the patients were still alive 5-years after diagnosis, independently of E-cadherin and CD44v6 expression.

In summary, this analysis showed that roughly 12% of patients from both cohorts (21/170 WE-C; 44/367 EA-C) present tumors where most cancer cells harbor concomitant E-cadherin LoF and very high CD44v6 expression. It also showed that these molecular features, known to be related to cancer aggressiveness, are associated with poor outcome in the WE-C, but not in the EA-C.

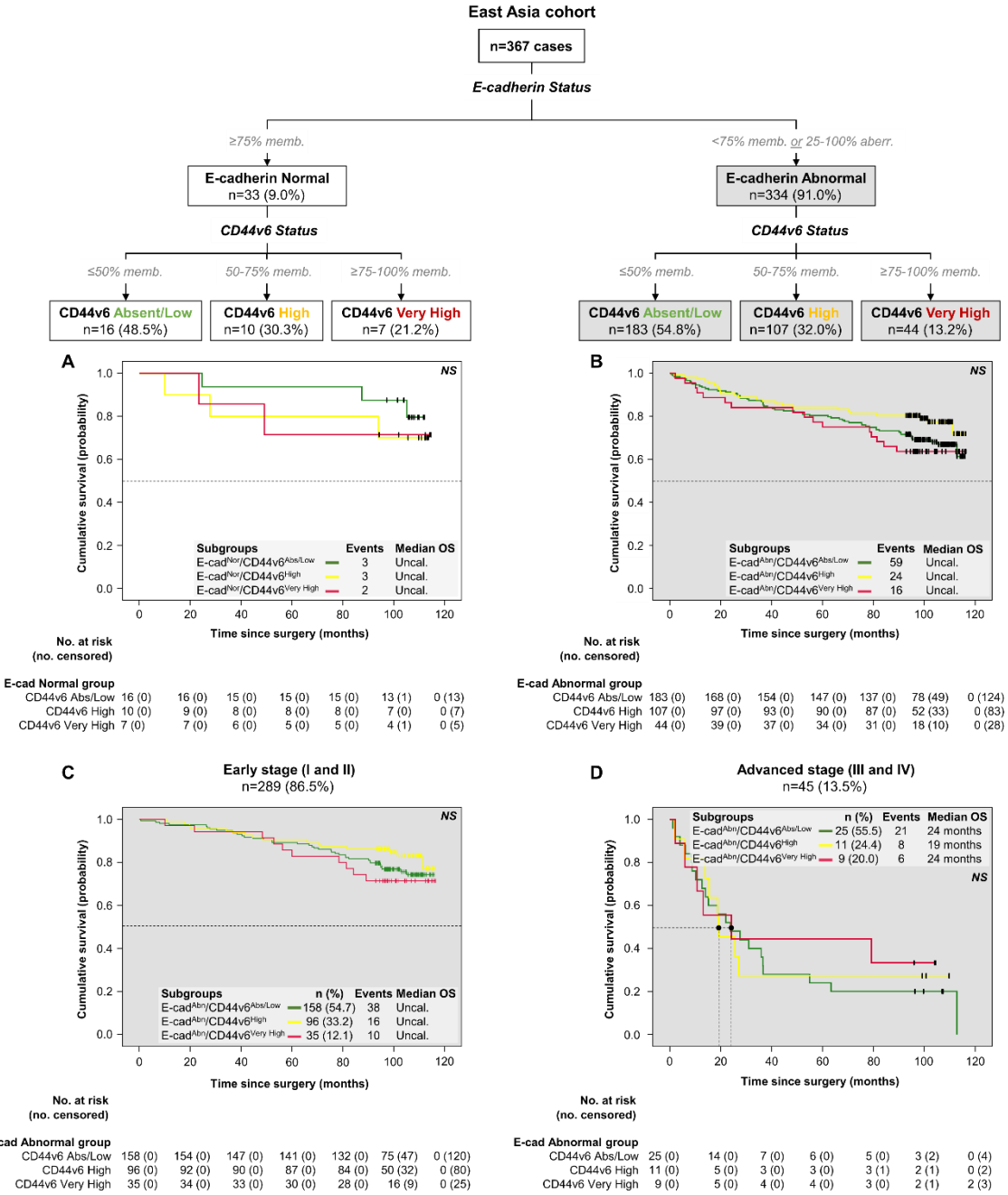


Figure 4. Survival analysis of East Asian patients stratified according to E-cadherin and CD44v6 expression. **(A-B)** Kaplan-Meier estimates showing OS of GC patients harboring normal or abnormal E-cadherin expression according to CD44v6 expression status (Absent/Low *vs.* High *vs.* Very High). **(C-D)** Details on the survival rates of GC patients harboring abnormal E-cadherin and CD44v6 *de novo* expression (Absent/Low vs. High vs. Very High) according to early or advanced stages. Dashed lines indicate the median OS. Pooled log-rank test indicates that there are no significant differences between the survival curves (NS) of any subgroups.

3. *Concomitant abnormal E-cadherin and very high CD44v6 expression determines disease aggressiveness in both cohorts*

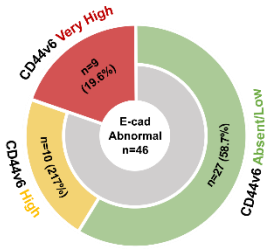
The survival discrepancy in this specific patient subset (GC cases bearing concomitant abnormal E-cadherin and very high CD44v6 expression) was seen as an opportunity to identify the factors underlying the poor survival outcome of early-stage patients from the WE-C.

For that, we explored the clinicopathological features of early stage GC patients in these two world regions. Based on the survival outcomes of WE-C patients (figure 2C), we grouped together cases presenting abnormal E-cadherin and absent/low CD44v6 and high CD44v6 (better OS) and compared them with the cases presenting abnormal E-cadherin and very high CD44v6 (worst OS), for each cohort (figure 5A and B).

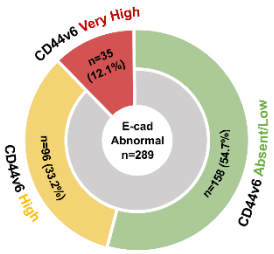
In both cohorts, early stage GC patients with tumors presenting abnormal E-cadherin and very high CD44v6 expression were significantly associated with features of aggressiveness, as compared to all remaining tumors in each series. In the WE-C, these tumors were more often associated to perineural (p=0.001) and vascular invasion (p=0.003) and invaded more frequently deeper layers of the stomach (p=0.020) (figure 5C). In the EA-C, those tumors were more frequently diagnosed at stage II (p=0.005) and permeated more often the lymphatic vessels (p<0.001), promoting LN metastasis (p=0.006) (figure 5D). Importantly, early stages GC cases stratified exclusively according to E-cadherin or CD44v6 expression presented milder features of aggressiveness in both cohorts, than those stratified based on the combination of both molecular features. Early stage GC patients from both cohorts whose tumors present CD44v6 very high expression (Figure S4A and B) were indeed associated with some of the clinicopathological features previously mentioned, but not all, while abnormal E-cadherin expression, by itself, was only associated with diffuse histotype in both cohorts (Figure S5A and B). These results reinforce the importance of combining molecular markers for patient stratification.

In sum, early stage GC patients from both world regions, whose GC tumors are enriched in cell populations concomitantly presenting abnormal E-cadherin and very high CD44v6 expression, were associated with aggressive clinicopathological features linked to invasion and dissemination. Despite that, the outcome of this particular patient subset in the EA-C was favorable.

A Western Europe cohort



B East Asia cohort



C

Variables	Total No. patients (n=46)	E-cad Abn / CD44v6 Absent/Low + High (n=37)	E-cad Abn / CD44v6 Very High (n=9)	*p* value
Age (years)				>0.05
Mean SD	70.5 ± 10.9	69.2 ± 10.8	76.0 ± 9.6	
Gender				>0.05
Male	23 (50·0%)	19 (51·4%)	4 (44·4%)	
Female	23 (50·0%)	18 (48·6%)	5 (55·6%)	
Lauren classification				>0.05
Intestinal	19 (41·3%)	15 (40·5%)	4 (44·4%)	
Diffuse	6 (13·0%)	5 (13·5%)	1 (11·1%)	
Mixed	14 (15·2%)	11 (29·7%)	3 (33·3%)	
Indeterminate	7 (30·4%)	6 (16·2%)	1 (11·1%)	
Growth pattern	44 *1	35 *1		>0.05
Expansive	14 (31.8%)	12 (34.3%)	2 (22.2%)	
Infiltrative	30 (68.2%)	23 (65.7%)	7 (77.8%)	
Lymphatic permeation				>0.05
Absent	22 (47.8%)	20 (54.1%)	2 (22.2%)	
Present	24 (52.2%)	17 (45.9%)	7 (77.8%)	
Perineural invasion				**0.009** *2
Absent	33 (71.7%)	30 (81.1%)	3 (33.3%)	
Present	**13 (28.3%)**	7 (18.9%)	**6 (66.7%)**	
Vascular invasion				**0.018** *2
Absent	28 (60.9%)	26 (70.3%)	2 (22.2%)	
Present	**18 (39.1%)**	11 (29.7%)	**7 (77.8%)**	
Depth of invasion (T)				**0.032**
pT1 (Mucosa + Submucosa)	17 (37.0%)	17 (45.9%)	0 (0.0%)	
pT2 (Muscular)	11 (23.9%)	7 (18.9%)	4 (44.4%)	
pT3-pT4 (Subserosa + Serosa)	**18 (39.1%)**	13 (35.1%)	**5 (55.6%)**	
Lymph node metastases (N)				>0.05
Absent (pN0)	36 (78.3%)	31 (83.8%)	5 (55.6%)	
Present (pN+)	10 (21.7%)	6 (16.2%)	4 (44.4%)	
Distant metastases (M)				-
Absent	46 (100.0%)	-	-	
Present	0 (0.0%)	-	-	
TNM staging				>0.05
I	25 (54.3%)	23 (62.2%)	2 (22.2%)	
II	21 (45.7%)	14 (37.8%)	7 (77.8%)	
Surgical margins				-
R0	46 (100.0%)	-	-	
R1/R2	0 (0.0%)	-	-	

D

Variables	Total No. patients (n=289)	E-cad Abn / CD44v6 Absent/Low + High (n=254)	E-cad Abn / CD44v6 Very High (n=35)	*p* value
Age (years)				**0.003**
Mean SD	62.4 ± 11.8	61.6 ± 11.8	68.0 ± 10.6	
Gender				>0.05
Male	190 (65.7%)	166 (65.4%)	24 (68.6%)	
Female	99 (34.4%)	88 (34.6%)	11 (31.4%)	
Lauren classification	288 *1	253 *1		>0.05
Intestinal	169 (58.7%)	147 (58.1%)	22 (62.9%)	
Diffuse	89 (30.9%)	81 (32.0%)	8 (10.8%)	
Mixed	30 (10.4%)	25 (9.9%)	5 (14.3%)	
Indeterminate	0 (0.0%)	0 (0.0%)	0 (0.0%)	
Growth pattern				>0.05
Expansive	35 (12.1%)	30 (11.8%)	5 (14.3%)	
Infiltrative	254 (87.9%)	224 (88.2%)	30 (85.7%)	
Lymphatic permeation				**0.003** *2
Absent	199 (68.9%)	183 (72.0%)	16 (45.7%)	
Present	**90 (31.1%)**	71 (28.0%)	**19 (54.3%)**	
Perineural invasion				>0.05
Absent	275 (95.2%)	242 (95.3%)	33 (94.3%)	
Present	14 (4.8%)	12 (4.7%)	2 (5.7%)	
Vascular invasion				>0.05
Absent	245 (84.8%)	213 (83.9%)	32 (91.4%)	
Present	44 (15.2%)	41 (16.1%)	3 (8.6%)	
Depth of invasion (T)				>0.05
pT1 (Mucosa + Submucosa)	176 (60.9%)	159 (62.6%)	17 (48.6%)	
pT2 (Muscular)	78 (27.0%)	66 (26.0%)	12 (34.3%)	
pT3-pT4 (Subserosa + Serosa)	35 (12.1%)	29 (11.4%)	6 (35.0%)	
Lymph node metastases (N)				**0.019** *2
Absent (pN0)	247 (85.5%)	222 (87.4%)	25 (71.4%)	
Present (pN+)	**42 (14.5%)**	32 (12.6%)	**10 (28.6%)**	
Distant metastases (M)				-
Absent	289 (100.0%)	-	-	
Present	0 (0.0%)	-	-	
TNM staging				**0.005** *2
I	235 (81.3%)	213 (83.9%)	22 (62.9%)	
II	**54 (18.7%)**	41 (16.1%)	**13 (37.1%)**	
Surgical margins				-
R0	334 (100.0%)	-	-	
R1/R2	0 (0.0%)	-	-	

*1 Remaining data not available; *2 Exact test of Fisher (2-sided);
R0 - clear surgical margins ; R1/R2 - presence of tumor cells at the surgical margins;

Figure 5. Clinicopathological features of early stage GC patients whose tumors harbor abnormal E-cadherin CD44v6 expression. **(A-B)** Radial map depicts the categories of patient subset in the WE-C **(A)** and EA-C **(B)** cohorts. **(C-D)** Clinicopathological associations of E-cadherin abnormal tumors according CD44v6 expression status (Absent/Low plus High *vs.* Very High) in WE-C **(C)** and EA-C **(D)** cohorts. *P* values below 0.05 indicate a significant difference between patients' subsets.

4. *Poor outcome in Western European GC patients with abnormal E-cadherin and very high CD44v6 expressing tumours is likely determined by late diagnosis and limited nodal dissection*

While one may consider that these aggressive clinicopathological features are insufficient to determine the disease outcome, a delayed disease diagnosis and inadequate surgical approach may indeed have a dismal impact in several ways. In line with this, we performed a comparative cohort analysis of the intrinsic features of GC patients with abnormal E-cadherin and very high CD44v6 expression, particularly age of onset, disease stage at diagnosis and the extension of LN dissection.

The comparison of both series revealed that early stage GC patients from EA-C with tumors bearing abnormal E-cadherin and very high CD44v6 were diagnosed, on average 8-years younger

than WE-C patients (68.0 ± 10.6 EA-C vs. 76.0 ± 9.6 WE-C, *p*=0.046, figure 6A). This difference likely arises from fact that South Korea offers population GC screenings programs, which detect a high number of GC cases at stage I, which are often asymptomatic (22/35 (62.9%) EA-C *vs.* 2/9 (22.2%) WE-C, *p*=0.057 figure 6B).

Given that E-cadherin LoF and CD44v6 overexpression determines tumor aggressiveness linked to invasion and dissemination, we hypothesized that the clinical intervention, particularly, the extension of nodal dissection could influence the outcome of this particular GC patients' subset. In general, we observed that GC patients from EA-C were submitted to more radical surgical procedures with extended lymphadenectomy, when compared to WE-C patients (24.8±13.5 WE-C *vs.* 39.0±15.8 EA-C, *p*<0.0001, Figure 6C), removing on average 1.6-times more LNs during the surgery. This tendency was kept in GC patients diagnosed at an early stage and expressing abnormal E-cadherin and very high CD44v6 expression (23.2±8.2 WE-C *vs.* 36.5±14.8 EA-C, *p*=0.005, Figure 6D).

On the basis of the favorable outcomes observed in the EA-C, our results indicate that the potential consequences of nodal metastization, in this cohort, were counteracted by an extensive LN dissection in the EA-C. The opportunity of comparing two cohorts with distinct surgical approaches and survival outcomes, allowed us to identify this particular patient subset that, if not detected at younger ages and at an early stage and treated with extensive LN dissection, may result in poor outcome, as is possible to observe in GC patients from the WE-C.

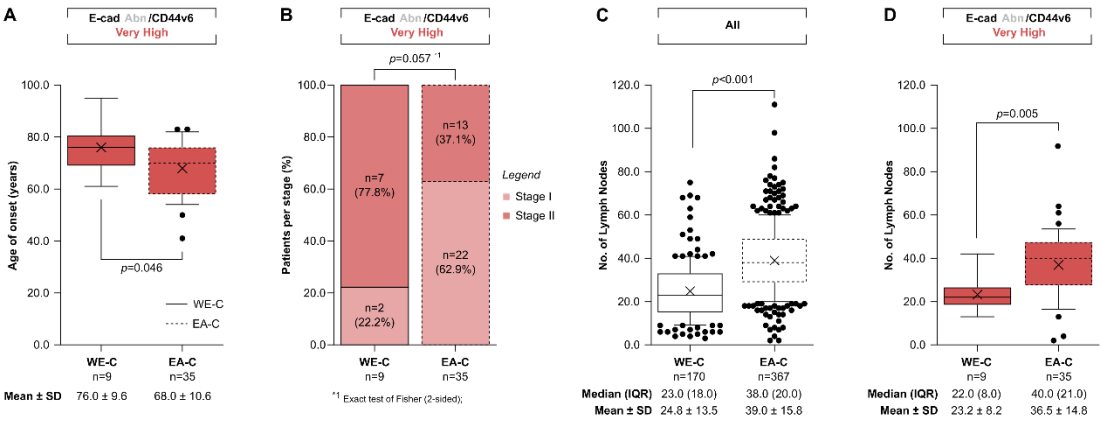


Figure 6. Comparative analysis of diagnostic and clinical intervention features of GC patients with abnormal E-cadherin and very high CD44v6. **(A)** Age of onset; **(B)** Patients per stage; **(C)** Total number of LNs removed in all patients; and **(D)** Total number of LNs removed in GC patients harboring abnormal E-cadherin and very high CD44v6 expression. *P* values below 0.05 indicate a significant difference between cohorts.

DISCUSSION

In this work, we performed a comparative cohort analysis using two independent and contemporary surgical cohorts of chemo-naïve patients from Portugal and South Korea. While South Korea has the highest GC incidence rates worldwide but impressively low mortality rates (37,266 new cases *vs.* 7,684 deaths), Portugal has the highest mortality rates among WE countries (2,885 new cases *vs.* 2,275 deaths) [1]. These differences are likely linked to earlier disease diagnosis through the implementation of GC screening programs in EA countries [24]. As such an approach is missing in WE countries, GC patients are frequently diagnosed at advanced/unresectable stages rendering poorer survival outcomes [5, 25]. We used this asymmetry as an opportunity to explore two molecular markers associated with clinical features of aggressiveness: E-cadherin and CD44v6.

Correlation of combined E-cadherin and CD44v6 expression patterns with patient OS allowed to stratify GC patients into subgroups with different clinical outcomes. One particular patient subset (~12% of all the patients, in each cohort), whose tumors were enriched in populations concomitantly harboring abnormal E-cadherin and very high CD44v6 expression, displayed extremely poor survival rates in the WE-C at early stages, but impressively high survival rates in the EA cohort (EA-C). Although these molecular features did not determine poor outcome for early stage GC patients from the EA-C, early stage tumors from this subset in both cohorts were significantly associated with features of aggressiveness related to invasion and dissemination, as compared to all other tumours in each series.

While one may consider that these molecular features are insufficient to determine poor outcome, if combined with a less aggressive surgical approach (e.g. less extensive LN dissection) neglecting removal of metastatic LN, may indeed have a dismal impact in several ways. Adequate LN dissection facilitates an appropriate staging and is essential to ensure that patients are correctly selected for treatment intervention, but it also eliminates the first site of disease dissemination thus rendering better survival outcomes [2, 26]. That is also what our analysis evidences for the EA-C, because the removal of LN is extensive in all patients and the probability of leaving behind metastatic LN is low. In the WE-C, because less LN are removed from GC patients, those early stage GC patients presenting abnormal E-cadherin and very high CD44v6 expression and more often presenting nodal metastization, have an increased risk of suffering consequences from local invasion.

The lymphatic route is indeed a special road of metastization for GC cells, much facilitated by their easy access along the gastric wall [27]. For this reason, Asian surgeons believed that GC patients' outcome improves if lymphatic spread is surgically prevented [7]. By the time these data were collected, D2 lymphadenectomy (i.e. removal of LNs surrounding the big vessels that supply the stomach) was routinely performed in Asian countries for all GC patients. In contrast, D1 lymphadenectomy (i.e. removal of the perigastric LNs) was the preferred approach in the Western world, due to alleged higher mortality/morbidity and lower survival rates associated to D2 lymphadenectomy [28]. Nowadays, the paradigm of extensive lymphadenectomies has shown signs of change. Western world surgeons have increasingly accepted the importance of performing more than a D1 dissection, and Eastern world surgeons are accepting that more than a D2 dissection does not always justify a certain degree of survival and morbidity.

The evidences collected from both cohorts suggest that the assessment of concomitant E-cadherin and CD44v6 expression in tumor biopsies, from early stage GC patients, may help stratifying those predicted to present more aggressive behaviors and with a higher likelihood to disseminate. Even at early stages, GC patients presenting these molecular profile should be indicated for more extensive lymphadenectomies, whenever patients are medically fit and hospitalized in specialized medical centers. This information is particularly important in WE countries, where routine GC screening strategies are far from being implemented, but also in EA countries where D1-plus approach (i.e. removal of LNs along hepatic and celiac artery) has been currently performed in early stage GC patients [29, 30]. Although we cannot rule out the impact of D1-plus lymphadenectomy in GC patients' outcome whose tumors harbor E-cadherin and CD44v6 dysfunction, our comparative analysis suggests that this patient subset should not be disregarded from extensive surgical approaches.

The mechanism by which the cells enter the vessels and proliferate is not completely understood, but so far, the activation of the epithelial-to-mesenchymal (EMT) transition program has been pointed out as a very efficient strategy to promote local invasion and dissemination [31]. As a cell-to-cell adhesion molecule and the central modulator of the EMT process, E-cadherin loss has been associated to increase LN dissemination propensity [32]. The involvement of CD44v6 in EMT process has also been described [33] and many studies reported increased CD44v6 expression in LN metastasis and linked this fact to worse prognosis.[34] The capacity of tumors to metastasize has also been associated with cancer-initiating cells (CICs) with self-renewal and drug/radiation-resistant properties [35] and there are evidences that CD44v6-expressing cells fulfill some of CICs tasks [15]. In colorectal cancer, CD44v6 overexpression was associated with LN metastasis, however, it is still not clear whether this result is mediated by CICs [36].

In a context of high likelihood of nodal metastization at early GC stages, the timing of diagnosis becomes extremely relevant. In the EA-C, this particular GC patient subset was diagnosed 8-years younger than GC patients from WE-C, when they are, in principle, more fit to perform more extensive surgeries. Despite etiologic and pathologic differences may exist at disease presentation, this is likely the result of the screening strategies implemented in EA countries due to high incidence rates [3]. Although the impact of screening programs in mortality rates is a question of debate, these initiatives detect indeed a large number of asymptomatic early stage cases potentially curable. The approval of a cut-off age for GC screening has been discussed, however the appropriate cut-off age depends on the regional incidence of GC, therefore the adoption of an universal cut-off age for screening is also under debate [37].

To the best of our knowledge, this is the first report describing the clinical usefulness of combined E-cadherin LoF and CD44v6 overexpression in GC. Although further studies with a single-cell resolution and at a molecular level may be essential to clarify E-cadherin and CD44v6 interplay, our co-expression analysis suggests that, it is very likely that E-cadherin LoF and *de novo* CD44v6 occur concomitantly in the same cancer cells' population, which supports patients' stratification based on both molecular features. Notwithstanding, whether this information can change clinical outcomes should be evaluated in a larger number of patients. Either way, international cooperation between

WE and EA countries should be encouraged to establish global standards for the diagnosis and GC patients' management.

Altogether, our study evidences the role of E-cadherin and CD44v6 as powerful biomarkers to identify aggressive early stage tumors with a particularly poorer survival outcome that may benefit from more aggressive surgical treatment.

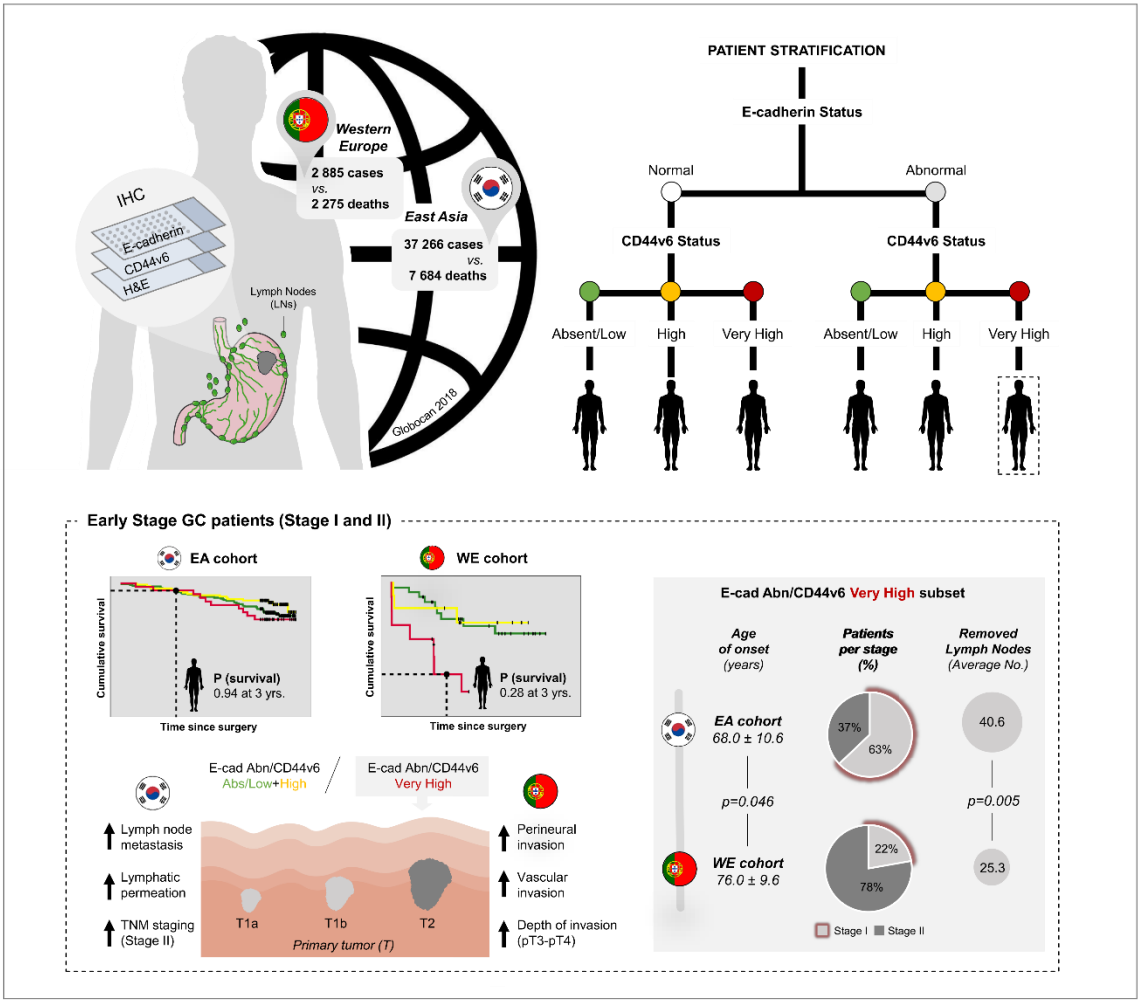


Figure 7. Implications of E-cadherin and CD44v6 dysfunction in tumor aggressiveness and GC patients' outcome. The opportunity of comparing two contemporary cohorts with distinct surgical approaches, allowed us to identify a particular patient subset, whose tumors harbor abnormal E-cadherin and very high CD44v6 expression, that if not detected younger, at an early stage and treated with extensive LN dissection may have a poorer outcome.

Contributors

CP, GMA and CO were involved in the conception and design of this study; RA, IG, FC, J-HP, WHK, H-KY were involved in the collection of clinicopathological data and/or tissue microarrays assembly; DL, GG, J-HP, AA and GMA were involved in immunohistochemical staining of E-cadherin and/or CD44v6 of both tissue microarrays; FC, SC, IG, DM, CP, GMA and CO were involved in the evaluation of E-cadherin and CD44v6 imuno-expression; LS and CW developed the TissUUmaps tool and established the pipelines for automated image registration; CL, CP and GMA were involved in the statistical analysis of patients' cohorts; CP, GMA and CO analyzed and interpreted the data, wrote the manuscript and produced the figures. CO sought institutional approval and funding and is responsible as guarantor for the overall content of the manuscript. All authors critically reviewed the manuscript and agreed to the published version.

Declaration of interests

We declare no competing interests.

Acknowledgments

This work was supported 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 – 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 and NORTE-01-0145-FEDER-000029 - 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, funded by ERDF, POCI and FCT, and project PTDC/CTM-NAN/120958/2010, from FCT. CP was supported by the grant SFRH/BD/113031/2015. GMA was supported by the Investigator FCT Program 2013 (IF/00615/2013), POPH - QREN Type 4.2, European Social Fund and Portuguese Ministry of Science and Technology (MCTES).

Ipatimup integrates the i3S Research Unit, which is partially supported by FCT, the Portuguese Foundation for Science and Technology. The development of TissUUmaps was funded by ERC CoG 682810 to CW.

The authors acknowledge the assistance of Mário Seixas in Slides Digitalization Service at Institute of Molecular Pathology and Immunology of the University of Porto (Ipatimup). CP acknowledges the support of the Doctoral Programme in Biomedicine at the Faculty of Medicine from University of Porto (FMUP).

REFERENCES

1. Bray, F.;Ferlay, J.;Soerjomataram, I.;Siegel, R.L.;Torre, L.A.;Jemal, A. Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA Cancer J Clin.* **2018**, 68, 394-424, DOI: 10.3322/caac.21492
2. Smyth, E.C.;Verheij, M.;Allum, W.;Cunningham, D.;Cervantes, A.;Arnold, D. Gastric cancer: ESMO Clinical Practice Guidelines for diagnosis, treatment and follow-up. *Ann Oncol.* **2016**, 27, v38-v49, DOI: 10.1093/annonc/mdw350
3. Leung, W.K.;Wu, M.-s.;Kakugawa, Y.;Kim, J.J.;Yeoh, K.-g.;Goh, K.L.;Wu, K.-c.;Wu, D.-c.;Sollano, J.;Kachintorn, U., et al. Screening for gastric cancer in Asia: current evidence and practice. *Lancet Oncol.* **2008**, 9, 279-287, DOI: 10.1016/S1470-2045(08)70072-X
4. Sugano, K. Screening of gastric cancer in Asia. *Best Pract Res Clin Gastroenterol.* **2015**, 29, 895-905, DOI: 10.1016/j.bpg.2015.09.013
5. De Angelis, R.;Sant, M.;Coleman, M.P.;Francisci, S.;Baili, P.;Pierannunzio, D.;Trama, A.;Visser, O.;Brenner, H.;Ardanaz, E., et al. Cancer survival in Europe 1999-2007 by country and age: results of EUROCARE--5-a population-based study. *Lancet Oncol.* **2014**, 15, 23-34, DOI: 10.1016/s1470-2045(13)70546-1
6. Bonenkamp, J.J.;Hermans, J.;Sasako, M.;van de Velde, C.J.;Welvaart, K.;Songun, I.;Meyer, S.;Plukker, J.T.;Van Elk, P.;Obertop, H., et al. Extended lymph-node dissection for gastric cancer. *N Engl J Med.* **1999**, 340, 908-914, DOI: 10.1056/nejm199903253401202
7. Zhang, C.-D.;Yamashita, H.;Seto, Y. Gastric cancer surgery: historical background and perspective in Western countries versus Japan. *Ann Transl Med.* **2019**, 7, 70, DOI:
8. Cunningham, D.;Allum, W.H.;Stenning, S.P.;Thompson, J.N.;Van de Velde, C.J.H.;Nicolson, M.;Scarffe, J.H.;Lofts, F.J.;Falk, S.J.;Iveson, T.J., et al. Perioperative Chemotherapy versus Surgery Alone for Resectable Gastroesophageal Cancer. *New England Journal of Medicine.* **2006**, 355, 11-20, DOI: 10.1056/NEJMoa055531
9. Ychou, M.;Boige, V.;Pignon, J.P.;Conroy, T.;Bouché, O.;Lebreton, G.;Ducourtieux, M.;Bedenne, L.;Fabre, J.M.;Saint-Aubert, B., et al. Perioperative chemotherapy compared with surgery alone for resectable gastroesophageal adenocarcinoma: an FNCLCC and FFCD multicenter phase III trial. *J Clin Oncol.* **2011**, 29, 1715-1721, DOI: 10.1200/jco.2010.33.0597
10. Brierley, J.;Gospodarowicz, M.;O'Sullivan, B. The principles of cancer staging. *Ecancermedicalscience.* **2016**, 10, ed61-ed61, DOI: 10.3332/ecancer.2016.ed61
11. Carneiro, P.;Fernandes, M.S.;Figueiredo, J.;Caldeira, J.;Carvalho, J.;Pinheiro, H.;Leite, M.;Melo, S.;Oliveira, P.;Simões-Correia, J., et al. E-cadherin dysfunction in gastric cancer--cellular consequences, clinical applications and open questions. *FEBS Lett.* **2012**, 586, 2981-2989, DOI: 10.1016/j.febslet.2012.07.045
12. 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, DOI: 10.1038/32918

13. Oliveira, C.;Pinheiro, H.;Figueiredo, J.;Seruca, R.;Carneiro, F. Familial gastric cancer: genetic susceptibility, pathology, and implications for management. *Lancet Oncol.* **2015**, 16, e60-70, DOI: 10.1016/s1470-2045(14)71016-2
14. Corso, G.;Carvalho, J.;Marrelli, D.;Vindigni, C.;Carvalho, B.;Seruca, R.;Roviello, F.;Oliveira, C. Somatic mutations and deletions of the E-cadherin gene predict poor survival of patients with gastric cancer. *J Clin Oncol.* **2013**, 31, 868-875, DOI: 10.1200/jco.2012.44.4612
15. Zöller, M. CD44: can a cancer-initiating cell profit from an abundantly expressed molecule? *Nat Rev Cancer.* **2011**, 11, 254-267, DOI: 10.1038/nrc3023
16. Lourenço, B.N.;Springer, N.L.;Ferreira, D.;Oliveira, C.;Granja, P.L.;Fischbach, C. CD44v6 increases gastric cancer malignant phenotype by modulating adipose stromal cell-mediated ECM remodeling. *Integr Biol (Camb).* **2018**, 10, 145-158, DOI: 10.1039/c7ib00179g
17. da Cunha, C.B.;Oliveira, C.;Wen, X.;Gomes, B.;Sousa, S.;Suriano, G.;Grellier, M.;Huntsman, D.G.;Carneiro, F.;Granja, P.L., et al. De novo expression of CD44 variants in sporadic and hereditary gastric cancer. *Lab Invest.* **2010**, 90, 1604-1614, DOI: 10.1038/labinvest.2010.155
18. Xing, X.;Tang, Y.B.;Yuan, G.;Wang, Y.;Wang, J.;Yang, Y.;Chen, M. The prognostic value of E-cadherin in gastric cancer: a meta-analysis. *Int J Cancer.* **2013**, 132, 2589-2596, DOI: 10.1002/ijc.27947
19. Xie, J.W.;Chen, P.C.;Zheng, C.H.;Li, P.;Wang, J.B.;Lin, J.X.;Lu, J.;Chen, Q.Y.;Cao, L.L.;Lin, M., et al. Evaluation of the prognostic value and functional roles of CD44v6 in gastric cancer. *J Cancer Res Clin Oncol.* **2015**, 141, 1809-1817, DOI: 10.1007/s00432-015-1964-8
20. Lopes, N.;Bergsland, C.;Bruun, J.;Bjornslett, M.;Vieira, A.F.;Mesquita, P.;Pinto, R.;Gomes, R.;Cavadas, B.;Bennett, E., et al. A panel of intestinal differentiation markers (CDX2, GPA33, and LI-cadherin) identifies gastric cancer patients with favourable prognosis. *Gastric Cancer.* **2020**, DOI: 10.1007/s10120-020-01064-6
21. Solorzano, L.;Partel, G.;Wählby, C. TissUUmaps: Interactive visualization of large-scale spatial gene expression and tissue morphology data. *Bioinformatics.* **2020**, DOI: 10.1093/bioinformatics/btaa541
22. Öfverstedt, J.;Lindblad, J.;Sladoje, N. Fast and Robust Symmetric Image Registration Based on Distances Combining Intensity and Spatial Information. *IEEE T Image Process.* **2019**, 28, 3584-3597, DOI: 10.1109/TIP.2019.2899947
23. Solorzano, L.;Pereira, C.;Martins, D.;Almeida, R.;Carneiro, F.;Almeida, G.;Oliveira, C.;Wahlby, C. Towards automatic protein co-expression quantification in immunohistochemical TMA slides. *IEEE J Biomed Health Inform.* **2020**, DOI: 10.1109/JBHI.2020.3008821
24. Suh, Y.S.;Yang, H.K. Screening and Early Detection of Gastric Cancer: East Versus West. *Surg Clin North Am.* **2015**, 95, 1053-1066, DOI: 10.1016/j.suc.2015.05.012
25. Verdecchia, A.;Francisci, S.;Brenner, H.;Gatta, G.;Micheli, A.;Mangone, L.;Kunkler, I. Recent cancer survival in Europe: a 2000-02 period analysis of EUROCARE-4 data. *Lancet Oncol.* **2007**, 8, 784-796, DOI: 10.1016/s1470-2045(07)70246-2

26. Coburn, N.G.;Swallow, C.J.;Kiss, A.;Law, C. Significant regional variation in adequacy of lymph node assessment and survival in gastric cancer. *Cancer.* **2006**, 107, 2143-2151, DOI: 10.1002/cncr.22229
27. Akagi, T.;Shiraishi, N.;Kitano, S. Lymph node metastasis of gastric cancer. *Cancers.* **2011**, 3, 2141-2159, DOI: 10.3390/cancers3022141
28. Cuschieri, A.;Weeden, S.;Fielding, J.;Bancewicz, J.;Craven, J.;Joypaul, V.;Sydes, M.;Fayers, P. Patient survival after D1 and D2 resections for gastric cancer: long-term results of the MRC randomized surgical trial. Surgical Co-operative Group. *Br J Cancer.* **1999**, 79, 1522-1530, DOI: 10.1038/sj.bjc.6690243
29. Lorenzon, L.;Giudicissi, R.;Scatizzi, M.;Balducci, G.;Cantafio, S.;Biondi, A.;Persiani, R.;Mercantini, P.;D'Ugo, D. D1-plus vs D2 nodal dissection in gastric cancer: a propensity score matched comparison and review of published literature. *BMC surgery.* **2020**, 20, 126-126, DOI: 10.1186/s12893-020-00714-x
30. Kimura, A.;Ogata, K.;Kogure, N.;Yanoma, T.;Suzuki, M.;Toyomasu, Y.;Ohno, T.;Mochiki, E.;Kuwano, H. Outcome of laparoscopic gastrectomy with D1 plus lymph node dissection in gastric cancer patients postoperatively diagnosed with locally advanced disease or lymph node metastasis. *Surgical endoscopy.* **2016**, 30, 2090-2096, DOI: 10.1007/s00464-015-4462-9
31. Ombrato, L.;Malanchi, I. The EMT universe: space between cancer cell dissemination and metastasis initiation. *Crit Rev Oncog.* **2014**, 19, 349-361, DOI: 10.1615/critrevoncog.2014011802
32. Xu, W.;Hu, X.;Chen, Z.;Zheng, X.;Zhang, C.;Wang, G.;Chen, Y.;Zhou, X.;Tang, X.;Luo, L., et al. Normal fibroblasts induce E-cadherin loss and increase lymph node metastasis in gastric cancer. *PloS one.* **2014**, 9, e97306-e97306, DOI: 10.1371/journal.pone.0097306
33. Saito, S.;Okabe, H.;Watanabe, M.;Ishimoto, T.;Iwatsuki, M.;Baba, Y.;Tanaka, Y.;Kurashige, J.;Miyamoto, Y.;Baba, H. CD44v6 expression is related to mesenchymal phenotype and poor prognosis in patients with colorectal cancer. *Oncol Rep.* **2013**, 29, 1570-1578, DOI: 10.3892/or.2013.2273
34. Lu, L.;Huang, F.;Zhao, Z.;Li, C.;Liu, T.;Li, W.;Fu, W. CD44v6: A metastasis-associated biomarker in patients with gastric cancer?: A comprehensive meta-analysis with heterogeneity analysis. *Medicine.* **2016**, 95, e5603-e5603, DOI: 10.1097/MD.0000000000005603
35. Elshamy, W.M.;Duhé, R.J. Overview: cellular plasticity, cancer stem cells and metastasis. *Cancer Lett.* **2013**, 341, 2-8, DOI: 10.1016/j.canlet.2013.06.020
36. Wang, J.-L.;Su, W.-Y.;Lin, Y.-W.;Xiong, H.;Chen, Y.-X.;Xu, J.;Fang, J.-Y. CD44v6 overexpression related to metastasis and poor prognosis of colorectal cancer: A meta-analysis. *Oncotarget.* **2017**, 8, 12866-12876, DOI: 10.18632/oncotarget.14163
37. Hamashima, C.;Group, S.R.;Guidelines, G.D.G.f.G.C.S. Update version of the Japanese Guidelines for Gastric Cancer Screening. *Jpn J Clin Oncol.* **2018**, 48, 673-683, DOI: 10.1093/jjco/hyy077

APPENDIX

This appendix includes supplementary figures and tables.

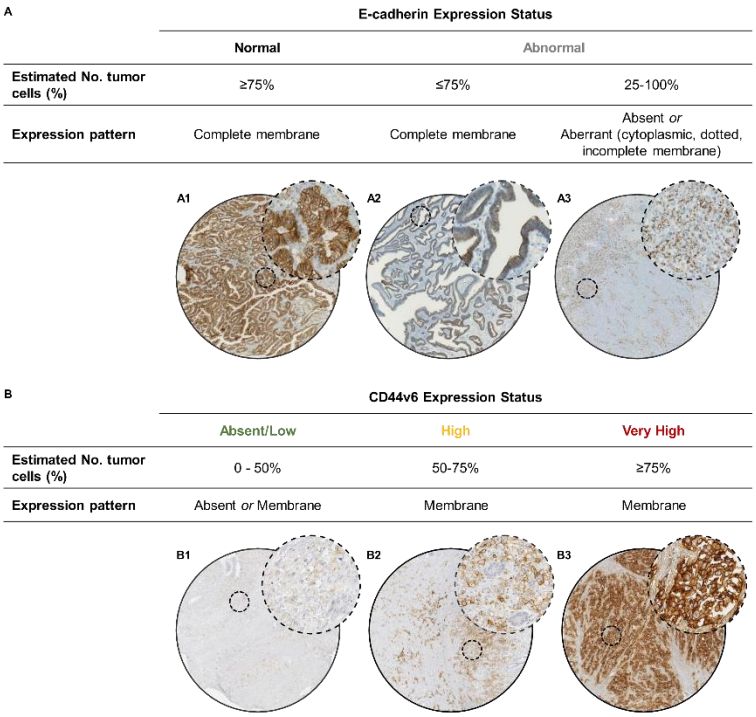


Figure S1. Categorization criteria for the classification of E-cadherin and CD44v6 expression. **(A)** E-cadherin expression was classified as normal when more than 75% of tumor cells presented a complete membranous expression pattern (A1). In case less than 75% of the tumor cells express membrane E-cadherin (A2), cases were classified as abnormal since at least 25% of tumor cells displays absent E-cadherin. Also, cases with completely absent or aberrant staining, in any extent, were classified as abnormal (A3). **(B)** CD44v6 expression was classified according to the extent of CD44v6 expression at the membrane of tumor cells. Cores were classified as: absent/low, if less than 50% of tumor cells presented CD44v6 positivity (B1); high, if 50-75% of tumor cells presented CD44v6 positivity (B2) and very high, if more than 75% of tumor cells presented CD44v6 positivity (B3). One representative core per category was selected to illustrate E-cadherin (A1-A3) and CD44v6 (B1-B3) expression profiles.

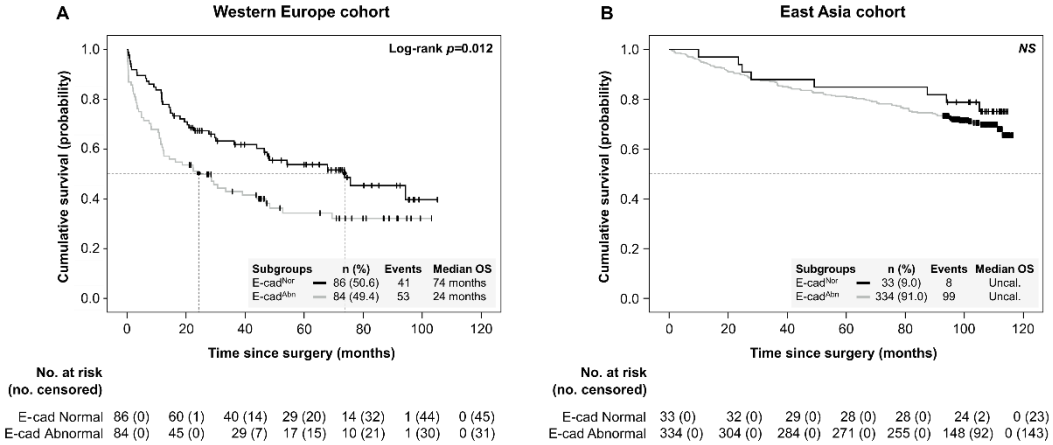


Figure S2. Kaplan-Meier survival analysis of GC patients according E-cadherin expression. **(A)** Western European (WE) GC patients displaying abnormal E-cadherin expression (n=84) presents significantly worst OS compared with patients harboring normal E-cadherin (24 *vs.* 74 months, *p*=0.012). **(B)** East Asian (EA) GC patients expressing abnormal E-cadherin (n=334) stand out for their poor OS, although this tendency is not significantly different from patients expressing normal E-cadherin.

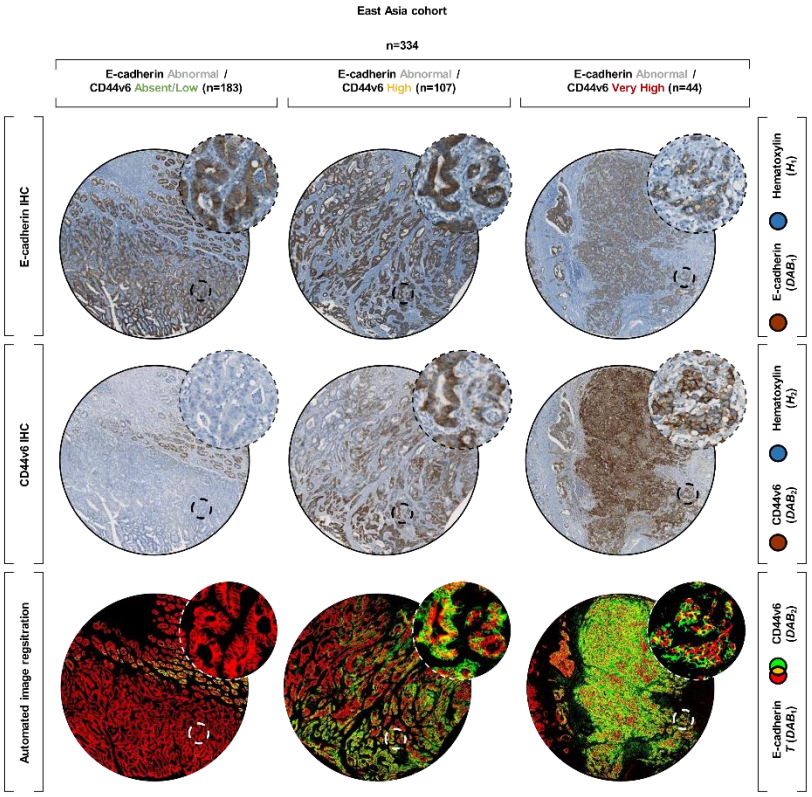


Figure S3. Analysis of E-cadherin and CD44v6 concomitant expression in the EA-C. Representative single staining IHC images and composite illustrating the automated overlap (yellow) of consecutive TMA slides of CD44v6 (green) and E-cadherin (red). Images are representative of GC patients harboring abnormal E-cadherin expression and CD44v6 *de novo* expression (Absent/Low, High and Very High).

A **Western Europe cohort**

Variables	Total No. patients (n=102)	CD44v6 Absent/Low+High (n=85)	CD44v6 Very High (n=17)	*p* value
Age (years)				>0.05
Mean ± SD	69.5 ± 10.9	68.7 ± 11.1	73.4 ± 9.5	
Gender				>0.05
Male	55 (53.9%)	47 (55.3%)	8 (47.1%)	
Female	47 (46.1%)	38 (44.7%)	9 (52.9%)	
Lauren classification				>0.05
Intestinal	59 (57.8%)	50 (58.8%)	9 (52.9%)	
Diffuse	7 (6.9%)	6 (7.1%)	1 (5.9%)	
Mixed	25 (24.5%)	19 (22.4%)	6 (35.3%)	
Indeterminate	11 (10.8%)	10 (11.8%)	1 (5.9%)	
Growth pattern				>0.05
Expansive	34 (33.3%)	29 (34.1%)	5 (29.4%)	
Infiltrative	68 (66.7%)	56 (65.9%)	12 (70.6%)	
Lymphatic permeation				>0.05
Absent	47 (46.1%)	41 (48.2%)	6 (35.3%)	
Present	55 (55.0%)	44 (51.8%)	11 (64.7%)	
Perineural invasion				>0.05
Absent	74 (72.5%)	65 (76.5%)	9 (52.9%)	
Present	28 (27.5%)	20 (23.5%)	8 (47.1%)	
Vascular invasion				**0.001**
Absent	63 (61.8%)	59 (69.4%)	4 (23.5%)	
Present	**39 (38.2%)**	26 (32.5%)	**13 (76.5%)**	
Depth of invasion (T)				**0.032**
pT1 (Mucosa + Submucosa)	44 (43.1%)	41 (48.2%)	3 (17.6%)	
pT2 (Muscular)	22 (21.6%)	15 (17.6%)	7 (41.2%)	
pT3-T4 (Subserosa + Serosa)	36 (35.3%)	29 (34.1%)	7 (41.2%)	
LN metastases (N)				>0.05
Absent (pN0)	80 (78.4%)	68 (80.0%)	12 (70.6%)	
Present (pN+)	22 (21.6%)	17 (20.0%)	5 (3.7%)	
Distant metastases (M)		-	-	-
Absent	102 (100.0%)			
Present	0 (0.0%)			
TNM Staging				>0.05
I	62 (60.8%)	54 (63.5%)	8 (47.1%)	
II	40 (39.2%)	31 (36.5%)	9 (52.9%)	
Surgical margins		-	-	-
R0	102 (100.0%)			
R1/R2	0 (0.0%)			

B **East Asia cohort**

Variables	Total No. patients (n=321)	CD44v6 Absent/Low+High (n=279)	CD44v6 Very High (n=42)	*p* value
Age (years)				>0.05
Mean ± SD	69.5 ± 10.9	61.8 ± 11.6	67.1 ± 10.7	
Gender				>0.05
Male	216 (67.3%)	187 (67.0%)	29 (69.0%)	
Female	105 (32.7%)	92 (33.0%)	13 (31.0%)	
Lauren classification				>0.05
Intestinal	191 (59.5%)	167 (59.9%)	24 (57.1%)	
Diffuse	97 (30.2%)	85 (30.5%)	12 (28.%)	
Mixed	32 (10.0%)	27 (9.7%)	5 (11.9%)	
Indeterminate	1 (0.3%)	0 (0.0%)	1 (2.4%)	
Growth pattern				>0.05
Expansive	39 (12.1%)	34 (12.2%)	5 (11.9%)	
Infiltrative	282 (87.9%)	245 (87.8%)	37 (88.1%)	
Lymphatic permeation				**0.007**
Absent	220 (68.5%)	199 (71.3%)	21 (50.0%)	
Present	**101 (31.5%)**	80 (28.7%)	**21 (50.0%)**	
Perineural invasion				>0.05
Absent	272 (84.7%)	235 (84.2%)	37 (88.1%)	
Present	49 (15.3%)	44 (15.8%)	5 (11.9%)	
Vascular invasion				>0.05
Absent	305 (95.0%)	265 (95.0%)	40 (95.2%)	
Present	16 (5.0%)	14 (5.0%)	2 (4.8%)	
Depth of invasion (T)				>0.05
pT1 (Mucosa + Submucosa)	197 (61.4%)	175 (62.7%)	22 (52.4%)	
pT2 (Muscular)	86 (26.8%)	73 (26.2%)	13 (31.0%)	
pT3-T4 (Subserosa + Serosa)	38 (11.8%)	31 (11.1%)	7 (16.7%)	
LN metastases (N)				>0.05
Absent (pN0)	275 (85.7%)	243 (87.1%)	32 (76.2%)	
Present (pN+)	46 (14.3%)	36 (12.9%)	10 (23.8%)	
Distant metastases (M)		-	-	-
Absent	321 (100.0%)			
Present	0 (0.0%)			
TNM Staging				**0.009**
I	263 (81.9%)	235 (84.2%)	28 (66.7%)	
II	**58 (18.1%)**	44 (15.8%)	**14 (33.3%)**	
Surgical margins		-	-	-
R0	321 (100.0%)			
R1/R2	0 (0.0%)			

*1 Remaining data not available; *2 Exact test of Fisher (2-sided)

Figure S4. Clinicopathological features of early stage GC patients according CD44v6 expression (Absent/Low plus High *vs.* Very High) from WE-C **(A)** and EA-C **(B)**.

A **Western Europe cohort**

Variables	Total No. patients (n=111)	E-cad Normal (n=65)	E-cad Abnormal (n=46)	*p* value
Age (years)				>0.05
Mean ± SD	69.5 ± 10.9	68.8 ± 11.0	70.5 ± 10.8	
Gender				>0.05
Male	60 (54.1%)	37 (56.9%)	23 (50.0%)	
Female	51 (45.9%)	28 (43.1%)	23 (50.0%)	
Lauren classification				**0.039**
Intestinal	**62 (55.9%)**	**43 (66.2%)**	19 (41.3%)	
Diffuse	**8 (7.2%)**	2 (3.1%)	**6 (13.0%)**	
Mixed	28 (25.2%)	14 (21.5%)	14 (30.4%)	
Indeterminate	13 (11.7%)	6 (9.2%)	7 (15.2%)	
Growth pattern	103 *1	59 *1	**44** *1	>0.05
Expansive	34 (33.0%)	20 (33.9%)	14 (31.8%)	
Infiltrative	69 (67.0%)	39 (66.1%)	30 (68.2%)	
Lymphatic permeation	110 *1	64 *1		>0.05
Absent	55 (50.0%)	33 (51.6%)	22 (47.8%)	
Present	55 (50.0%)	31 (48.4%)	24 (52.2%)	
Perineural invasion				>0.05
Absent	82 (73.9%)	49 (75.4%)	33 (71.7%)	
Present	29 (26.1%)	16 (24.6%)	13 (28.3%)	
Vascular invasion	110 *1	64 *1		>0.05
Absent	70 (63.6%)	42 (65.6%)	28 (60.9%)	
Present	40 (36.4%)	22 (34.4%)	18 (39.1%)	
Depth of invasion (T)				>0.05
pT1 (Mucosa + Submucosa)	50 (45.0%)	33 (50.8%)	17 (37.0%)	
pT2 (Muscular)	25 (22.5%)	14 (21.5%)	11 (23.9%)	
pT3-T4 (Subserosa + Serosa)	36 (32.4%)	18 (27.7%)	18 (39.1%)	
LN metastases (N)				>0.05
Absent (pN0)	88 (79.3%)	52 (80.0%)	36 (78.3%)	
Present (pN+)	23 (20.7%)	13 (20.0%)	10 (21.7%)	
Distant metastases (M)		-	-	-
Absent	111 (100.0%)			
Present	0 (0.0%)			
TNM Staging				>0.05
I	71 (64.0%)	46 (70.8%)	25 (54.3%)	
II	40 (36.0%)	19 (29.2%)	21 (45.7%)	
Surgical margins		-	-	-
R0	111 (100.0%)			
R1/R2	0 (0.0%)			

B **East Asia cohort**

Variables	Total No. patients (n=322)	E-cad Normal (n=33)	E-cad Abnormal (n=289)	*p* value
Age (years)				>0.05
Mean ± SD	62.5 ± 11.6	62.4 ± 11.8	63.8 ± 10.1	
Gender				>0.05
Male	216 (67.1%)	26 (78.8%)	190 (65.7%)	
Female	106 (32.9%)	7 (21.2%)	99 (34.3%)	
Lauren classification				**0.018**
Intestinal	**191 (59.5%)**	**22 (66.7%)**	169 (58.7%)	
Diffuse	**97 (30.2%)**	8 (24.2%)	**89 (30.9%)**	
Mixed	32 (10.0%)	2 (6.1%)	30 (10.4%)	
Indeterminate	1 (0.3%)	1 (3.0%)	0 (0.0%)	
Growth pattern				>0.05
Expansive	39 (12.1%)	4 (12.1%)	35 (12.1%)	
Infiltrative	283 (87.9%)	29 (87.9%)	254 (87.9%)	
Lymphatic permeation				>0.05
Absent	**220 (68.3%)**	21 (63.6%)	**199 (68.9%)**	
Present	**102 (31.7%)**	12 (36.4%)	**90 (31.1%)**	
Perineural invasion				>0.05
Absent	273 (84.8%)	28 (84.8%)	245 (84.8%)	
Present	49 (15.2%)	5 (15.2%)	44 (15.2%)	
Vascular invasion				>0.05
Absent	305 (94.7%)	30 (90.9%)	275 (95.2%)	
Present	17 (5.3%)	3 (9.1%)	14 (4.8%)	
Depth of invasion (T)				>0.05
pT1 (Mucosa + Submucosa)	197 (61.2%)	21 (63.6%)	176 (60.9%)	
pT2 (Muscular)	86 (26.7%)	8 (24.2%)	78 (27.0%)	
pT3-T4 (Subserosa + Serosa)	39 (12.1%)	4 (12.1%)	35 (12.1%)	
LN metastases (N)				>0.05
Absent (pN0)	276 (85.7%)	29 (87.9%)	247 (85.5%)	
Present (pN+)	46 (14.3%)	4 (12.1%)	42 (14.5%)	
Distant metastases (M)		-	-	-
Absent	322 (100.0%)			
Present	0 (0.0%)			
TNM Staging				>0.05
I	263 (81.7%)	28 (84.8%)	235 (81.3%)	
II	59 (18.3%)	5 (15.2%)	54 (18.7%)	
Surgical margins		-	-	-
R0	322 (100.0%)			
R1/R2	0 (0.0%)			

*1 Remaining data not available, *2 Exact test of Fisher (2-sided)

Figure S5. Clinicopathological features of early stage GC patients according E-cadherin expression (Normal *vs.* Abnormal) from WE-C **(A)** and EA-C **(B)**.

Table S1. Characterization of clinicopathological features of all GC patients included in WE-C and EA-C.

	Western Europe cohort	East Asia cohort
Variables	**Total No. patients (%) (n= 170)**	**Total No. patients (%) (n= 367)**
Age (years)		
Mean ± SD	71.9 ± 10.8	63.6 ± 11.7
Gender		
Male	86 (50.6%)	245 (66.8%)
Female	84 (49.4%)	122 (33.2%)
Lauren classification		366 *1
Intestinal	87 (51.2%)	211 (57.7%)
Diffuse	17 (10.0%)	114 (31.1%)
Mixed	45 (26.5%)	39 (10.7%)
Indeterminate	21 (12.4%)	2 (0.5%)
Growth pattern	162 *1	
Expansive	40 (24.7%)	43 (11.7%)
Infiltrative	122 (75.3%)	324 (88.3%)
Lymphatic permeation	169 *1	
Absent	56 (33.1%)	224 (61.0%)
Present	112 (66.3%)	143 (39.0%)
Perineural invasion		
Absent	97 (57.1%)	292 (79.6%)
Present	73 (42.9%)	37 (10.1%)
Vascular invasion	168 *1	
Absent	75 (44.6%)	330 (89.9%)
Present	93 (55.4%)	37 (10.1%)
Depth of invasion (T)		
pT1 (Mucosa + Submucosa)	51 (30.0%)	197 (53.7%)
pT2 (Muscular)	25 (14.7%)	86 (23.4%)
pT3-T4 (Subserosa + Serosa)	94 (55.3%)	84 (22.9%)
LN metastases (N)		
Absent (pN0)	91 (53.5%)	276 (75.2%)
Present (pN+)	79 (56.5%)	91 (24.8%)
Distant metastases (M)		
Absent	145 (85.3%)	356 (97.0%)
Present	25 (14.7%)	11 (3.0%)
TNM Staging		
I	71 (41.8%)	263 (71.7%)
II	40 (23.5%)	59 (16.1%)
III	34 (20.0%)	34 (9.3%)
IV	25 (14.7%)	11 (3.0%)
Surgical margins		
R0	157 (92.4%)	367 (100.0%)
R1/R2	13 (7.6%)	0 (0.0%)

*1 Remaining data not available;
LN – lymph node; R0 – clear surgical margins;
R1/21 – presence of tumor cells at the surgical margins;

Table S2. Pairwise comparisons of GC patients in the WE-C, according to pTNM staging. The log-rank test indicates the significant differences between survival curves.

	Stage I	**Stage II**	**Stage III**	**Stage IV**
Stage I	-	0.088	<0.001	<0.001
Stage II	0.088	-	<0.001	<0.001
Stage III	<0.001	<0.001	-	0.001
Stage IV	<0.001	<0.001	0.001	-

Table S3. Pairwise comparisons of GC patients in the EA-C, according to pTNM staging. The log-rank test indicates the significant differences between survival curves.

	Stage I	**Stage II**	**Stage III**	**Stage IV**
Stage I	-	<0.001	<0.001	<0.001
Stage II	<0.001	-	<0.001	<0.001
Stage III	<0.001	<0.001	-	0.030
Stage IV	<0.001	<0.001	0.030	-

Table S4. Pairwise comparisons of GC patients in the WE-C, according to E-cadherin/ CD44v6 subgroups. The log-rank test indicates the significant differences between survival curves.

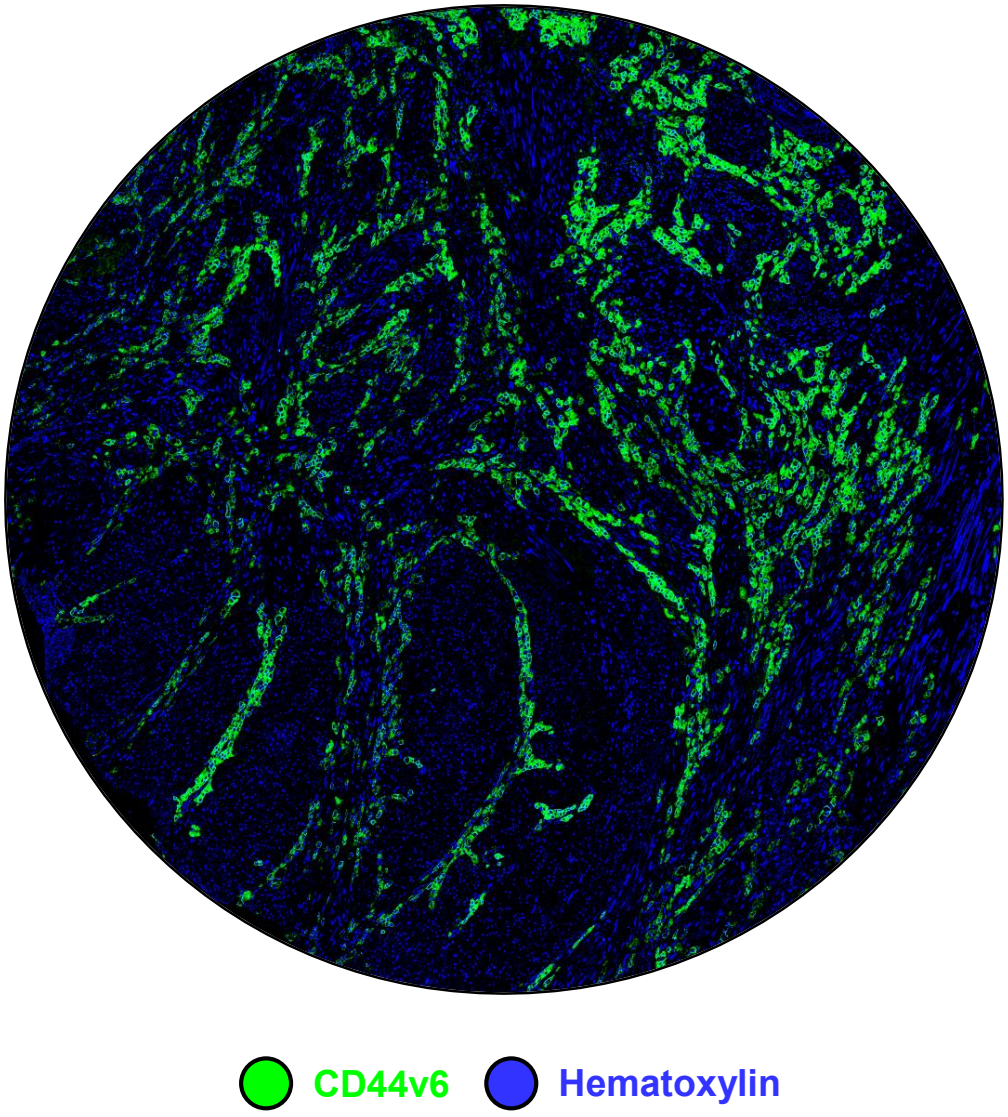
	E-cad ^{Abnormal}/ CD44v6 $^{Absent/Low}$	**E-cad ^{Abnormal}/ CD44v6 ^{High}**	**E-cad ^{Abnormal}/ CD44v6 $^{Very\ High}$**
E-cad ^{Abnormal}/ CD44v6 $^{Absent/Low}$	-	0.182	<0.001
E-cad ^{Abnormal}/ CD44v6 ^{High}	0.182	-	0.096
E-cad ^{Abnormal}/ CD44v6 $^{Very\ High}$	<0.001	0.096	-

Table S5. Pairwise comparisons of GC patients in the EA-C, according to E-cadherin/ CD44v6 subgroups. The log-rank test indicates the significant differences between survival curves.

	E-cad ^{Abnormal}/ CD44v6 $^{Absent/Low}$	**E-cad ^{Abnormal}/ CD44v6 ^{High}**	**E-cad ^{Abnormal}/ CD44v6 $^{Very\ High}$**
E-cad ^{Abnormal}/ CD44v6 $^{Absent/Low}$	-	0.175	0.552
E-cad ^{Abnormal}/ CD44v6 ^{High}	0.175	-	0.144
E-cad ^{Abnormal}/ CD44v6 $^{Very\ High}$	0.552	0.144	-

Chapter 4

Mechanisms of Therapy Response Mediated by CD44v6



Original Paper 3

Expression of CD44v6 containing isoforms influences cisplatin response in gastric cancer cells

Carla Pereira*, Daniel Ferreira*, Nuno Mendes, Pedro Granja, Gabriela M Almeida, Carla Oliveira

**These authors contributed equally to this work.*

Cancers **2020**, 12(4), 858, DOI: 10.3390/cancers12040858

ABSTRACT

CD44v6-containing isoforms are frequently de novo expressed in gastric cancer (GC). Whether CD44v6 has a central role in GC transformation and/or progression, whether it conditions response to therapy or whether it is only a bystander marker is still not known. Therefore, we aimed to clarify the role of CD44v6 in GC. We generated GC isogenic cell lines stably expressing CD44s or CD44v6 and tested them for different cancer hallmarks and response to cisplatin, and we further confirmed our findings in cells that endogenously express CD44v6. No correlation between overexpression of CD44v6 and the tested cancer hallmarks was observed, suggesting CD44v6 is not a driver of GC progression. Upon cisplatin treatment, CD44v6+ cells survive better and have lower apoptosis levels than CD44v6− cells, possibly due to concomitant activation of STAT3 and P38. In co-culture experiments, we discovered that CD44v6+ cells are involved in GC cell overgrowth after cisplatin treatment. In conclusion, we show that CD44v6 expression increases cell survival in response to cisplatin treatment in GC cells and that these cells override CD44v6-negative cells after cisplatin-treatment. This suggests that tumor expression of CD44v6-containing variants may condition the outcome of GC patients treated with chemotherapy.

Stomach cancer; chemotherapy; CD44 variants; apoptosis; cell population dynamics

INTRODUCTION

CD44 is a ubiquitous membrane receptor that was first described for its role as an organ-specific, lymphocyte homing, cell surface molecule [1]. It has since been associated with a large number of key cellular adhesion processes, such as homotypic and heterotypic cellular adhesion [2,3]. Additionally, CD44 mediates cell to extracellular matrix (ECM) adhesion, as shown by its function as the major receptor for hyaluronic acid, a major component of the ECM [4], and also by its ability to interact with fibronectin [5] and collagen [6]. CD44 has also been described as a co-receptor with a pivotal role in intracellular signaling, mediating key aspects of cellular behavior such as motility and cell survival [7–9].

The human CD44 gene (NG_008937) encodes a polymorphic group of transmembrane glycoproteins generated by alternative splicing. The standard CD44 isoform (CD44s) includes only the constitutive exons, while the variant isoforms (CD44v) contain one or more variable exons (in addition to the constitutive ones) [7,9–11].

Regarding gastric tissue, we have previously shown that CD44s is widely expressed in both normal and diseased gastric epithelial cells [12]. In contrast, CD44v6-containing isoforms are *de novo* expressed in stomach premalignant and malignant lesions and in ~70% of all gastric cancers [12], the third leading cause of cancer related mortality worldwide [13]. Aberrant expression of CD44v isoforms has been associated with several cancer-related features like invasion and metastization [10,14,15]. Moreover, CD44 and its CD44v isoforms are expressed as surface markers of cancer stem cells (CSCs), influencing key CSC-associated properties such as tumor initiation, self-renewal, metastasis and chemoresistance [10,16–18]. Indeed, increased expression of CD44 has been observed in trastuzumab resistant gastric [19] and breast [20] cancer cells, as well as in paclitaxel resistant ovarian cancer cells [21]. Likewise, expression of some CD44v has been associated with chemoresistance, like CD44v6 in pancreatic/prostate/colon cancer cells [22–24] and CD44v9 in gastric cancer cells [25]. Moreover, an association between increased CD44 or CD44v6 tumor expression and worse overall survival (OS) in gastric cancer (GC) patients has previously been reported [26–28]. In sum, much is known about the molecule CD44v6 (in general) and several studies suggest that it can have a function in GC. However, it is still not known whether the observed *de novo* expression of CD44v6 has a central role in the transformative process and/or progression of GC, if it conditions response to therapy or if it is only a bystander marker resulting from all or some of those events. Therefore, in the present work, we aimed to clarify the role of CD44v6 using both an isogenic model of GC cells stably expressing CD44s or CD44v6, and GC cells that endogenously express CD44v6.

MATERIALS AND METHODS

1. Cell lines and culture conditions

Human GC cell lines MKN74 and MKN45 were purchased from the JCRB Cell Bank (Japanese Collection of Research Bioresources Cell Bank), and the non-commercial cell line GP202 cell line was established at Ipatimup [**39**]. All reagents were purchased from ThermoFisher Scientific (Waltham, MA, USA), unless otherwise stated. All cell lines were cultured in RPMI 1640 medium with 10% heat inactivated fetal bovine serum (FBS) (Biowest, Nuaillé, France). Cell lines were maintained at 37 °C and 5% CO_2 in a high humidity atmosphere and sub-cultured every 3 to 4 days. Cells were grown in the absence of antibiotics except for cell selection in MKN74 cells, where G418 was used. Cells were never continuously cultured for more than 4 months. Cell identification was confirmed by STR analysis and cells were confirmed to be free of mycoplasma contamination.

2. Flow cytometry

Flow cytometry was used to determine the expression of CD44 and/or CD44v6 in GC cell lines. All reagents were purchased from ThermoFisher Scientific (Waltham, MA, USA), unless stated otherwise. Cells were seeded in 6-well plates to a density of 5×10^5 cells/well and allowed to grow to approximately 90% confluency. Cells were then washed with phosphate buffered saline (PBS), versinized for 5 min and washed in ice-cold PBS-3%BSA (bovine serum albumin). Cells were then blocked in PBS-3%BSA for 30 min and incubated with the appropriate primary antibody: mouse monoclonal antibody against CD44 (clone 156-3C11; 1:100 dilution; 60 min incubation; Cell Signaling Technology (Beverly, MA, USA)) or mouse monoclonal antibody against CD44v6 (MA54; 1:100; 60 min). Cells were subsequently washed with PBS-3%BSA and incubated with a secondary antibody: anti-mouse Alexa Fluor 488 or anti-mouse Alexa Fluor 564 (1:500; 60 min). Fluorescence was analyzed using a FACS Calibur or an Accuri C6 cytometer (both from BD Biosciences, Franklin Lakes, NJ, USA). The mean fluorescence intensity was measured for at least 20,000 gated events per sample and data were analyzed using the software Flow Jo, version 10.

3. RNA extraction and cDNA synthesis

All reagents were purchased from ThermoFisher Scientific (Waltham, MA, USA), unless otherwise stated. RNA was isolated using the mIRVana™ extraction Kit by following the supplier recommended protocol. Extraction efficiency and quality was assessed using a NanoDrop ND-1000. First Strand cDNA synthesis was performed using SuperScript® II reverse transcriptase, random hexamer primers and 1 µg of template RNA, according to the manufacturers' protocol.

4. Characterization of v6 containing transcripts in GP202 cell line

To characterize the full spectrum of v6 containing transcripts present in the GP202 cell line, we used a set of primers that amplifies from exon 5 through to exon 20. Forward primers included one overlapping exon 5 (primer A, **Table S1**) and another overlapping exon v6 (primer B). A reverse orientation set of primers was designed to overlap exon v6 (primer D), exon boundary 17/18 (primer F), as well as two primers that amplify either the short tailed isoform variant, overlapping exon 19 (primer G) or the long tailed isoform variant, overlapping exon 20 (primer H). To identify the full length sequence of the v3–v10 transcript, we used a nested PCR approach to amplify the variable region of said transcript in the GP202 cell line. The first round of amplification included a forward primer overlapping exon 5 (primer A) and a reverse primer overlapping exon 20 (primer H). The second round of amplification was performed with a set of primers spanning the exon 5/exon v3 boundary (primer C, **Table S1**) and the exon 16/exon 17 boundary (primer E). The resulting amplicons were resolved on a 3% agarose gel. All bands were excised and DNA extracted using a band extraction kit (GE Healthcare, Chicago, IL, USA). Sequencing was achieved with a primer walking strategy, with each primer designed to correctly map each inter-exon region between exons 5 and 20. The primers used for this task were primers, A, B, D, F, G and H (**Table S1**). The sequence of this CD44v3–v10 (i.e. v6 containing) variant, endogenously expressed in the GP202 cell line, was then used to generate the isogenic cell model mentioned below.

5. Generation of an isogenic cell line model of tumor CD44v6 status

All reagents were from ThermoFisher Scientific (Waltham, MA, USA), unless otherwise stated. The sequences coding for CD44s (variant CD44-03-ENST00000263398) and CD44v6 (variant CD44-04-ENST00000415148; Ensembl release 68, July 2012) in a pCMV6-XL5 and pCMV6-AC backbone respectively, were purchased from OriGene. CD44s was excised with EcoRI and SmaI and cloned directly into a pIRES-EGFP2 plasmid. CD44v6 was excised in two steps, first with EcoRI and XhoI and sub-cloned into a pIRES-EGFP2 plasmid with a custom made MCS constructed by Bordeira Carriço et al. [40] and subsequently cleaved with SacI and SalI and subcloned into a pIRES-EGFP2 plasmid. Restriction enzymes were all purchased from New England Biolabs (Ipswich, MA, USA). The correct sequence of each transcript was confirmed by Sanger sequencing. The pIRES-EGFP2 empty vector was used as the mock control. The MKN74 cell line, was transfected using Lipofectamine 2000 reagent with either the empty pIRES-EGFP2 vector (MKN74_Mock), the pIRES-EGFP2_CD44v6 (MKN74_CD44v6) or the pIRES-EGFP2_CD44s (MKN74_CD44s). Selective pressure to isolate stably expressing cells, was applied with 1 mg/mL of G418, 48 h after transfection. In order to obtain pure populations of CD44v6 and CD44s cells, magnetic bead sorting with the Magnetic Separation kit CELLection Pan Mouse IgG kit from Dynabeads® was performed, according to the manufacturers' instructions. Briefly, following transfection, cells were sorted according to the expression of

CD44v6 or CD44s. After the first round of separation, cells were allowed to grow to confluence, and two more rounds of magnetic bead sorting performed. The efficiency of cell selection was assessed by flow cytometry, as described above. Correct expression and localization, of the desired CD44 isoform, at the cell membrane was determined by immunofluorescence, as described below.

6. Immunofluorescence

All immunofluorescence reagents were from ThermoFisher Scientific (Waltham, MA, USA), and all antibodies were from Cell Signaling Technology (Beverly, MA, USA) unless stated otherwise.

Cells were grown on glass coverslips for the desired amount of time under normal growth conditions or with specific treatments before being washed twice in PBS and fixed with 4% paraformaldehyde (Merck, Darmstadt, Germany) for 20 min. Post fixation, cells were washed with PBS, incubated 10 min with NH_4Cl, washed twice with PBS and incubated 5 min in PBS-0.2%Triton X-100 for membrane permeabilization. Blocking was performed for 30 min with PBS-5% BSA. Coverslips were subsequently incubated with the desired primary antibodies. Detection of phosphorylated forms required methanol permeabilization for 10 min on ice prior to incubation with the primary antibody. Primary antibodies, diluted in PBS-5% BSA were: (i) mouse monoclonal antibodies–anti-CD44v6 (clone MA541; 1:100 dilution; 60 min incubation; ThermoFisher Scientific), anti-CD44 antibody (156-3C11; 1:100; 60 min), STAT3 antibody (124H6; 1:100); (ii) rabbit monoclonal antibodies–anti-P38 (D13E1; 1:100; ON incubation), anti-phospho-P38 (12F8; 1:100; ON), and phospho-STAT3 (D3A7; 1:100; ON). Cells were then washed twice with PBS and incubated in the dark with secondary antibodies: Alexa Fluor 488 Donkey Anti-Mouse secondary antibody (1:500; 60 min; Life Technologies, (Carlsbad, CA, USA) or Alexa Fluor 594 anti-rabbit (1:500, 60 min; Life Technologies), after which a final PBS wash was performed. In all cases, the cell nuclei were stained using Vectashield mounting media with DAPI (Vector Laboratories, Burlingame, CA, USA). Cells were analyzed by fluorescence microscopy (Imager.Z1, AxioCam fluorescence microscope or Eclipse TE-2000, both from Zeiss, Gottingen, Germany) using AxioVision software (Rockville, MD, USA).

7. Quantitative Real-Time Reverse Transcription PCR (qRT-PCR)

Gene expression levels were assessed by real-time qRT-PCR. RNA Isolation and cDNA synthesis was performed as described above. Real time qRT-PCR was performed in duplicate with 200 ng of template cDNA on an ABI Prism 7000 Sequence Detection System, using probes specific for CD44v6 (exon span V5–V6, Hs.PT.58.45400024) and total CD44 (exon span V2–V3, Hs.PT.58.4880087) (both from iDT, Corallvile, IA, USA). The relative expression level of CD44v6 and total CD44 was determined by the comparative $2^{-\Delta\Delta C_T}$ method using the housekeeping gene

GADPH (4352934E; ThermoFisher Scientific, Waltham, MA, USA), to normalize expression results between samples.

8. Protein extraction and Western Blot

Cells were cultured for the desired amount of time under normal growth conditions or with specific treatments before being washed twice with PBS and lysed with cold catenin lysis buffer [1% (*v/v*) Triton X-100, 1% (*v/v*) IGEPAL in PBS], supplemented with phosphatase (Sigma-Aldrich, Poole, UK) and protease (Roche, Basel, Switzerland) inhibitor cocktails. Total protein content of the samples was quantified by a modified Bradford Assay (Bio-Rad, Hercules, CA, USA). Thirty-five µg of total protein lysate were resolved by electrophoresis in a 7.5% or a 10% SDS polyacrylamide gel. The protein was transferred onto a Hybond nitrocellulose membrane (GE Healthcare, Chicago, IL, USA) and stained with Ponceau S solution (Sigma-Aldrich, Poole, UK) to confirm the efficacy of the transfer. Membrane blocking was performed by incubating membranes for 20 to 30 min with 5% non-fat milk diluted in 0.5% (*v/v*) PBS-Tween-20, or with 5% BSA diluted in 0.5% (*v/v*) TBS-Tween-20 or 5% non-fat milk diluted in 0.1% (*v/v*) TBS-Tween-20 (for the phosphorylated protein forms). Then, incubation of membranes with primary antibodies was performed overnight at 4 °C. Antibodies were all from Cell Signaling Technology (Beverly, MA, USA), unless stated otherwise. Primary antibodies used were: (i) mouse monoclonal antibody against CD44v6 (MA54; 1:500; ThermoFisher Scientific, Waltham, MA, USA), CD44 (156-3C11; 1:1000); (ii) rabbit monoclonal antibodies against AKT (1:1000), EGFR (D38B1; 1:1000), p44/42 MAPK (Erk1/2) (1:1000), P38 MAPK (D13E1; 1:1000), STAT3 (124H6; 1:1000), phospho-AKT (Ser473) (D9E; 1:1000), phospho-EGFR (D7A5; 1:1000), phospho-p44/42 MAPK (Thr202/Tyr204), (1:1000), phospho-p38 MAPK (Thr180/Tyr182) (12F8; 1:1000), phospho-STAT3 (Tyr705) (D3A7; 1:1000). α-tubulin (1:10,000; Sigma-Aldrich, Poole, UK) or α-actin (1:1000; Santa Cruz Biotechnology, Dallas, TX, USA) were used as loading controls. Membranes were further incubated with horseradish peroxidase-conjugated secondary antibodies (GE Healthcare and Santa Cruz) and labelled protein specific signal was detected by ECL (GE Healthcare, Chicago, IL, USA). Bands were quantified with the Quantity One software (Bio-Rad, Hercules, CA, USA). Protein for Western blotting was extracted from at least three independent biological replicates.

9. Cell Growth Assays

The Sulforhodamine B (SRB) assay or Presto Blue (PB; ThermoFisher Scientific, Waltham, MA, USA), a resazurin-based assay, were used to assess cell growth. All reagents were purchased from Sigma-Aldrich (Poole, UK) unless otherwise stated. Cells were seeded in 96-well plates (2.5×10^3 per well) under normal conditions (5% CO2 humidified

atmosphere at 37 °C) and allowed to adhere for 24 h. Cells were then allowed to grow for a further 48 h. At the required times: (i) for the SRB assay, cells were fixed in 10% trichloroacetic acid (TCA) for 1 h on ice, proteins stained with 4% SRB solution for 30 min, wells washed repeatedly with 1% acetic acid to remove the unbound dye, and the protein-bound stain was solubilized with 10 mM Tris solution. The SRB absorbance was measured at 560 nm and background corrected at 655 nm; (ii) for the PB assay, cells were incubated with PB solution (1×) for 45 min at 37 °C and fluorescence assessed (with an excitation wavelength of 560 nm and emission of 590 nm). Both absorbance and fluorescence were measured using a microplate reader (PowerWave HT Microplate Spectrophotometer; BioTek, Bad Friedrichshall, Germany). At least three independent experiments were performed, each measured in triplicate. The doubling time was calculated using the SRB assay as follows: Doubling Time = (0.301 × t)/(Log ODt/OD0), where ODt is the SRB absorbance at 48 h, OD0 is the SRB absorbance at an initial time and t is the time elapsed between the two (in this case, 48 h). Experiments to determine the IC50 for cisplatin were performed using the PB assay for all GC cell lines used (**Figure S3**): MKN74 IC50 ~ 3.6 μM; MKN45 IC50 ~ 6 μM and; GP202 IC50 ~ 17 μM. Cisplatin concentrations used in subsequent experiments were all performed within these ranges, depending on the assay. The selection of 5-FU concentrations used was performed according to Nakamura et al. [41].

10. Invasion and Slow Aggregation Assays

Cell invasion assays were carried out using matrigel invasion chambers (BD Biosciences, Franklin Lakes, NJ, USA), according to manufacturer´s specifications. Briefly, prior to cell seeding, membranes were hydrated on both sides with complete culture medium for 1 h at 37 °C. Cells were seeded (1×10^5 per invasion chamber) and incubated for 24 h at 37 °C. Cells on the top portion of the filter were removed with a wet cotton swab. Invading cells, on the bottom portion of the membrane, were washed with PBS and fixed with ice cold methanol for 15 min. Each filter was detached from the plastic insert and mounted on a glass slide with mounting medium containing DAPI (Vector Laboratories, Burlingame, CA, USA). The total number of invading cells, from four independent biological replicates, was assessed by counting individual nuclei on a Zeiss Imager.Z1, AxioCam fluorescence microscope.

Slow aggregation assays, were performed by coating each well of a 96-well plate with 50 μL of a 6.7 mg/mL solution of bacto-agar in PBS. A total of 2×10^4 cells were plated per well in triplicate for each experimental condition. Aggregation was monitored under an inverted microscope and photographed at 0, 24 and 48 h post seeding with a Nikon Digital Camera. At least three independent biological replicates were performed.

11. Wound Healing Assay

Wound healing assays were performed by seeding 2.1 × 10^4 cells per condition in a culture insert µ-Dish (iBiDi, Martinsried Germany), according to manufacturers' specifications. Briefly, after detachment of the cell culture insert, cell migration was monitored with a Nikon Digital Camera every 6 h. The last time point recorded was at 32 h, as the gap between the two migrating regions was already closed. At least three biological replicates were performed.

12. Tumor growth kinetics

All procedures involving animals were performed in accordance with the European Guidelines for the Care and Use of Laboratory Animals, Directive 2010/63/UE and the National Regulation published in 2013 (Law number: 113/2013, from August 7th) and approved by the national Ethics Committee from the Portuguese DGAV (ref. n° 013042/2017-05-08). The authors involved in these experiments have an accreditation for animal research given by the Portuguese National Authorities (Ministerial Directive 1005/92). NIH(S)II: nu/nu mice, a strain described by Azar et al. 1980 [42], were generated under Ipatimup/i3S supervision. MKN74_Mock and MKN74_CD44v6 cells (5 × 10^6) were subcutaneously injected bilaterally into the back of four 6–8 week old male nude mice. Animals' health was monitored daily. Tumor size was measured using external digital calipers, every 3 to 6 days, to assess tumor growth kinetics. Tumor volume was calculated using the following formula: length × $width^2$ × 0.5. The experiment was terminated and mice humanely euthanized by cervical dislocation preceded by anesthesia with isoflurane when individual tumor volumes reached ~1000 mm^3. Results are expressed as the average +/−SEM of three animals/tumors (each animal harbored one tumor from MKN74_Mock cells and one tumor from MKN74_CD44v6 cells).

13. Assessment of cell survival upon drug treatments

The effects of cisplatin and 5-FU (both from Sigma Aldrich, Poole, UK) on cell survival of the MKN74_Mock, MKN74_CD44v6 and MKN74_CD44s cells were evaluated by the PB assay, mentioned above. Briefly, cells were seeded in 96-well plates (2500 cells per well), allowed to adhere for 24 h and then incubated with cisplatin or 5-FU for 48 h. Cells were then processed for the PB assay, as described above. Cell survival for each drug treatment was calculated, as a percentage, in relation to the respective vehicle-treated control, for each cell line. At least three independent experiments were performed, each measured in triplicate.

14. Assessment of cisplatin-induced apoptosis

MKN74_Mock, MKN74_CD44v6 and MKN74_CD44s cells were seeded in 6-well plates (1.0 × 10^5 cells/ well), left to adhere for 24 h, and incubated with 10 μM of cisplatin for 24 and 48 h. Cisplatin-induced apoptosis was assayed by labelling cells with Annexin V-APC antibody (ImmunoTools, Friesoythe, Germany), according to the manufacturer's instructions. Briefly, cells were trypsinized, pelleted, washed with PBS, resuspended in Annexin V binding buffer (10 mM HEPES, pH 7.4; 140 mM NaCl; 2.5 mM $CaCl_2$) and incubated with Annexin V APC-conjugate for 15 min. Cells were then washed twice in PBS and resuspended in Annexin V binding buffer. Measurement of phosphatidylserine externalization was analyzed using an Accuri C6 flow cytometer and Accuri C6 software (BD Biosciences, Franklin Lakes, NJ, USA), plotting at least 20,000 events per sample. Results represent the average of at least three independent experiments.

15. CD44v6 expression inhibition by siRNA

GP202 and MKN45 cells were transfected with siRNAs for CD44v6 using Lipofectamine® RNAiMax Transfection Reagent (ThermoFisher Scientific, Waltham, MA, USA), according to the manufacturers' instructions. Briefly, cells were seeded in 6-well plates (2 × 10^5 cells/well and 1.5 × 10^5 cells/well, respectively GP202 and MKN45 cell lines). After 24 h, lipid based conjugates were prepared by mixing 20 nM of non-targeting siRNA (negative control DS NC1; iDT, Leuven, Belgium), or human CD44v6 siRNA (Sense strand: 5'-GCGUCAGGUUCCAUAGGAAUCCUTT-3' and Antisense strand: 5'-AAAGGAUUCCUAUGGAACCUGACGCAG-3', custom made from iDT, Leuven, Belgium), to diluted Lipofectamine® RNAiMax Reagent in 1:1 ratio. The conjugates were incubated for 5 min at room temperature and added dropwise to the cells. Upon 48 h of incubation, protein extraction and Western blots were performed to evaluate the efficacy of CD44v6 silencing. In addition, in vitro cisplatin treatments (10 or 20 μM, for MKN45 and GP202 respectively) were carried out and apoptosis assessed (as described above).

16. Assessment of percentage of CD44v6+ cells upon cisplatin incubation

MKN74_Mock and MKN74_CD44v6 were mixed and seeded in 6-well plates with an initial density of 5 × 10^4 cells/well (50:50 proportion), allowed to grow for 24 h and incubated with 10 μM of cisplatin or vehicle (0.9% (*v/v*) NaCl) for 6 h (preliminary experiments indicated 6 h cisplatin incubation to be suitable for subsequent long-term experiments, as it induced a reasonable degree of apoptosis and cells were still able to recover/proliferate a few days after cisplatin incubation). Cells were then washed with PBS, growth medium was replaced with cisplatin-free media and co-cultures were maintained during 15 days. CD44v6 cell enrichment was assessed at several time points by flow cytometry using a CD44v6 conjugated antibody. Briefly, at each time point, cells were washed with PBS, versinized (ThermoFisher Scientific, Waltham, MA, USA) for 15 min, blocked in PBS-2% FBS for 30 min and incubated with CD44v6-APC conjugated antibody (2F10; 10 μL/106 cells; 30 min; R&D Systems, McKinley Place, MN, USA). Cells were

subsequently washed with PBS-2% FBS and analysed by flow cytometry. Fluorescence was analyzed using a FACS Canto II (BD Biosciences, Franklin Lakes, NJ, USA). The mean fluorescence intensity was measured for at least 20,000 gated events per sample and data were analyzed using the software FlowJo, version 10.

17. Statistical Analysis

For statistical significance assessment, a Two-way ANOVA with Tukey's Post Hoc Test was performed. A *p*-value < 0.05 was considered significant and all analyses were performed using GraphPad Prism version 8.2.1 for Windows.

RESULTS

1. CD44v6 overexpression is not a driver of GC development

In an attempt to clarify the role of CD44v6 overexpression in GC, we generated isogenic cell lines by stably expressing in a CD44-null MKN74 GC cell line (**Figure S1**), the following sequences: (i) a CD44v6-containing isoform highly expressed in GC cells (GP202 cell line, **Figure S2**) [12], from now on mentioned as MKN74_CD44v6 cells (**Figure 1A**); (ii) the CD44s isoform (MKN74_CD44s cells, **Figure 1A**), or; (iii) the empty plasmid (MKN74_Mock cells, **Figure 1A**). Characterization of the generated cells confirmed the expression of the intended transcripts at the transcript and protein levels, as well as its localization at the cell membrane (**Figure 1B–D**). Moreover, the CD44v6 expression levels, obtained here, are physiologically relevant since they are similar to those endogenously expressed in the GP202 cell line (data not shown).

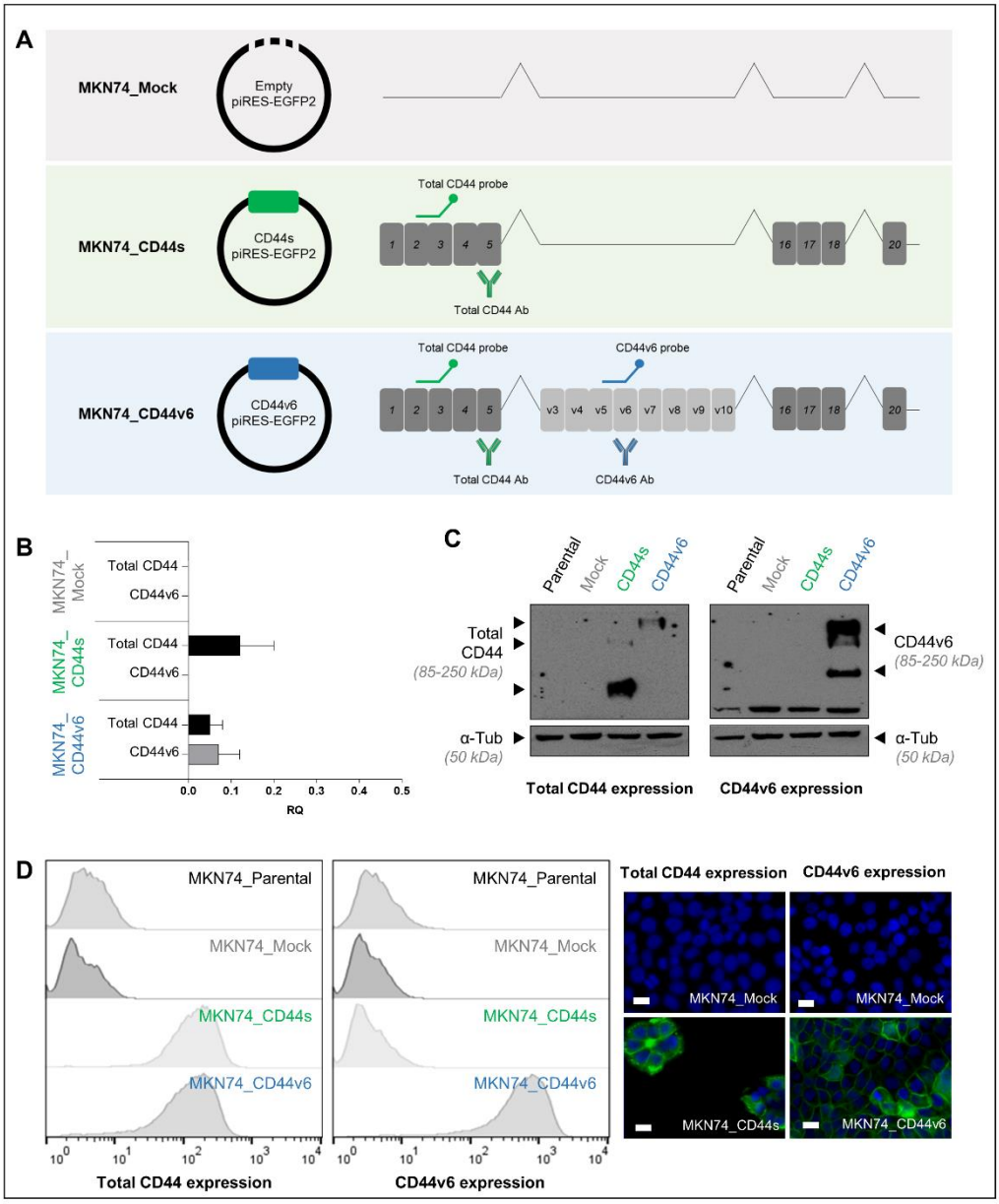


Figure 1. The generated MKN74 cells express the intended CD44 transcripts. (**A**) Transcripts transfected into the MKN74 (CD44v6 negative) gastric cancer (GC) cell line. The CD44 canonical form is the long cytoplasmic tail, exon 20 containing isoform (CD44-03 – ENST00000263398). The exon v6 containing transcript expresses the variable exons between v3-v10 (variant CD44-04 – ENST00000415148). Ensembl release 68, July 2012. It is highlighted where the probes and antibodies, used below, bind to the cDNA or protein, respectively; (**B**) Real time qRT-PCR representing the fold change expression of total CD44 and CD44v6 in transfected cells. Results are shown as average + SD and are representative of three independent experiments. The parental and MKN74_Mock transfected cells were used as negative controls; (**C**) Western Blot depicting total CD44 and CD44v6 protein levels in the generated MKN74 cells; (**D**) Expression was assessed at the post-translational level by immunofluorescence and flow cytometry against total CD44 and CD44v6. Nuclei are stained with DAPI (represented in blue) and white scale bars represent a distance of 20 µm.

These three isogenic cell lines were further characterized and no differences in growth rate, invasion, cell-cell aggregation or migration capacity were observed between them (**Figure 2A–D**). In addition, non-stimulated isogenic cell lines presented no differences regarding the expression of known CD44 interactors and downstream signaling partners (AKT, EGFR, ERK 1/2, P38 and STAT3) at the post-translational level (**Figure 2E, F**). In accordance with these results, tumor growth kinetics was also similar between MKN74_CD44v6 and MKN74_Mock cells (**Figure 2G**). The lack of an obvious correlation between overexpression of CD44v6 and the tested cancer hallmarks, suggests that CD44v6 is not a driver of GC progression.

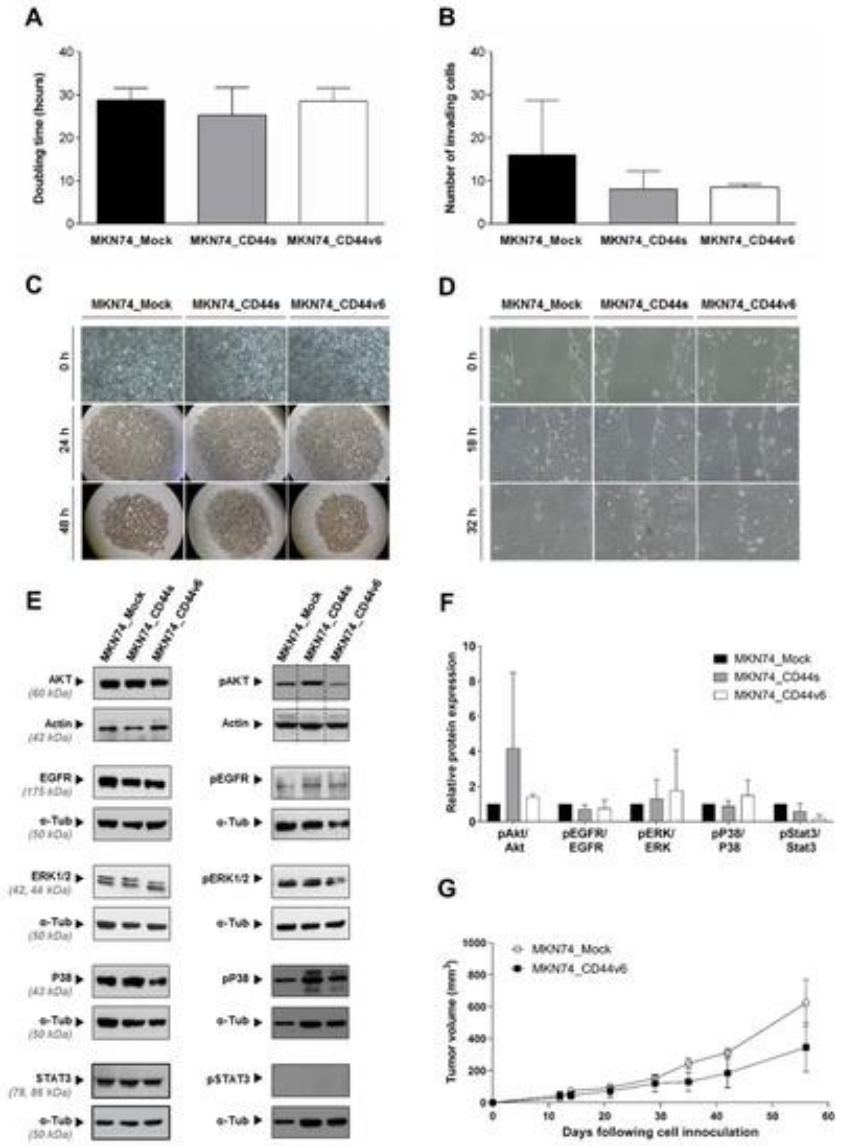


Figure 2. Functional characterization of the generated MKN74_Mock, MKN74_CD44s and MKN74_CD44v6 cells regarding: (**A**) Doubling time; (**B**) Invasion capacity; (**C**) Slow aggregation assay; (**D**) Migration capacity; (**E**) Protein expression of common CD44 interactors: AKT, EGFR, ERK 1/2, P38 and STAT3 and respective phosphorylated forms; (**F**) Relative protein quantification of pAKT, pEGFR, pERK 1/2, pP38 and pSTAT3 with respect to the respective total protein amount. All data are represented as average + SD and/or are representative of three independent experiments; (**G**) Tumor growth kinetics of MKN74_Mock and MKN74_CD44v6 xenografts in nude mice, presented as average +/− SEM. No statistical differences were observed.

2. CD44v6 influences response to cisplatin treatment in GC cells

We then investigated whether CD44v6 influences response to conventional chemotherapeutic agents, namely cisplatin and 5-fluorouracil (5-FU), often included in the chemotherapy regimens provided to GC patients. We verified that MKN74_CD44v6 cells survive better to a clinically relevant cisplatin concentration (10 µM) (**Figure 3A**), compared to MKN74_Mock cells ($p < 0.05$), whereas no differential response was observed in response to 5-FU treatment (**Figure 3B**). Moreover, MKN74_CD44v6 cells also exhibited decreased apoptosis levels in response to cisplatin in comparison to both isogenic counterparts, $p < 0.001$ (**Figure 3C**). These results indicate that de novo expression of CD44v6 in GC cells allows tolerating cisplatin treatment.

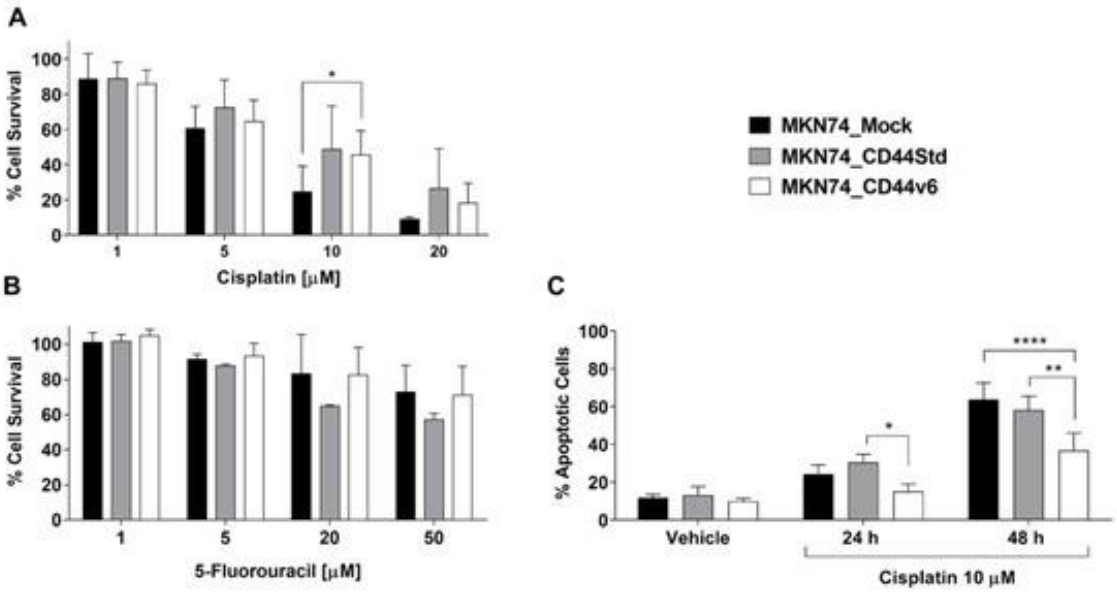


Figure 3. Assessing the response of the isogenic MKN74 cells to conventionally used chemotherapeutic agents: **(A)** Percentage cell survival upon incubation with cisplatin or **(B)** 5-FU for 48 h (compared to vehicle control) in MKN74 cells; **(C)** Percentage of apoptotic cells in MKN74 cells incubated with 10 µM cisplatin or vehicle (0.9% NaCl) for 48 h. Results are expressed as the average + SD of at least three independent experiments. Statistically significant results were determined by Two-way ANOVA with Tukey's multiple comparisons test (* $p < 0.05$; ** $p < 0.001$; **** $p < 0.0001$).

To confirm these results in a model that better mimics the natural context, we modeled the expression of CD44v6 in two non-edited GC cells that endogenously overexpress CD44v6 in all their cells (GP202 and MKN45; **Figure S1**) and treated them with cisplatin. Specific siRNAs were used to inhibit CD44v6 expression, with 50% and 90% inhibition levels being obtained for MKN45 and GP202, respectively (**Figure 4A**). CD44v6 depleted cells displayed higher cisplatin-induced apoptosis levels when compared to siRNA scramble controls, $p < 0.001$ (**Figure 4B**), indicating that CD44v6 expression inhibition renders cancer cells more sensitive to the effects of this drug. Although the siRNAs we used in this experiment can efficiently inhibit the expression of CD44v6, our results could have been reinforced if an additional set of CD44v6 siRNAs had also been used.

Nevertheless, taken together, these results consistently indicate that, in these GC controlled models, CD44v6 modulates response to cisplatin treatment.

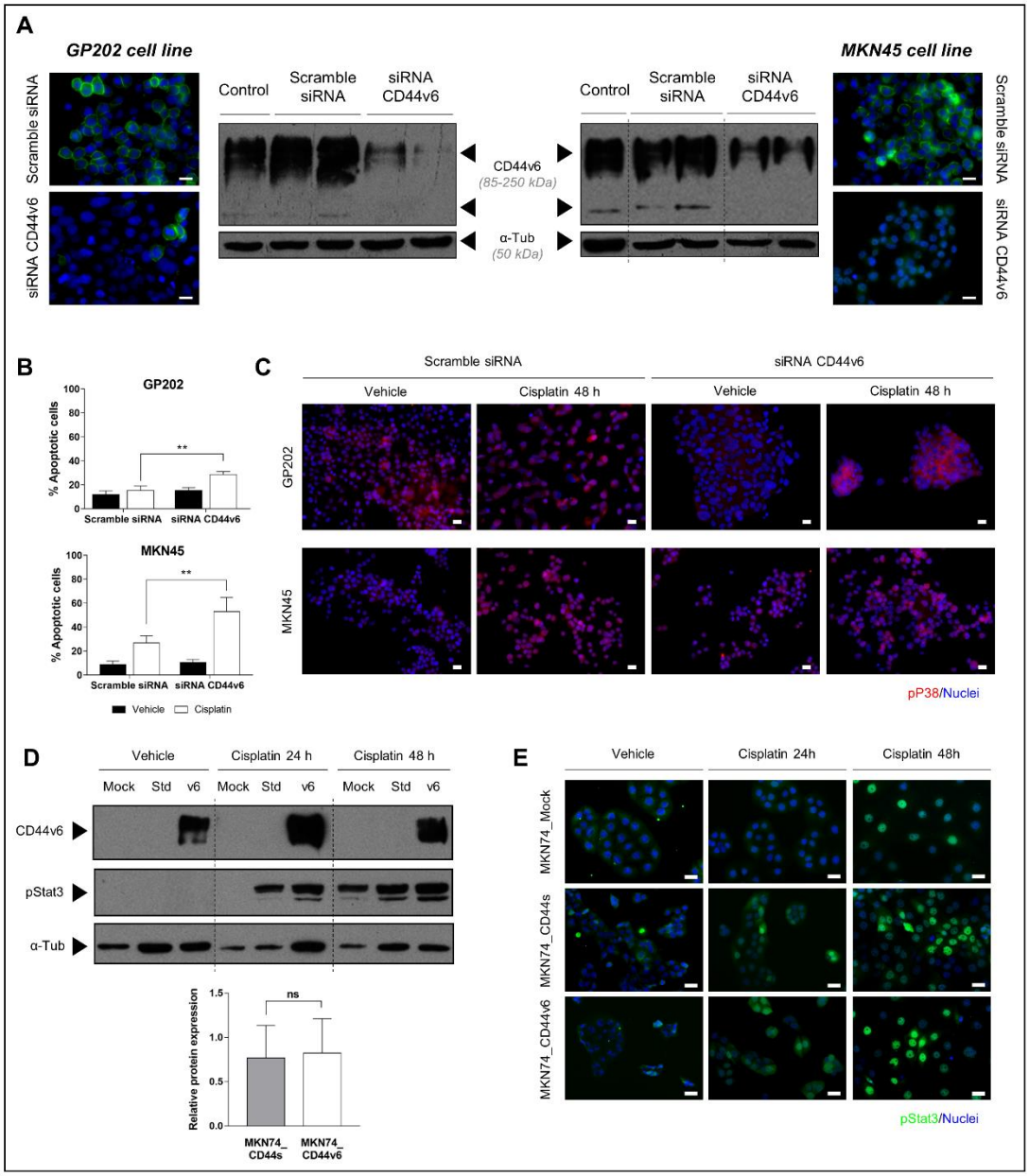


Figure 4. CD44v6-induced modulation to cisplatin response in GC cells is possibly mediated by pSTAT3 or pP38. (**A**) Immunofluorescence and Western blotting of CD44v6 in GP202 (left) and MKN45 (right) upon incubation with CD44v6 specific siRNAs, compared with incubation with scramble siRNA. Expression inhibition levels were, on average, ~90% and ~50% for GP202 and MKN45, respectively, in two independent experiments for each cell line. CD44v6 is represented in green; (**B**) Percentage of apoptotic cells in GP202 and MKN45 cell lines in response to 48 h treatment with cisplatin (concentrations selected according to **Figure S3**) or vehicle, following a 24 h incubation with scramble or CD44v6 siRNAs. Results are expressed as the average + SD of at least three independent experiments. Statistically significant results were determined by Two-way ANOVA with Tukey's multiple comparisons test (* $p < 0.05$; ** $p < 0.001$; **** $p < 0.0001$); (**C**) Immunofluorescence of pP38 (seen in red) in GC cell lines treated with vehicle control

or with cisplatin for 48 h (following a 24 h pre-incubation with scramble or CD44v6 siRNAs). Percentage of cells with nuclear pP38 expression is shown underneath the corresponding experimental conditions; **(D)** Western blotting of pSTAT3 in MKN74 isogenic cell lines in vehicle control and following 24 and 48 h with 10 µM cisplatin treatment. CD44v6 and pSTAT3 were run in the same gel against the same tubulin; **(E)** Immunofluorescence of pSTAT3 (seen in green) in vehicle and cisplatin treated cell lines. In all immunofluorescence images, nuclei are stained with DAPI (seen in blue) and white scale bars represent a distance of 50 µm.

3. CD44v6-induced modulation to cisplatin response in GC cells is likely mediated by pSTAT3 or pP38

Since activation of STAT3 or P38 signaling pathways have been described as mediators of survival in response to cisplatin, we evaluated the levels of pSTAT3 and/or pP38 in the MKN74 isogenic model and in the GC cells that endogenously overexpress CD44v6, GP202 and MKN45 (**Figure 4C–E**). Regarding the isogenic model, none of the three cell lines present the phosphorylated form of STAT3, pSTAT3, at basal level (**Figure 4D,E**; and previously observed in **Figure 2E**). However, following 24 h of cisplatin treatment, pSTAT3 levels increased in MKN74_CD44v6 overexpressing cells, but also in MKN74_CD44s cells, but not in Mock cells (**Figure 4D,E**). Only after 48 h following cisplatin treatment, all three cell lines presented high pSTAT3 levels. Although STAT3 activation, through phosphorylation, becomes a general survival response after 48 h of cisplatin treatment, CD44v6 cells seem to more rapidly induce and maintain STAT3 activation. Regarding P38, the activated form of this protein is already present in the parental MKN74 cell line and no changes were observed in the remaining conditions (**Figure S4 A,B**). We then used GP202 and MKN45 GC cell lines (that endogenously express CD44v6) to test STAT3 and P38 activation following cisplatin treatment in CD44v6 depleted cells. As pSTAT3 is constitutive in both cell lines, we could not use it as a readout of CD44v6 overexpression, nor of cisplatin treatment (**Figure S4C**). The CD44v6 expressing GP202 cells constitutively present pP38 (**Figure 4C**). Following cisplatin treatment, CD44v6 expressing cells show a small, although not statistically significant, increase in the activation of P38 (**Figure 4C**, upper left panel), without an evident difference in terms of cisplatin-induced apoptosis compared to vehicle control for this tested concentration (**Figure 4B**). However, upon CD44v6 depletion, pP38 expression is no longer observed in the nucleus of vehicle nor of cisplatin treated cells (**Figure 4C**, upper right panel), indicating that, in this cell line, nuclear pP38 expression is dependent on the presence of CD44v6. We can then hypothesize that the increased cisplatin sensitivity observed upon depletion of CD44v6 in GP202 cells, in **Figure 4B**, may be due to the loss of CD44v6 mediated pP38. MKN45 parental cells endogenously expressing CD44v6, also present P38 constitutively activated. Similarly to GP202, in the presence of CD44v6, there is only a small increase in P38 activation in response to cisplatin treatment (**Figure 4C**, lower left panel). However, contrary to GP202, in CD44v6 expressing MKN45 cells, there is a significant level of cisplatin-induced apoptosis, compared to vehicle (**Figure 4B**), which is probably related to the cisplatin concentrations used. Regarding P38 activation, while in GP202 cells this depends on the

presence of CD44v6, in MKN45 cells P38 remains activated (and in the nucleus) of most cells upon CD44v6 depletion (**Figure 4C**, lower right panel). Interestingly, in MKN45 cells, CD44v6 also seems to be relevant for P38 activation, at least in the presence of cisplatin, since the percentage of cells with nuclear pP38 expression is significantly reduced when CD44v6 is depleted (**Figure 4C**, lower right panel). Nevertheless, in the presence of both CD44v6 and pP38, MKN45 cells cope better with cisplatin treatment than with only pP38. Overall, these results indicate that in the three GC cell lines, concomitant CD44v6 expression, STAT3 activation and P38 activation is associated with increased survival after cisplatin treatment.

4. CD44v6 expressing cells are involved in GC cell overgrowth after cisplatin treatment

As CD44v6-positive (CD44v6+) cells seem to survive better after cisplatin treatment than CD44v6-negative cells (CD44v6−), it is likely that the proportion of CD44v6+ and CD44v6− cells in heterogeneous tumors (often found in GC, as shown in [29]) influences response to chemotherapy. We set out to test this hypothesis in vitro, by mimicking the effects of chemotherapy on a CD44v6 heterogeneous tumor. For that, we co-cultured CD44v6+ and CD44v6− cells (in a 50:50 ratio), treated them with cisplatin and then allowed them to recover for 15 days, without drug treatment. We then analyzed the proportion of the two cell populations at intermediate time-points and at the end of the experiment by flow cytometry (**Figure 5A**). During recovery from cisplatin treatment, and from day 9 onwards, the CD44v6+ cell population constitutes more than 70–80% of the entire cell culture, $p < 0.0001$ (**Figure 5B**). These results indicate that CD44v6 is involved in GC cell overgrowth after cisplatin treatment. If formally demonstrated in patients, this may suggest that cancer relapses after platinum-based chemotherapy may also be enriched in CD44v6+ cells.

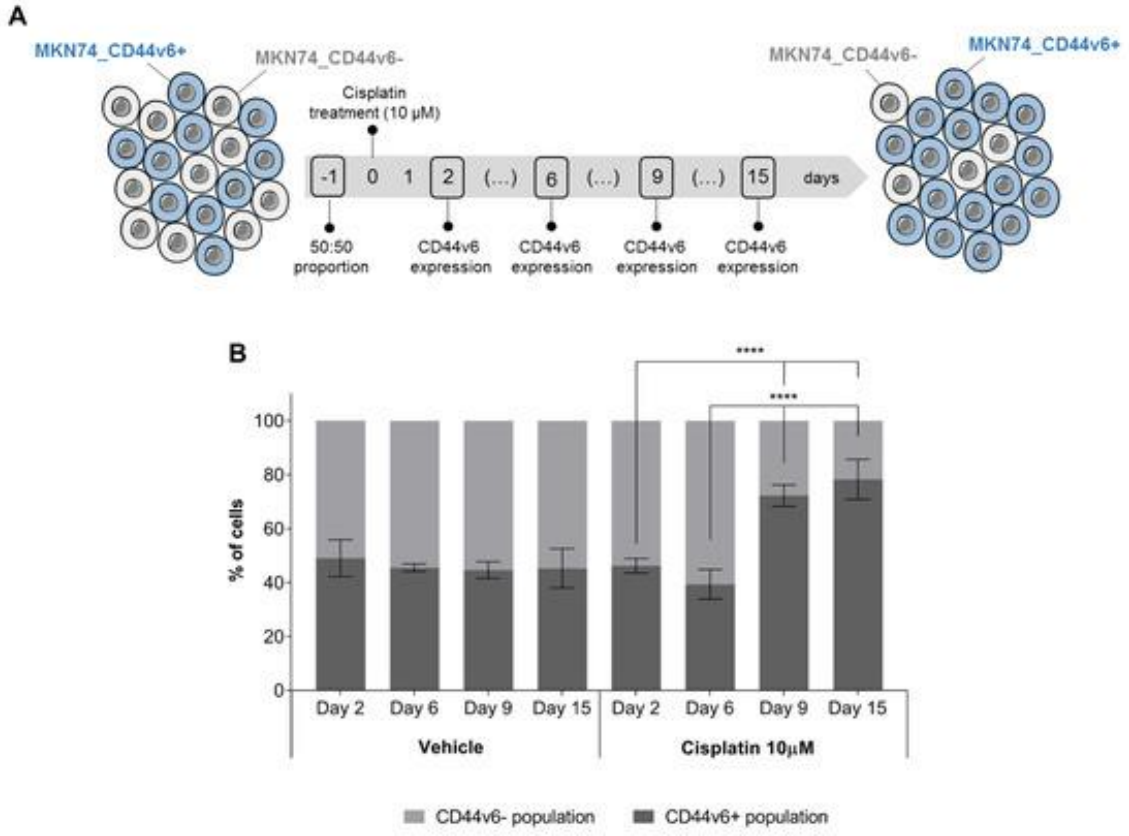


Figure 5. CD44v6+ expressing cells are involved in GC cell overgrowth after cisplatin treatment. (**A**) Scheme depicting the longitudinal assessment of CD44v6+ and CD44v6− population dynamics following cisplatin incubation, in a co-culture setting; (**B**) Percentage of CD44v6+ cells in co-cultures of CD44v6+ and CD44v6− MKN74 cells (MKN74_CD44v6 and MKN74_Mock cells, respectively) following a 6 h incubation with cisplatin (or vehicle control) and up to 15 days of recovery. At the end of the experiment > 80% of the cell culture is composed of CD44v6+ cells. The cisplatin concentration used is clinically relevant. Results expressed as the average +/− SD of at least three independent experiments. Statistically significant results were determined by Two-way ANOVA with Tukey's multiple comparisons test (**** $p < 0.0001$).

DISCUSSION

Aiming to unveil the role of CD44v6 in GC, we generated an isogenic model of GC cells stably expressing CD44s or CD44v6. Expression of CD44v6 or CD44s induced no differences in terms of doubling-time, invasion, cell-adhesion, cell motility, and tumorigenic potential in vivo. These experiments were performed in a tumorigenic GC cell line, which is a limitation of our study regarding possible conclusions on the capacity of CD44v6 de novo expression to influence gastric tumor initiation. Nevertheless, this lack of an obvious correlation between overexpression of CD44v6 and the tested cancer hallmarks, together with our earlier report showing that CD44v6 was already overexpressed in pre-malignant lesions [12] (most of which do not lead to cancer), suggests that CD44v6 is not a driver of GC progression.

Importantly, our data identify CD44v6 as a modulator of response to cisplatin treatment. Indeed, in the isogenic cell line model, overexpression of CD44v6 led to increased cell survival and decreased cisplatin-induced apoptosis. Likewise, downregulation of CD44v6-containing isoforms, in GC cell lines that endogenously express it, led to increased sensitivity to cisplatin-induced apoptosis. Similar results have been described in other cancer types [24,30], suggesting that expression of these variant isoforms could somehow mediate an escape mechanism to programmed cell death. In fact, expression of CD44v has been described to have a role in promoting chemoresistance through the upregulation of lyn kinase, via the Pi3K/AKT pathway in colon carcinoma cells [31] and CD44v6 in particular, was shown to block Fas mediated apoptosis [32]. The mechanism downstream of CD44v6, or cooperating with CD44v6, by which CD44v6 expressing cells have increased cisplatin survival is still unknown, however, our data show that in GC cells this may occur due to the concomitance of CD44v6 expression and activation of STAT3 or P38, depending on the cellular context. Indeed, high expression of activated STAT3 has been shown to contribute to cisplatin resistance in ovarian cancer [33]. Moreover, abrogation of STAT3 activity has been shown to circumvent cisplatin resistance in ovarian cancer cells [34], making it a promising target to reverse cisplatin resistance in ovarian and other cancer types [33,34,35]. It is therefore likely that the early activation of STAT3 we observed in response to cisplatin, in CD44-expressing cells, and mainly in CD44v6-expressing cells, is contributing to increase apoptosis resistance to this drug in the isogenic MKN74 cell line model. The mechanism by which CD44v6 is capable of modulating cisplatin response in GC cells seems to depend on the cellular context. In the two cell lines that endogenously express CD44v6 (MKN45 and GP202), it seems that activation of P38, which our data showed to be CD44v6 dependent in GP202 cells, is contributing to cisplatin resistance. Therefore, when we down-regulate the expression of CD44v6 in GP202 cells, P38 is no longer activated and cells become more sensitive to cisplatin. This is not surprising considering that inhibition of P38 has been shown to sensitize tumor cells to cisplatin-induced apoptosis [36]. In the MKN45 cells, where expression of pP38 is independent of CD44v6, it is not by losing pP38 that the CD44v6 depleted cells become more sensitive to cisplatin. This is probably due to another unexplored pathway. Nevertheless, what does become evident from our experiments is that both GP202 and MKN45 cells respond better to cisplatin treatment when they are expressing both CD44v6 and pP38. Indeed, the same can also be observed in the isogenic MKN74 cells that constitutively express pP38 (regardless of cisplatin treatment) but where the CD44v6 expressing cells cope better with cisplatin treatment. Regarding activation of STAT3, our data are less informative; however, as STAT3 becomes or remains activated upon cisplatin treatment in all tested cell lines, we may hypothesize that it also contributes, in a context of CD44v6 expression and P38 activation, to promote survival of cancer cells when these are challenged by cisplatin. Indeed, a regulatory mechanism between STAT3 and P38 has been described in head and neck cancer, whereby STAT3 phosphorylation requires P38 [37]; however, additional experiments would be required to assess if this is also the case in GC cells. Importantly our study has some limitations since it only provides indications that activation of STAT3 and P38 may be implicated in CD44v6 mediated response to cisplatin. To

confirm these findings, additional studies would have to be performed using, for instance, STAT3 inhibitors to see if in their presence MKN74_CD44v6 cells regain sensitivity to cisplatin, equal to that observed in the MKN74_Mock cells. Interestingly, our experiments using co-cultures of CD44v6+ and CD44v6− cells show that CD44v6+ cells could have some selective advantage when a heterogeneous gastric tumor is treated with cisplatin-based chemotherapy. This selective advantage of CD44v6+ cells was only observed following 9 days of incubation with cisplatin, and appeared to be a sudden increase when compared with the previous time-point tested (6 days). Nevertheless, and considering those results represent the balance between the cells that are dying and the cells that are proliferating in each of the CD44v6+ and CD44v6− cell populations, we believe this was indeed a gradual increase in CD44v6+ cells that took place between day 6 and day 9. However, from the previous results described herein, we can assume that the molecular changes, that ultimately gave rise to this result, started to occur much earlier during/shortly after the cisplatin treatment. This overgrowth of CD44v6+ cells after chemotherapy, may indicate that this cell population can get enriched in GC relapses that arise after chemotherapy. This is plausible as CD44/CD44v are considered CSC markers in several tumor types, including GC [16,17]. Further studies in patients' samples and supported by patients' clinical data are required to address this subject. If our findings are confirmed, strategies to specifically eliminate CD44v6+ cells could lead to decreased recurrence and improved patient survival. One possible strategy is the targeting of CD44v6 downstream effectors, as direct targeting of CD44v6 ought to be avoided due to the lethal skin toxicity side effects described in a Phase I clinical trial using a highly potent antimicrotubule agent coupled to a monoclonal antibody against CD44v6 [38]. Our data also shed light into GC molecular heterogeneity and its relation with therapy, by suggesting that the CD44v6+ population is responsible for tumor overgrowth and may potentially help drive recurrence following conventional chemotherapy, in tumors composed of CD44v6+ and CD44v6− cell populations. Indeed, considering that more than half of GC tumors possess some degree of heterogeneity regarding CD44v6 expression (with CD44v6+ and CD44v6− cell populations co-existing in the same tumor) [29], this may have important consequences in terms of GC patients' clinical outcome.

CONCLUSIONS

In conclusion, we show that expression of CD44 isoforms containing exon v6 in the presence of activated STAT3 and P38, increase cell survival in response to cisplatin treatment in GC cells and that these cells override CD44v6-negative cells after cisplatin treatment. This suggests that tumor expression of CD44v6-containing variants may condition the outcome of GC patients treated with chemotherapy.

Funding

This work was supported 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 – 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 and NORTE-07-0124-FEDER-000029 - 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, funded by ERDF, POCI and FCT, and project PTDC/CTM-NAN/120958/2010, from FCT. CP was supported by the grant SFRH/BD/113031/2015 and DF by the grant PD/BD/105976/2014. GMA was supported by the Investigator FCT Program 2013 (IF/00615/2013), POPH - QREN Type 4.2, European Social Fund and Portuguese Ministry of Science and Technology (MCTES).

Acknowledgments

The authors acknowledge the support of the i3S Scientific Platforms “Advanced Light Microscopy”, member of the PPBI (PPBI-POCI-01-0145-FEDER-022122), and “Translational Cytometry Unit”. CP acknowledges the support of Faculty of Medicine from University of Porto, specifically by the Doctoral Programme in Biomedicine. DF acknowledges the support of the Institute for Biomedical Sciences Abel Salazar from the University of Porto, specifically by the Doctoral Programme on Cellular and Molecular Biotechnology Applied to Health Sciences (reference PD/00016/2012).

Author Contributions

G.M.A. and C.O. conceived, designed and supervised the study; C.P., D.F., N.M., and G.M.A. acquired and analyzed the data; C.P., D.F., G.M.A. and C.O. interpreted the data; C.P., D.F., G.M.A. and C.O. drafted the manuscript; All authors critically revised the manuscript for important intellectual content; C.O. and P.L.G. obtained funding. All authors have read and agreed to the published version of the manuscript.

Conflicts of Interest

The authors declare no conflict of interest.

REFERENCES

1. Gallatin, W.M.; Weissman, I.L.; Butcher, E.C. A cell-surface molecule involved in organ-specific homing of lymphocytes. *Nature* **1983**, *304*, 30–34.
2. Koopman, G.; Vankooyk, Y.; Degraaff, M.; Meyer, C.J.L.M.; Figdor, C.G.; Pals, S.T. Triggering of the CD44 Antigen on Lymphocytes-T Promotes T-Cell Adhesion through the Lfa-1 Pathway. *J. Immunol.* **1990**, *145*, 3589–3593.
3. Shimizu, Y.; Vanseventer, G.A.; Siraganian, R.; Wahl, L.; Shaw, S. Dual Role of the CD44 Molecule in T-Cell Adhesion and Activation. *J. Immunol.* **1989**, *143*, 2457–2463.

4. Aruffo, A.; Stamenkovic, I.; Melnick, M.; Underhill, C.B.; Seed, B. CD44 Is the Principal Cell-Surface Receptor for Hyaluronate. *Cell* **1990**, *61*, 1303–1313.
5. Jalkanen, S.; Jalkanen, M. Lymphocyte CD44 Binds the Cooh-Terminal Heparin-Binding Domain of Fibronectin. *J. Cell Biol.* **1992**, *116*, 817–825.
6. Wayner, E.A.; Carter, W.G.; Piotrowicz, R.S.; Kunicki, T.J. The Function of Multiple Extracellular-Matrix Receptors in Mediating Cell-Adhesion to Extracellular-Matrix - Preparation of Monoclonal-Antibodies to the Fibronectin Receptor That Specifically Inhibit Cell-Adhesion to Fibronectin and React with Platelet Glycoproteins Ic-IIa. *J. Cell Biol.* **1988**, *107*, 1881–1891.
7. Ponta, H.; Sherman, L.; Herrlich, P.A. CD44: From adhesion molecules to signalling regulators. *Nat. Rev. Mol. Cell Biol.* **2003**, *4*, 33–45.
8. Naor, D.; Sionov, R.V.; Ish-Shalom, D. CD44: Structure, function, and association with the malignant process. *Adv. Cancer Res.* **1997**, *71*, 241–319.
9. Ponta, H.; Wainwright, D.; Herrlich, P. The CD44 protein family. *Int. J. Biochem. Cell Biol.* **1998**, *30*, 299–305.
10. Prochazka, L.; Tesarik, R.; Turanek, J. Regulation of alternative splicing of CD44 in cancer. *Cell. Signal.* **2014**, *26*, 2234–2239.
11. Ruiz, P.; Schwarzler, C.; Gunthert, U. CD44 Isoforms during Differentiation and Development. *Bioessays* **1995**, *17*, 17–24.
12. da Cunha, C.B.; Oliveira, C.; Wen, X.; Gomes, B.; Sousa, S.; Suriano, G.; Grellier, M.; Huntsman, D.G.; Carneiro, F.; Granja, P.L.; et al. De novo expression of CD44 variants in sporadic and hereditary gastric cancer. *Lab. Invest.* **2010**, *90*, 1604–1614.
13. Bray, F.; Ferlay, J.; Soerjomataram, I.; Siegel, R.L.; Torre, L.A.; Jemal, A. Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA Cancer J. Clin.* **2018**, *68*, 394–424.
14. Naor, D.; Wallach-Dayan, S.B.; Zahalka, M.A.; Sionov, R.V. Involvement of CD44, a molecule with a thousand faces, in cancer dissemination. *Semin. Cancer Biol.* **2008**, *18*, 260–267.
15. Senbanjo, L.T.; Chellaiah, M.A. CD44: A Multifunctional Cell Surface Adhesion Receptor Is a Regulator of Progression and Metastasis of Cancer Cells. *Front. Cell Dev. Biol.* **2017**, *5*, 18.
16. Yan, Y.; Zuo, X.; Wei, D. Concise Review: Emerging Role of CD44 in Cancer Stem Cells: A Promising Biomarker and Therapeutic Target. *Stem Cells Transl. Med.* **2015**, *4*, 1033–1043.
17. Todaro, M.; Gaggianesi, M.; Catalano, V.; Benfante, A.; Iovino, F.; Biffoni, M.; Apuzzo, T.; Sperduti, I.; Volpe, S.; Cocorullo, G.; et al. CD44v6 is a marker of constitutive and reprogrammed cancer stem cells driving colon cancer metastasis. *Cell Stem Cell* **2014,** *14*, 342–356.
18. Zöller, M. CD44, Hyaluronan, the Hematopoietic Stem Cell, and Leukemia-Initiating Cells. *Front. Immunol.* **2015**, *6*, 235.

19. Yang, Z.; Guo, L.; Liu, D.; Sun, L.; Chen, H.; Deng, Q.; Liu, Y.; Yu, M.; Ma, Y.; Guo, N.; et al. Acquisition of resistance to trastuzumab in gastric cancer cells is associated with activation of IL-6/STAT3/Jagged-1/Notch positive feedback loop. *Oncotarget* **2015**, *6*, 5072–5087.
20. Boulbes, D.R.; Chauhan, G.B.; Jin, Q.; Bartholomeusz, C.; Esteva, F.J. CD44 expression contributes to trastuzumab resistance in HER2-positive breast cancer cells. *Breast Cancer Res. Treat.* **2015**, *151*, 501–513.
21. Gao, Y.; Foster, R.; Yang, X.; Feng, Y.; Shen, J.K.; Mankin, H.J.; Hornicek, F.J.; Amiji, M.M.; Duan, Z. Up-regulation of CD44 in the development of metastasis, recurrence and drug resistance of ovarian cancer. *Oncotarget* **2015**, *6*, 9313–9326.
22. Jung, T.; Gross, W.; Zöller, M. CD44v6 coordinates tumor matrix-triggered motility and apoptosis resistance. *J. Biol. Chem.* **2011**, *286*, 15862–15874.
23. Ni, J.; Cozzi, P.J.; Hao, J.L.; Beretov, J.; Chang, L.; Duan, W.; Shigdar, S.; Delprado, W.J.; Graham, P.H.; Bucci, J.; et al. CD44 variant 6 is associated with prostate cancer metastasis and chemo-/radioresistance. *Prostate* **2014**, *74*, 602–617.
24. Lv, L.; Liu, H.G.; Dong, S.Y.; Yang, F.; Wang, Q.X.; Guo, G.L.; Pan, Y.F.; Zhang, X.H. Upregulation of CD44v6 contributes to acquired chemoresistance via the modulation of autophagy in colon cancer SW480 cells. *Tumour Biol.* **2016**, *37*, 8811–8824.
25. Miyoshi, S.; Tsugawa, H.; Matsuzaki, J.; Hirata, K.; Mori, H.; Saya, H.; Kanai, T.; Suzuki, H. Inhibiting xCT Improves 5-Fluorouracil Resistance of Gastric Cancer Induced by CD44 Variant 9 Expression. *Anticancer Res.* **2018**, *38*, 6163–6170.
26. Chen, Y.; Fu, Z.; Xu, S.; Xu, Y.; Xu, P. The prognostic value of CD44 expression in gastric cancer: A meta-analysis. *Biomed. Pharmacother.* **2014**, *68*, 693–697.
27. Xie, J.W.; Chen, P.C.; Zheng, C.H.; Li, P.; Wang, J.B.; Lin, J.X.; Lu, J.; Chen, Q.Y.; Cao, L.L.; Lin, M.; et al. Evaluation of the prognostic value and functional roles of CD44v6 in gastric cancer. *J. Cancer Res. Clin. Oncol.* **2015**, *141*, 1809–1817.
28. Xu, Y.Y.; Guo, M.; Yang, L.Q.; Zhou, F.; Yu, C.; Wang, A.; Pang, T.H.; Wu, H.Y.; Zou, X.P.; Zhang, W.J.; et al. Regulation of CD44v6 expression in gastric carcinoma by the IL-6/STAT3 signaling pathway and its clinical significance. *Oncotarget* **2017**, *8*, 45848–45861.
29. Pereira, C.; Ferreira, D.; Lemos, C.; Martins, D.; Mendes, N.; Almeida, D.; Granja, P.; Carneiro, F.; Almeida, R.; Almeida, G.M.; et al. CD44v6 expression is a novel predictive marker of therapy response and poor prognosis in gastric cancer patients. *bioRxiv* **2018**, 468934; doi: 10.1101/468934.
30. Mooney, K.L.; Choy, W.; Sidhu, S.; Pelargos, P.; Bui, T.T.; Voth, B.; Barnette, N.; Yang, I. The role of CD44 in glioblastoma multiforme. *J. Clin. Neurosci.* **2016**, *34*, 1–5.
31. Wang, J.L.; Su, W.Y.; Lin, Y.W.; Xiong, H.; Chen, Y.X.; Xu, J.; Fang, J.Y. CD44v6 overexpression related to metastasis and poor prognosis of colorectal cancer: A meta-analysis. *Oncotarget* **2017**, *8*, 12866–12876.
32. Bendardaf, R.; Lamlum, H.; Ristamäki, R.; Pyrhönen, S. CD44 variant 6 expression predicts response to treatment in advanced colorectal cancer. *Oncol. Rep.* **2004**, *11*, 41–45.

33. Wu, C.J.; Sundararajan, V.; Sheu, B.C.; Huang, R.Y.; Wei, L.H. Activation of STAT3 and STAT5 Signaling in Epithelial Ovarian Cancer Progression: Mechanism and Therapeutic Opportunity. *Cancers (Basel)* **2020**, *12*, 24.
34. Ji, T.; Gong, D.; Han, Z.; Wei, X.; Yan, Y.; Ye, F.; Ding, W.; Wang, J.; Xia, X.; Li, F.; et al. Abrogation of constitutive Stat3 activity circumvents cisplatin resistant ovarian cancer. *Cancer Lett.* **2013**, *341*, 231–239.
35. Sun, C.Y.; Nie, J.; Huang, J.P.; Zheng, G.J.; Feng, B. Targeting STAT3 inhibition to reverse cisplatin resistance. Biomed. *Pharmacother* **2019**, *117*, 109135.
36. Pereira, L.; Igea, A.; Canovas, B.; Dolado, I.; Nebreda, A.R. Inhibition of p38 MAPK sensitizes tumour cells to cisplatin-induced apoptosis mediated by reactive oxygen species and JNK. *EMBO Mol. Med.* **2013**, *5*, 1759–1774.
37. Riebe, C.; Pries, R.; Schroeder, K.N.; Wollenberg, B. Phosphorylation of STAT3 in head and neck cancer requires p38 MAPKinase, whereas phosphorylation of STAT1 occurs via a different signaling pathway. *Anticancer Res.* **2011**, *31*, 3819–3825.
38. Riechelmann, H.; Sauter, A.; Golze, W.; Hanft, G.; Schroen, C.; Hoermann, K.; Erhardt, T.; Gronau, S. Phase I trial with the CD44v6-targeting immunoconjugate bivatuzumab mertansine in head and neck squamous cell carcinoma. *Oral Oncol.* **2008**, *44*, 823–829.
39. Gärtner, F.; David, L.; Seruca, R.; Machado, J.C.; Sobrinho-Simões, M. Establishment and characterization of two cell lines derived from human diffuse gastric carcinomas xenografted in nude mice. *Virchows Arch.* **1996**, *428*, 91–98.
40. Bordeira-Carriço, R.; Ferreira, D.; Mateus, D.D.; Pinheiro, H.; Pêgo, A.P.; Santos, M.A.; Oliveira, C. Rescue of wild-type E-cadherin expression from nonsense-mutated cancer cells by a suppressor-tRNA. *Eur. J. Hum. Genet.* **2014**, *22*, 1085–1092.
41. Nakamura, A.; Nakajima, G.; Okuyama, R.; Kuramochi, H.; Kondoh, Y.; Kanemura, T.; Takechi, T.; Yamamoto, M.; Hayashi, K. Enhancement of 5-fluorouracil-induced cytotoxicity by leucovorin in 5-fluorouracil-resistant gastric cancer cells with upregulated expression of thymidylate synthase. *Gastric Cancer* **2014**, *17*, 188–195.
42. Azar, H.A.; Hansen, C.T.; Costa, J. N:NIH(S)-nu/nu mice with combined immunodeficiency: A new model for human tumour heterotransplantation. *J. Natl. Cancer Inst.* **1980**, *65*, 421–430.

APPENDIX

This appendix includes supplementary figures and tables.

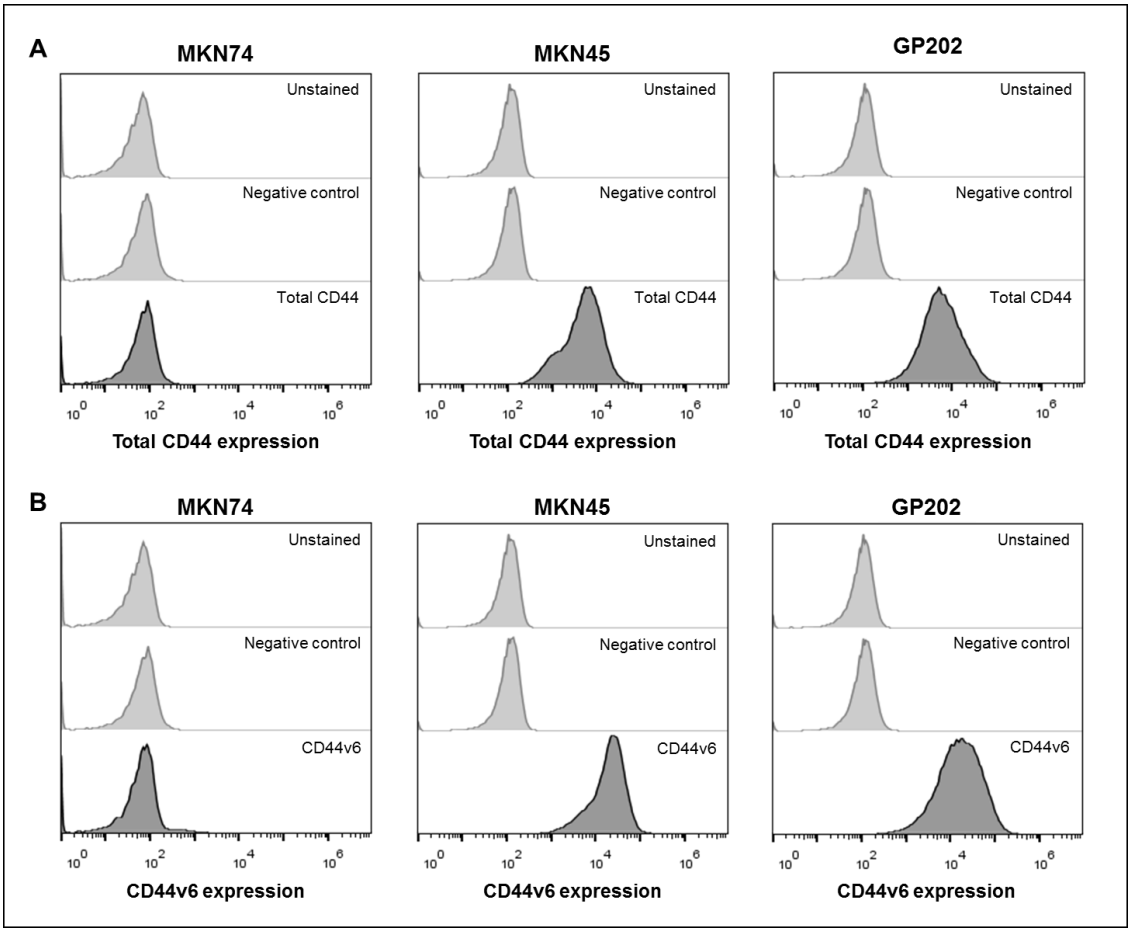


Figure S1: Confirmation that MKN74 cell line is CD44-null and does not express CD44, as determined by flow cytometry. **(*A*)** No expression of the standard version of total CD44 is detected in MKN74 cells (MKN45 and GP202 cells were used as a positive control for the expression of total CD44); **(*B*)** No expression of CD44v6 is detected in MKN74 cells (MKN45 and GP202 cells were used as a positive control for CD44v6 expression).

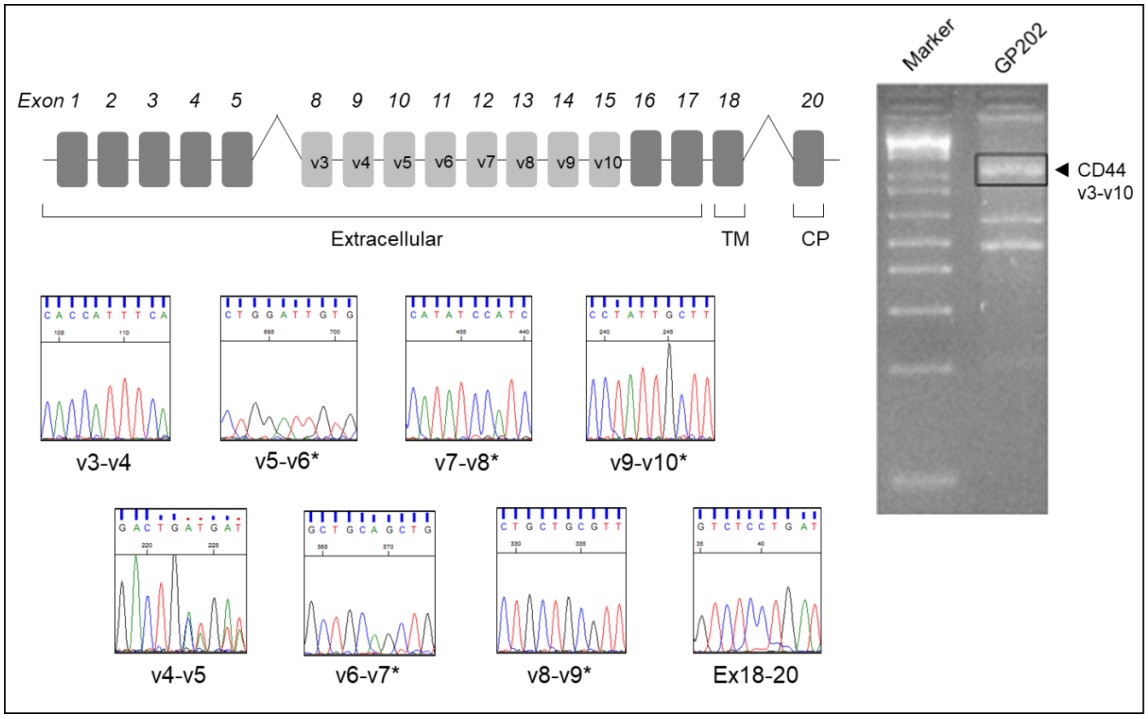


Figure S2: Identification of CD44v3-v10 variant endogenously expressed in the GP202 cell line. The band outlined in the gel represents the CD44v3-v10 transcript. Identity of the transcript was confirmed by Sanger sequencing. Each exon-exon junction in the boundary region was positively identified. It should be noted that the v4-v5 exon boundary displays a double reading frame after the Exonv5 start. One of the sequences has a CAG insertion identified using the mutation surveyor software. *These sequences were obtained with a reverse primer and are shown as the reverse complement of the forward strand.

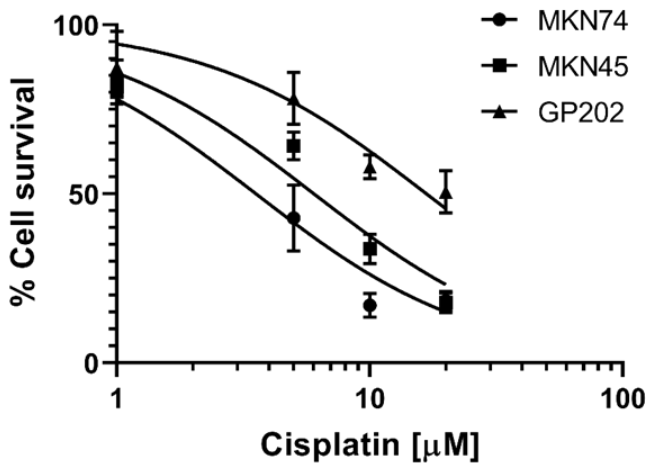


Figure S3: Determination of IC50 values in MKN74, MKN45 and GP202 cell lines. Cells were treated for 48 h with various cisplatin concentrations (1, 5, 10 and 20 μM) and cell survival was determined using a resazurin-based assay (PB). IC50 values were determined by non-linear regression analysis, as follows: MKN74 IC50 ~ 3.6 μM; MKN45 IC50 ~ 6 μM and; GP202 IC50 ~ 17 μM.

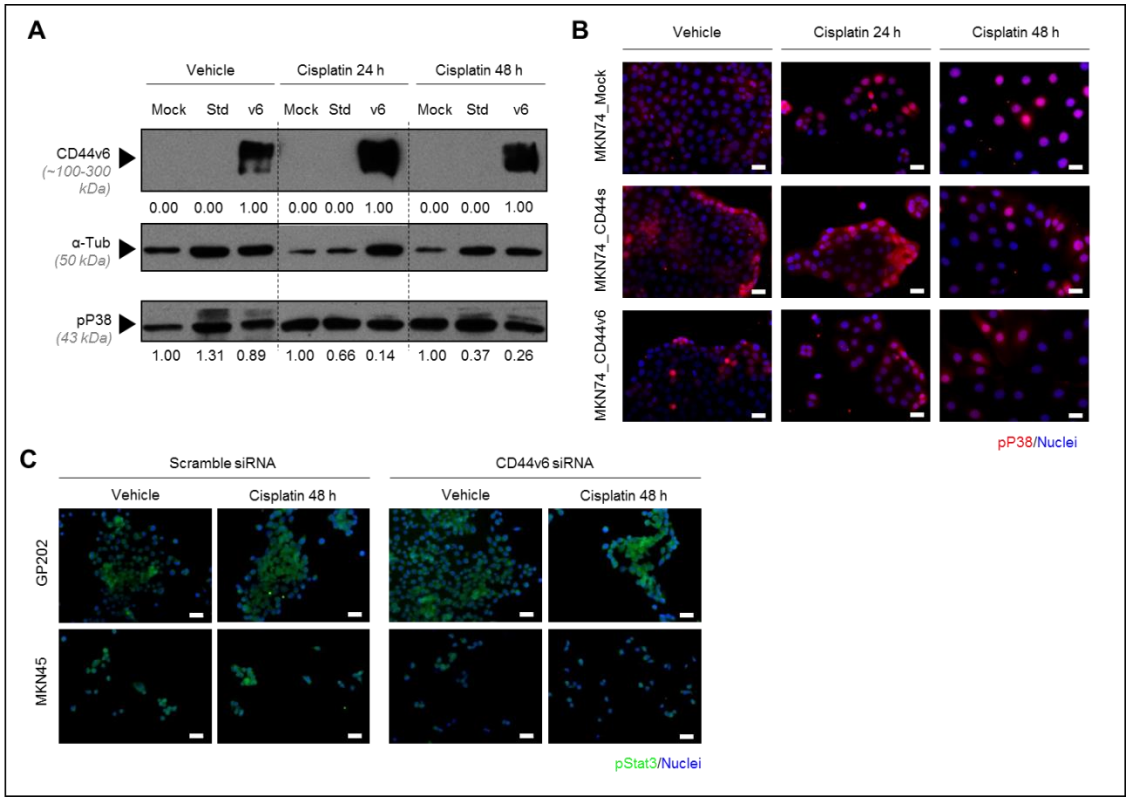


Figure S4: Expression of pP38 and pSTAT3 in MKN74 and GP202/MKN45 cells, respectively. **(A)** Western blotting of pP38 in MKN74 cell lines. CD44v6 and pP38 were run in the same gel as pSTAT3 (shown in Figure 4D) against the same tubulin; **(B)** Immunofluorescence of pP38 (seen in red) in untreated cell lines and upon treatment with cisplatin for 24 and 48 hours; **(C)** Immunofluorescence of pSTAT3 (seen in green) in vehicle and cisplatin treated GP202 and MKN45 cell lines (following a 24 h pre-incubation with scramble or CD44v6 siRNAs). Nuclei are stained with DAPI (seen in blue) and white scale bars represent a distance of 50 µm.

Table S1: Primers used in the characterization of v6 containing transcripts in the GP202 cells.

Orientation	In-text Primer Name	Binding Site	Primer Sequence (5' – 3')	Melting Temperature	Comment
Forward	A	Exon 5	CATCCCAGACGAAGACAGTCC	Tm = 66.0 °C	
	B	Exon v6	CCCATTCGACAACAGGGACA	Tm = 69.1 °C	
	C	Exon 5/ v3	GAATCCCTGCTACCAGTACG	Tm = 61.1 °C	
Reverse	D	Exon v6	TTGGCGATATCCCTCATGCC	Tm = 68.3 °C	
	E	Exon 16/ 17	GTGTCCATCTGATTCAGATCCA	Tm = 63.9 °C	
	F	Exon 17/ 18	TGATCAGCCATTCTGGAATTTG	Tm = 65.9 °C	
	G	Exon 19	GAATCTCTTCAACTTCTTCGAC	Tm = 58.6 °C	Amplifies specifically the short tailed isoform
	H	Exon 20	TCTTCATGTCCACATTCTGC	Tm = 61.4 °C	Amplifies the long and short tailed isoforms

Skipping exon-v6 from CD44v6-containing isoforms influences chemotherapy response and self-renewal capacity of gastric cancer cells

Silvana Lobo*, **Carla Pereira***, Gabriela M. Almeida, Carla Oliveira

**These authors contributed equally to this work.*

Cancers ***2020****,* 12(9), 2378, DOI: 10.3390/cancers12092378

ABSTRACT

De novo expressed CD44 isoforms containing exon-v6 are frequently associated with gastric cancer (GC) aggressiveness, and may predict chemotherapy response in vitro. Whether exon-v6 itself is responsible for conferring these properties to CD44v6-containing isoforms remains to be elucidated. CRISPR/Cas9 and Phosphorodiamidate Morpholino oligomers (PMOs) were used to induce specific exon-v6 skipping, maintaining the CD44 reading frame, in two GC cell lines endogenously expressing CD44v6. Cisplatin and 5-fluorouracil treatment response, and self-renewal ability was compared between CRISPR/Cas9-edited, CD44v6 knockdown and mock cells. We obtained homozygous genome-edited cell lines with exon-v6 deletion. Edited cells transcribed CD44v isoforms presenting in frame v5–v7 splicing, mimicking exon-v6 skipping. Results showed that removing specifically exon-v6 sensitizes cells to cisplatin and impairs cells' self-renewal ability, similarly to CD44v6 knockdown. In parallel, we also tested a clinically feasible approach for transient exon-v6 skipping with a PMO-based strategy. We demonstrate that exon-v6 specific removal from CD44v isoforms increases cell sensitivity to cisplatin and impairs GC cells self-renewal. We trust that a PMO approach designed towards CD44v6 overexpressing GC cells may be a suitable approach to sensitize tumor cells for conventional therapy.

CRISPR/Cas9; Chemoresistance; Exon skipping; Morpholinos; Cell survival; Stomach neoplasms

INTRODUCTION

Gastric cancer (GC), the third leading cause of cancer-related death worldwide, is often diagnosed in advanced/unresectable stages, where patients are treated with conventional chemotherapy and have an expected survival of only 1 year [1]. Thus, it is crucial to improve treatment effectiveness and patients' outcome. CD44 is a family of glycoproteins encoded by CD44 gene (locus 11p13) [2, 3], that mediate cell-cell and cell-matrix contact essential for tissue integrity and maintenance. Human CD44 gene (NG_008937) is composed of two terminal constitutively expressed portions (exons 1–5/16–18/20), and in between, by different combinations of alternatively spliced exons (exons 7–15/v2–v10) [2, 4, 5]. CD44v6-containing isoforms encompass a group of isoforms, including variant exon-6 (v6) [3, 6], a cancer stem cell (CSC) marker widely associated with poorer patient prognosis, increased invasion, metastization and drug resistance in several cancers, including GC [7-12]. Blockade of CD44v6 delays tumor growth and metastization in several in vivo models of pancreatic [13] and colorectal cancer [14]. Therefore, targeting CD44v6-expressing cancer cells arises as a suitable cancer therapy candidate. This strategy was tested in head and neck squamous cell carcinoma with the development of a monoclonal antibody against CD44v6 (BIWA) [15], combined with mertansine, a chemotherapeutic cytotoxic drug. This clinical trial was discontinued, as one participant passed away due to severe skin toxicity, since keratinocytes express CD44v6-containing isoforms [16]. CD44v6-containing isoforms are not expressed in normal gastric mucosa, but become overexpressed in pre-neoplastic gastric lesions and tumors [17]. Recently, we reported that GC cell lines expressing CD44v6-containing isoforms are more resistant to cisplatin than cells lacking CD44v6 expression [18]. Still, it remains to be elucidated whether it is exon-v6 itself, or the isoforms containing exon-v6, that modulate chemotherapy response in cancer cells. Herein, we hypothesize that the presence/absence of exon-v6 in CD44v6-containg isoforms conditions the chemotherapy response of GC cells, and that specific exon-v6 skipping can sensitize GC cells to chemotherapy.

In this study, we used CRISPR/Cas9 to remove exon-v6 from the CD44 locus, by targeting splice-sites and/or adjacent regions, to generate specific exon-v6 skipping from CD44v6-containing isoforms in two GC cell lines. Our results show that specific exon-v6 skipping induces increased cell sensitivity to cisplatin and decreased self-renewal capacity, to the same extent as the full depletion of CD44v6-containing isoforms. In sum, our data highlight a feasible alternative to full CD44v6-containing isoforms depletion, raising the possibility of using exon-v6 skipping as targeted therapy to delay GC cells self-renewal and sensitize gastric tumors to chemotherapy. This is expected to decrease off-target effects observed in previous approaches.

MATERIALS AND METHODS

All reagents were purchased from ThermoFisher Scientific (Waltham, MA, USA), unless otherwise stated.

1. Cell Culture

Human GC cell line MKN45 was purchased from the Japanese Collection of Research Bioresources Cell Bank, and the non-commercial GP202 cell line was established at IPATIMUP [19]. Both cell lines were cultured in RPMI with 10% fetal bovine serum (FBS) (Biowest, Nuaillé, France), and maintained at 37 °C under 5% CO2 humidified atmosphere. Cells were grown in the absence of antibiotics (unless when puromycin selection was performed) and confirmed to be mycoplasma free.

2. Generation of a Permanent Exon-v6 Skipping Model by CRISPR/Cas9

CRISPR/Cas9 is a gene editing tool to permanently modify the genome, either by insertions, deletions or point mutations. sgRNAs were specifically designed to target a specific DNA portion via Watson-Crick base paring, and thus precisely guide the Cas9 protein to the region to be edited. The Cas9 protein recognizes a protospacer adjacent motif (PAM) sequence, causing a double-strand break (DSB) in the target DNA, that can then be repaired by the non-homologous end-joining (NHEJ) or the homology-directed repair (HDR) pathways [20]. All sgRNAs were designed to target the exon-v6 adjacent areas using Benchling online platform, and were purchased from Sigma-Aldrich (Poole, UK). Six sgRNAs were designed (**Figure 1B, C**) and individually cloned into pSpCas9(BB)-2A-Puro (PX459) V2.0 plasmid (Addgene plasmid #62988, Watertown, MA, USA), in the Bbs I restriction site, according to the "Morrisey Lab Protocol" [21], with few adaptations. Each Sg plasmid was transformed into Escherichia coli one shot Match1TM-T1R (Invitrogen, Carlsbad, CA, USA), according the manufacturer's instructions. A colony PCR was performed to screen the colonies using the Primers A and D from Table 1. The positive colonies were amplified and sequenced using the same primers. The sgRNA were transfected in pairs to MKN45 and GP202 GC cell lines. As control, GC cell lines were transfected with an empty vector (mock). Briefly, cells were plated in 12 well plates and allowed to grow for 24 h, until approximately 70% confluency. Afterwards, Sgs' plasmid or the empty plasmid were complexed with 1.75 µL of Lipofectamine 3000 transfection reagent, according to the manufacturer's instructions. After 48 h, puromycin (Merck, Darmstadt, Germany) was added to the cells to select vector positive cells. Puromycin was renewed every 72 h, until the non-transfected cells, which do not have the antibiotic selection gene to puromycin carried by the plasmid, were dead.

3. Genotyping Analysis of CRISPR/Cas9 Skipping Models

gDNA was extracted using NZY Tissue gDNA Isolation kit (NZYTech, Lisbon, Portugal) and RNA was isolated using TriPure Isolation Reagent (Sigma-Aldrich, Poole, UK), according to the manufacturers' protocol. NanoDrop® ND-1000 UV-Vis Spectrophotometer (ThermoFisher Scientific, Waltham, MA, USA), was used to determine gDNA and RNA quality and concentration. cDNA was synthesized using 1 µg of template RNA and SuperScript® II reverse transcriptase, according to the manufacturers' protocol. Non-RT negative controls were produced replacing SuperScript® II for

sterile water. gDNA and cDNA were amplified with multiplex PCR kit (Qiagen, Venlo, The Netherlands) and primers designed to flank exon-v6, in the flanking introns (primers C/F were used to characterize gDNA) or flanking exons (primers B/E were used to characterize cDNA). PCR products were analyzed through 2% agarose gel electrophoresis and sanger sequencing.

4. CD44/CD44v Expression Analysis

CD44v6 and total CD44 gene expression were assessed by RT-qPCR, using specific probes for CD44v6 (exon span v5–v6, Hs.PT.58.45400024) and CD44 total (exon span 2–3, Hs.PT.58.4880087) (both from iDT, Coralville, IA, USA). RT-qPCR was performed in triplicate, using 1 µL of template cDNA per well. TaqMan Master Mix was used to amplify PCR product and ABI Prism 7500 Fast Sequence Detection System was used to quantify the expression. Relative expression was calculated by comparative $2^{-\Delta\Delta C_T}$ method using the housekeeping gene 18 s (custom assay) (iDT, Coralville, IA, USA).

CD44v6 and total CD44 protein expression were assessed by flow cytometry. Cells were detached using versene reagent and blocked with 3% bovine serum albumin (BSA)–phosphatase buffer saline (PBS), for 30 min. Cells were incubated with the following primary antibodies: mouse monoclonal antibody against total CD44 (156-3C11; 1:100 dilution; 60 min; Cell Signaling Technology, Beverly, MA, USA) and CD44v6 (MA54; 1:100; 60 min). Cells were washed and subsequently incubated with secondary antibody anti-mouse Alexa Fluor 647 (1:500; 30 min). Fluorescence was measured using FACS Canto II (BD Biosciences, Franklin Lakes, NJ, USA). Flow Jo version 10 software was used to analyze the data.

Total CD44, CD44v6 and CD44v9 protein expression levels and subcellular location were assessed by immunofluorescence. After cell growth in glass coverslips, cells were fixed with 4% paraformaldehyde (PFA) (Merck, Darmstadt, Germany). Cells were permeabilized using 0.2% Triton X-100 (Sigma-Aldrich) and subsequently blocked with 5% BSA-PBS. The primary antibodies used were mouse monoclonal antibody against CD44v6 (MA54; 1:100; overnight (ON); 4 °C), rat monoclonal antibody against CD44v9 (RV3; 3 µg/mL; ON; 4 °C; AB Nova, Taipé, Taiwan), and mouse monoclonal antibody against CD44 total (156-3C11; ON; 1:100 dilution; Cell Signaling Technology, Beverly, MA, USA). After washings with PBS, cells were incubated with the appropriate secondary antibodies: Alexa Fluor 594 Goat Anti-Mouse secondary antibody (1:500; 2 h; Life Technologies, Carlsbad, CA, USA) and Alexa Fluor 488 Goat Anti-Rat secondary antibody (1:500; 2 h; Life Technologies). Nuclei were stained with DAPI (1:1000; 5 min; Sigma-Aldrich). Vectashield mounting media (Vector Laboratories, Burlingame, CA, USA) was used to mount the cover slips. Fluorescence was measured (AxioCam fluorescence microscope, Zeiss, Gottingen, Germany) and data were analyzed using AxioVision software version 4.8. (Rockville, MD, USA).

5. Inhibition of CD44v6 Expression by siRNA

Mock cells and CRISPR/Cas9 edited cells were seeded in 6 well plates (1.5×10^5 and 2×10^5 cells/well for MKN45 and GP202, respectively) and allowed to adhere ON. In the following day, Mock cells were transfected with 20 nM scramble (negative control; DS NC1; iDT, Coralville, IA, USA) or CD44v6 siRNA (Sense strand: 5′-GCGUCAGGUUCCAUAGGAAUCCUTT-3′; Antisense strand: 5′-AAAGGAUUCCUAUGGAACCUGACGCAG-3′; iDT, Leuven, Belgium), using Lipofectamine® RNAiMax transfection reagent, according to the manufacturers' instructions. CRISPR/Cas9 edited cells were also transfected with scramble siRNA to ensure the same conditions between mock, KD and CD44v6 edited cells for the following assays. After 24 h of incubation, transfected cells were detached, counted and re-plated in 96- (2500 and 4000 cells for MKN45 and GP202 respectively) and 6-well plates (1000 and 3000 cells for MKN45 and GP202 respectively), for short-and long-term assays, respectively.

6. Cell Survival Assays

After transfection, and 24 h after cell re-plating, cells were treated with cisplatin (reconstituted in 0.9% NaCl), 5-FU (reconstituted in sterile water) or vehicle (0.9% NaCl). The concentrations used in short-term assays were: 2.5 µM (MKN45) and 20 µM (GP202) for cisplatin and 5 µM (MKN45 and GP202) for 5-FU. The short-term effects of cisplatin and 5-FU on cell survival were assessed after 48 h of treatment, using PB (Invitrogen) and SRB (Sigma-Aldrich) assays. PB reagent was diluted in RPMI to 1x and added to cells. Cells were incubated for 45 min at 37 °C and fluorescence measured (with an excitation wavelength of 560 nm and emission of 590 nm). Afterwards, cells were fixed in 10% trichloroacetic acid (TCA) (Merck, Darmstadt, Germany) for 1 h on ice, and proteins were stained with 4% SRB solution (Sigma-Aldrich), for 30 min at room temperature. Wells were repeatedly washed with 1% acetic acid to remove unbound dye. Protein stain was solubilized with 10 mM Tris solution (Calbiochem, San Diego, CA, USA) and absorbance read at 560 nm, with background correction at 655 nm. Fluorescence and absorbance were assessed in a microplate reader PowerWave HT Microplate Spectrophotometer (BioTek, Bad Friedrichshall, Germany). To calculate the percentage of cell survival, fluorescence (Fluo)/absorbance (Abs) readings of treatment conditions were normalized to the vehicle (% cell survival = (Treatment Fluo/Abs * 100)/Vehicle Fluo/Abs).

The long-term effects of cisplatin and 5-FU treatment were analyzed using the clonogenic assay. Briefly, after transfection and 24 h after cell re-plating, cells were treated with 0.5 µM cisplatin, 2.5 µM or 2 µM of 5-FU (for MKN45 and GP202, respectively) and vehicle. After 48 h drug treatment, RPMI medium was renewed, and cells were incubated under normal conditions (5% CO2 humidified atmosphere at 37 °C) for ~10 days, until visible colonies had about 50 cells each. Afterwards, cells were fixed with methanol (Fisher Scientific, Hampton, NH, USA), stained with 0.5% crystal violet (Sigma-Aldrich), and colonies counted. To calculate the percentage of cell survival, the number of

colonies of each treatment conditions was normalized to vehicle (% cell survival = (Treatment No. of colonies * 100)/Vehicle No. of colonies).

7. Transient Exon-v6 Skipping Using Morpholinos

To understand the therapeutic potential of exon-v6 skipping, we tested a transient strategy approved in the clinics for other diseases. To this end, we checked the possible splice sites near exon-v6, using Human Splicing Finder online tool [22], and designed PMOs to target the specific splice acceptor and donor at the 5′ and 3′ sites of exon-v6, respectively. Two PMOs were designed: "CD44v6_BEG" (5′-CTGGACTGTGAGAAGAATATCAGTT-3′), at the 5′ site of exon-v6 and "CD44v6_END" (5′-CTTGTTAAACCATCCATTACCAGCT-3′), at the 3′ site of exon-v6. Both GC cell lines were seeded in 12-well plates and allowed to adhere ON, until approximately 60–70% confluency. In the following day, culture medium was replaced and PMOs (alone or mixed) were transfected into cells using Endo-Porter transfection reagent (Gene Tools, Philomath, OR, USA) and incubated for 48 h. The concentrations of each PMO used to transfect cells were 4 µM along with 2 µM of Endo–Porter. After this period, cells were collected, either for genotyping or immunofluorescence analysis.

8. Statistical Analysis

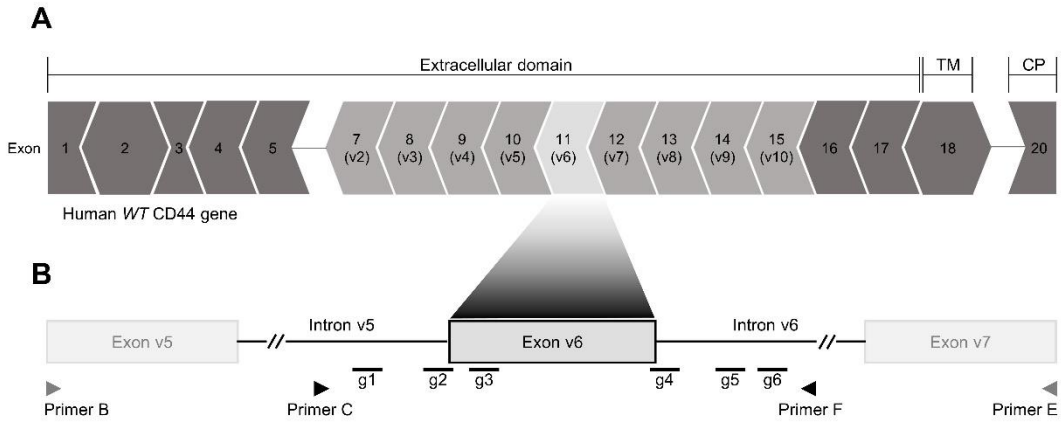
Statistical analysis was performed using GraphPad Prism version 7.00 software (GraphPad Software Inc., San Diego, CA, USA). A normality test was performed using a Shapiro-Wilk test. Two-way ANOVA with Tukey's post hoc test for multiple comparison analysis was used, with 95% confidence interval. Significant differences were considered significant when $p < 0.05$.

RESULTS

We aimed at understanding if the presence of exon-v6 in CD44v isoforms may affect GC chemotherapy response and cell self-renewal, and for this, we developed an exon-v6 specific skipping strategy.

1. CRISPR/Cas9 Editing at Exon-v6 Splice-Sites Generates CD44v Isoforms Lacking Exon-v6

We established a CRISPR/Cas9 strategy to target the adjacent areas of exon-v6 (**Figure 1A**), by designing single guide RNAs (sgRNAs) (g1–g6) in the exon-v6 vicinity (**Figure 1B, C**), to achieve exon-v6 skipping. After genome editing, we obtained two edited cell lines for MKN45 (O26 and O15) and GP202 (O26 and O25). To evaluate whether our strategy successfully removed exon-v6 from CD44v isoforms, genomic DNA (gDNA) was amplified, using primers for exon-v6 flanking introns (Primers C/F, **Table 1**).



C

sgRNA	Sequence	PAM	Location
g1	5' CACCGTATGTTGACAGCTATTGGTG 3'	AGG	Intron v5
g2	5' CACCGTGATATTCTTCTCACAGTCC 3'	AGG	Intron v5 ^ Exon v6
g3	5' CACCGGGCAACTCCTAGTAGTACAA 3'	CGG	Exon v6
g4	5' CACCGCAGGGACAGCTGGTAATGGA 3'	TGG	Exon v6 ^ Intron v6
g5	5' CACCGGGGTTAAACGGTAGACATTG 3'	AGG	Intron v6
g6	5' CACCGAGGAATTGTCACGAGATGTT 3'	AGG	Intron v6

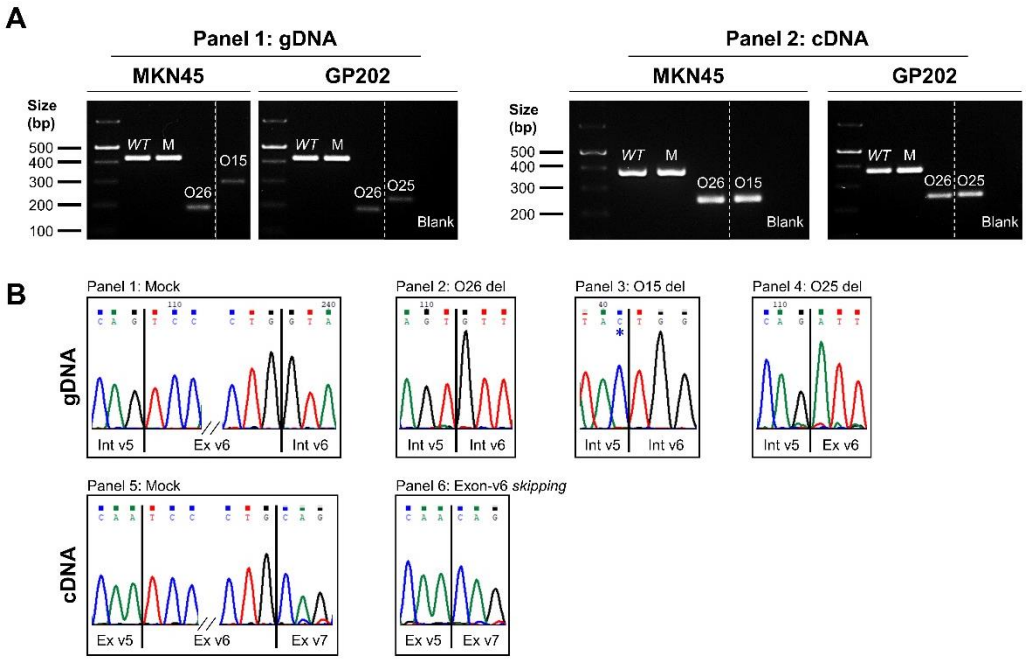
Figure 1. Human CD44 locus and CRISPR/Cas9 single guide RNA (sgRNA) design strategy. (A) The human CD44 canonical transcripts annotated with 18 exons (ENST00000428726.8). The dark grey exons comprise the constitutively expressed portion of the gene (exons 1–5, 16–18 and 20), and the light grey corresponds to the alternatively expressed portion of the gene in humans (exons v2–v10). Each exon is docked to adjacent exons in the mRNA sequence. Every set of three nucleotides encodes the amino acids that will comprise the CD44 protein. In the exon boundaries, the resultant amino acid may enclose the last nucleotide of the former exon and the two first nucleotides from the next exon (>), or contain the two last nucleotides of the former exon and the first nucleotide from the next exon (<). This indicates that if exon-v6 is removed, exons v5 and v7 are spliced in-frame. (B) Genomic spots of the sgRNA designed to target the adjacent areas to exon-v6. Primers B and E (represented in grey) were used to genotype the complementary DNA (cDNA) and primer C and F (represented in black) were used to genotype the genomic DNA (gDNA) of the edited cells. (C) Specifications of the sgRNAs (g1–g6) designed near the exon-v6 vicinity to permanently induce exon-v6 skipping. TM—Transmembrane domain; CP—Cytoplasmic domain; WT—wild-type; PAM—Protospacer Adjacent Motif.

Table 1. Details of the primers designed to validate CRISPR/Cas9 constructs (primer A and D) and genotype the exon-v6 skipping cell models obtained by CRISPR/Cas9 and Phosphorodiamidate Morpholino Oligomers (PMOs) approaches (primer B, C, E and F).

Orientation	In-Text Name	Binding Site	Primer Sequence (5′–3′)	Melting Temperature
Forward	A	pSpCas9(BB)_Bbs I	5′-GGGCCTATTTCCCATGATTCCTT-3′	Tm = 68.5 °C
	B	CD44 Exon 10 (v5)	5′-ATGTAGACAGAAATGGCACCAC-3′	Tm = 62.7 °C
	C	CD44 Intron 10	5′-ATCAGTGGCCTGTTTCCTTG-3′	Tm = 64.0 °C

	D	pSpCas9(BB)_Bbs I	5′-GACTCGGTGCCACTTTTTCAA-3′	Tm = 66.2 °C
Reverse	E	CD44 Exon 12 (v7)	5′-CCATCCTTCTTCCTGCTTGATG-3′	Tm = 66.8 °C
	F	CD44 Intron 11	5′-TTTGGCTCTGTGTGAACTGC-3′	Tm = 64.1 °C

Resulting products were smaller in edited cells than in wild-type (WT) and mock controls (Figure 2A, left panel). This was confirmed by Sanger sequencing **(Figure 2B**, panel 1–4), showing that all CRISPR/Cas9 edited cell lines presented homozygous deletions encompassing exon-v6 (**Figure 2C**). To evaluate specific exon-v6 skipping, complementary DNA (cDNA) was amplified with primers in v5 and v7 adjacent exons (Primers B/E, **Table 1**). All edited cell lines presented a smaller transcript, corresponding to splicing between exons v5-v7 and in-frame exon-v6 skipping (**Figure 2A**, right panel; **Figure 2B**, panels 5–6).



C

Edited cell line	Genotyping	Genomic Coordinates (GRCh38/hg38)	Combination of sgRNAs
MKN45_Mock	WT CD44v6	Chr 11: 35,139,171 – 35,232,402	Empty vector
MKN45_O26	Homozygous deletion of 236 bp	Del: 35,204,513 – 35,204,748	g2 + g6
MKN45_O15	Homozygous deletion of 131 bp + insertion of 2 bp	Ins C: 35,204,448 Del: 35,204,449 – 35,204,580 Ins A: 35,204,704	g1 + g5
GP202_Mock	WT CD44v6	Chr 11: 35,139,171 – 35,232,402	Empty vector
GP202_O26	Homozygous deletion of 236 bp	Del: 35,204,513 – 35,204,748	g2 + g6
GP202_O25	Homozygous deletion of 190 bp	Del: 35,204,512 – 35,204,702	g2 + g5

Figure 2. Genotyping analysis of CRISPR/Cas9 edited GC cells. (A) Analysis of gDNA (Panel 1) and cDNA (Panel 2) from wild-type (WT), Mock and edited GC cell lines. In Panel 1, MKN45/GP202_Mock cells present a transcript of the same size of WT (422 bp), while the edited cell lines present a transcript with smaller sizes (MKN45_O26 and GP202_O26–186 bp; MKN45_O15–293 bp; GP202_O25–232 bp). In Panel 2,

MKN45/GP202_WT and Mock cells present a transcript with a size of 378 bp, while all edited cell lines present a smaller transcript (249 bp) corresponding to the occurrence of exon-v6 skipping. All PCR products were run in the same 2% agarose gel (**Figure S1**). (B) Representative examples of sequencing analysis: Panel 1—gDNA of Mock cells; Panel 2—gDNA deletion of 235 bp of O26 clones. (The present sequencing analysis was performed with gDNA extracted from MKN45_O26, but GP202_O26 presents the same deletion); Panel 3—gDNA deletion of 131 bp of O15 clone (* C nucleotide is a single nucleotide insertion); Panel 4—gDNA deletion of 190 bp of O25 clone; Panel 5—cDNA of Mock cells; Panel 6—cDNA of all edited clones, which represents the skipping of exon-v6. (C) Specifications of each edited cell line by CRISPR/Cas9 with the respective edition genomic coordinates (GRCh38/hg38) and the combination of sgRNAs transfected to create each edited cell line. WT–wild-type; M–Mock; gDNA–genomic DNA; cDNA–complementary DNA; Chr—Chromosome; Del—Deletion; Ins—Insertion.

By real-time quantitative PCR (RT-qPCR), we verified that our editing strategy was specific for CD44v6-containing isoforms, as no exon-v6 expression was detected in edited cells, while total CD44 expression was maintained (**Figure 3A**). By flow cytometry, we showed that edited cell lines presented depletion of the exon-v6 encoded peptide, while maintaining equivalent total CD44 protein levels as mock cells (**Figure 3B**). To demonstrate that only the v6-encoded domain was skipped, we performed co-immunofluorescence for CD44v6 and CD44v9, a variant exon downstream of exon-v6 that is often included in CD44v6-containing isoforms. As expected, exon-v6 and exon-v9 were co-detected in mock cells, while in edited cells, only exon-v9 was detected (**Figure 3C**). When using v6-directed siRNAs (CD44v6 KD cells), both exon-v6 and exon-v9 expression disappeared, demonstrating that exon-v6 and exon-v9 are both present in the same CD44v6-containing isoforms detected with this approach. These data support that, exon-v6 skipping in edited cells occurs without interfering with downstream exon-v9, demonstrating that CD44v isoforms expression remains unchanged.

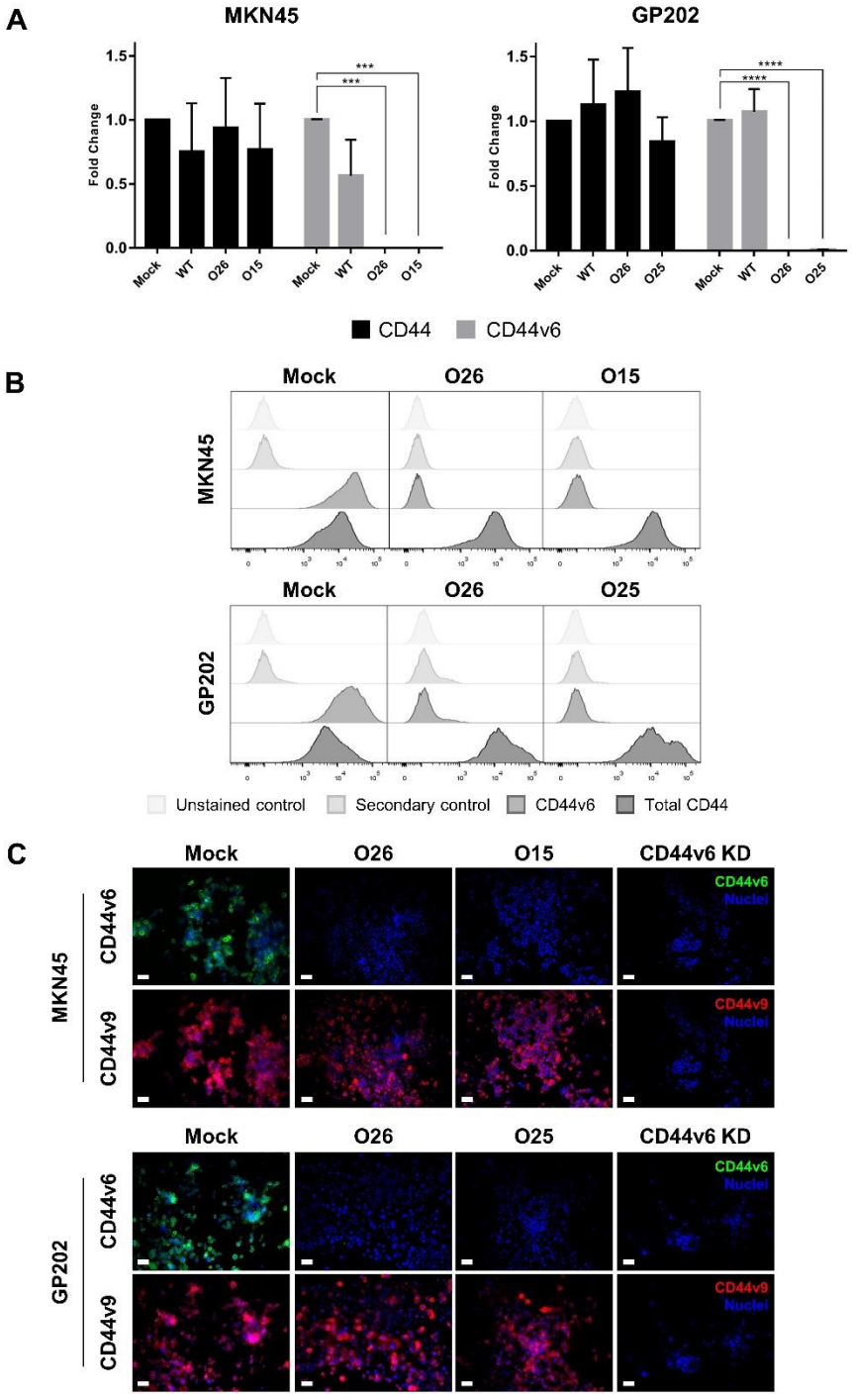


Figure 3. Characterization of CRISPR/Cas9 edited cells and CD44v6 knockdown (KD) cells (using siRNA) at the RNA and protein level for two GC cell lines. (A) RT-qPCR representing the fold-change of CD44v6 and total CD44 gene expression in MKN45 and GP202 WT, Mock and edited cells. Results represent the mean + SD of three independent experiments. Statistically significant results were determined by Two-way ANOVA with Tukey's Post Hoc Test for multiple comparison analysis (*** $p < 0.001$; **** $p < 0.0001$). (B) Protein expression analysis by flow cytometry for CD44v6 and total CD44 expression in GP202 and MKN45 cells; (C) Co-immunofluorescence analysis for CD44v6 (green) and CD44v9 (red) expression in GP202 and MKN45 cells. To facilitate the visualization, CD44v6 and CD44v9 panels are depicted in separate, but were performed in the same slide, and membranous CD44v6 and CD44v9 staining are shown in greater detail in **Figure S2**. Nuclei

were stained with DAPI (blue) and white scale bars represent a distance of 50 µm. WT—wild-type; KD—Knockdown.

2. *CD44v6-Containing Isoforms and Exon-v6 Itself are Positive Modulators of Cisplatin Response*

We previously demonstrated that MKN45 and GP202 GC cell lines, which express CD44v6-containing isoforms, are less sensitive to cisplatin than CD44v6-depleted counterparts [18]. To evaluate if exon-v6 is, itself, responsible for modulation of chemotherapy response in these cell lines, we performed short and long-term assays in the presence of cisplatin and 5-fluorouracil (5-FU). For that, we compared mock cells treated with scramble siRNA (scramble), with CD44v6 knockdown (KD) cells, or with exon-v6 edited cells. In both cell lines, 48 h after treatment, we observed that CD44v6 KD cells survived better than scramble in presto blue (PB) and sulforhodamine B (SRB) assays (**Figure 4**, left panel). The outcome of short-term treatment in MKN45 and GP202 exon-v6 edited cells mimicked the observations for CD44v6 KD for at least one clone per cell line, but only for cisplatin treatments, and mainly for the PB assay. For SRB assay, only MKN45_O26, treated either with cisplatin or 5-FU, mimics the significant cell survival increase observed in CD44v6 KD. Regarding the long-term survival, clonogenic assay demonstrated the opposite, as CD44v6 KD cell survival significantly decreased when compared to scramble, in response to both drugs (**Figure 4**, left panel). The same was true for both GP202 edited cells, which presented a statistically significant or strong tendency for decreased cell survival to cisplatin compared to control (**Figure 4**, bottom right panel). Overall, these results demonstrate that exon-v6 skipping mimics to some extent the effect that CD44v6 KD have in GC cell survival. Moreover, it consolidates the role of both CD44v6-containing isoforms, and exon-v6 alone, as positive modulators of GC survival.

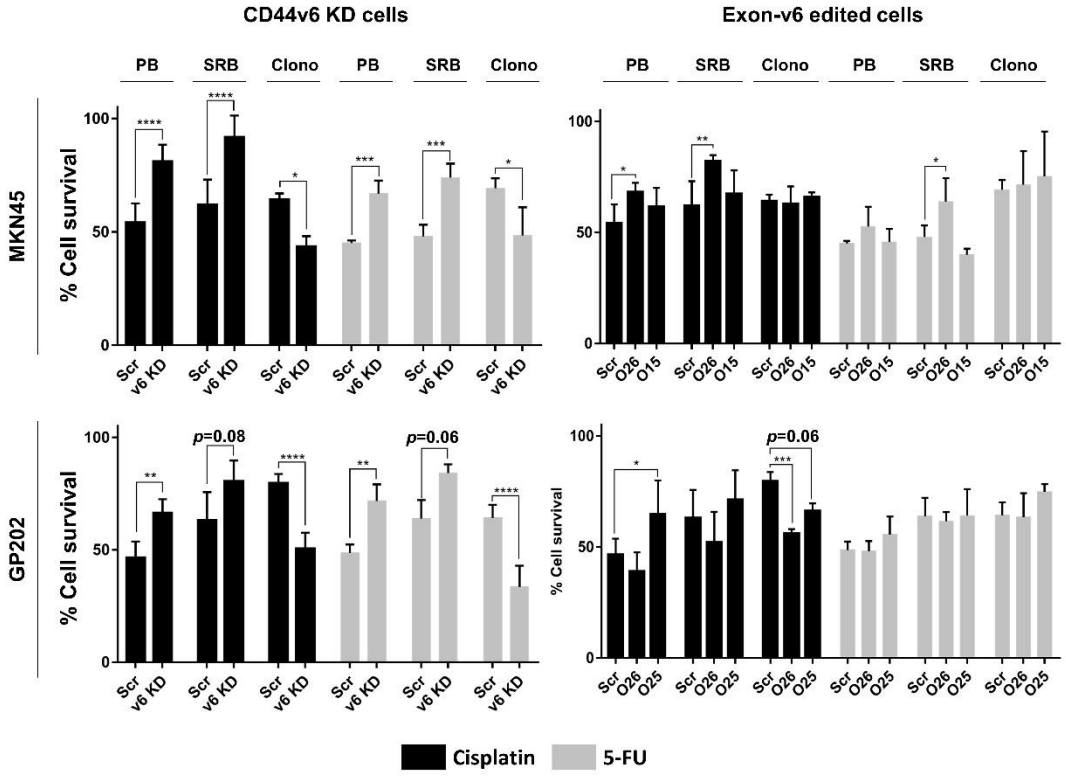


Figure 4. Cell survival in response to cisplatin or 5-fluorouracil (5-FU) in MKN45 and GP202 cells. Percentage cell survival of CD44v6 expressing cells (Scramble) vs. CD44v6 KD cells and Scramble vs. exon-v6 depleted cells (O26, O15, O25) to cisplatin (represented in black) and 5-FU (represented in grey) treatment analyzed by Presto Blue (PB) and Sulforhodamine B (SRB) assays, after 48 h and clonogenic assay, after 10 days. The concentrations used in PB and SRB assays were: 2.5 µM (MKN45) and 20 µM (GP202) of cisplatin and 5 µM (MKN45 and GP202) of 5-FU. The concentrations used for the clonogenic assay were: 0.5 µM (MKN45 and GP202) for cisplatin and 2.5 µM (MKN45) and 2 µM (GP202) for 5-FU. Results represent the mean + SD of three independent experiments. Statistically significant results were determined by two-way ANOVA with Tukey's Post Hoc Test for multiple comparison analysis (* $p < 0.05$; ** $p < 0.01$; *** $p < 0.001$; **** $p < 0.0001$). Src–Scramble; v6 KD–CD44v6 Knockdown; PB–PrestoBlue; SRB–Sulforhodamine B; Clono–Clonogenic; KD–Knockdown.

3. *Exon-v6 Depletion Decreases Self-Renewal Capacity of GC Cells to the Same Extent as Knocking-Down CD44v6-Containing Isoforms*

Next, we tried to understand whether self-renewal capacity, which can be assessed by calculating basal colony formation efficiency, was different between mock, CD44v6 KD and edited cells. We observed that both CD44v6 KD cell lines presented extremely low colony formation efficiency compared to scramble (**Figure 5**). Interestingly, the two edited clones that behaved similarly to CD44v6 KD (MNK45_O26 and GP2020_O25), in short- and/or long-term assays with cisplatin and 5-FU, were also those presenting a statistically significant decrease in colony formation efficiency (**Figure 5**). We tried to understand if deletion size or break-points' location could explain the different results obtained for the edited clones. Most likely, this does not provide an explanation,

as MKN45_O26 and GP202_O26 share exactly the same deletion (**Figure 2C**), and behave differently regarding drug response and colony formation efficiency. In summary, the latter results clearly demonstrate that exon-v6 skipping can have the same effect on cell self-renewal as full KD of CD44v6-containing isoforms; however, in the future, it would be important to extend this analysis to additional GC cell lines.

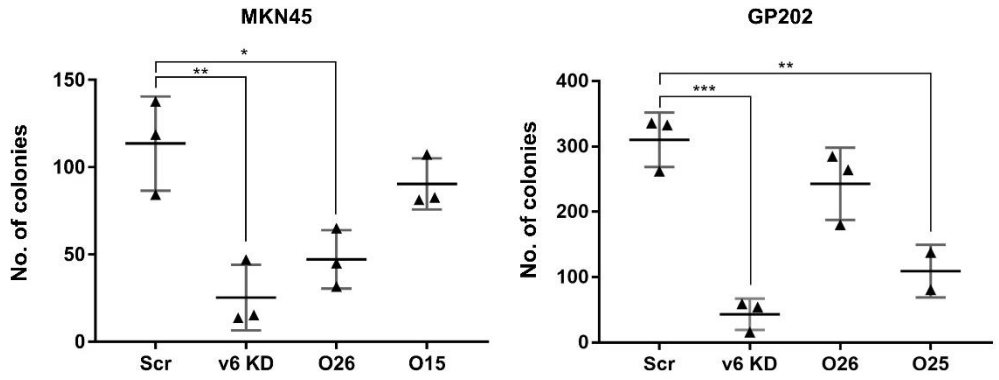


Figure 5. Self-renewal capacity of MKN45 and GP202 cells in response to the KD of CD44v6 isoforms or specific exon-v6 skipping. Colony formation capacity of Scramble, CD44v6 KD cells and exon-v6 edited cells lines. Results show the mean ± SD of three independent experiments. Statistically significant results were determined by two-way ANOVA with Tukey's Post Hoc Test for multiple comparison analysis (* $p < 0.05$; ** $p < 0.01$; *** $p < 0.001$; **** $p < 0.0001$). Src–Scramble; v6 KD–CD44v6 Knockdown; KD—Knockdown.

4. Transient Exon-v6 Skipping Using Phosphorodiamidate Morpholino Oligomers (PMOs)

Our results suggest that specific exon-v6 skipping significantly decreases GC cells self-renewal, while sensitizing GC cells to cisplatin in long-term assessment. Therefore, we hypothesize that using exon-v6 skipping therapy in CD44v6 expressing gastric tumors could improve GC patients' response to chemotherapy. Hence, we tried to use a morpholino-based transient strategy, which is already approved in the clinics for other diseases, to perform CD44 exon-v6 skipping. We designed PMOs directed to exon-v6 flanking splice-sites (**Figure 6A, B**). PMOs were expected to mask splice-sites, leading to the production of CD44v transcripts lacking exon-v6. We successfully reproduced the exon-v6 CRISPR/Cas9 editing in GP202 cells, either by using a single PMO or combining two PMOs. RNA analysis revealed the production of a smaller transcript (**Figure 6C**) compared to the one observed in parental cells, of which sequencing was confirmed to represent exon-v6 skipping (**Figure 6D**). Using immunofluorescence, we demonstrated that exon-v6 is undetectable, as a result of exon skipping, whereas total CD44 remains expressed (**Figure 6E**). These results indicate that the use of at least one PMO targeting an exon-v6 splice-site efficiently induces exon-v6 splicing out, creating an opportunity for clinical intervention.

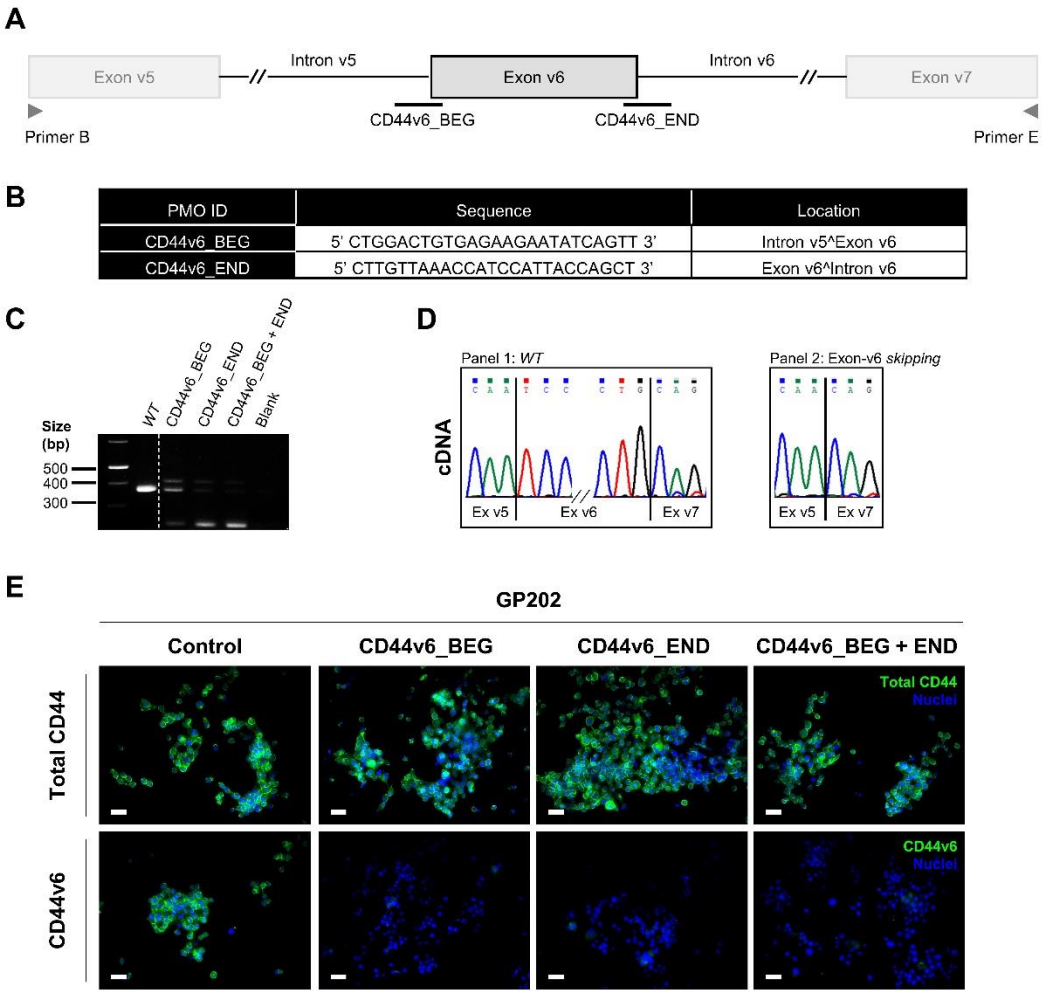


Figure 6. Characterization of transfected cells established from WT GP202 GC cell line, using a PMOs approach. (A) Scheme of the PMOs designed to target the splice-sites of exon-v6. Primers B and E (represented in grey) were used to genotype the cDNA of cells. (B) Specifications of PMOs (CD44v6_BEG and CD44v6_END) designed to target the splice-sites of exon-v6 and cause exon-v6 skipping. (C) Analysis of cDNA from WT and transfected cells with either single or mixed PMOs. Transfected cells present a smaller transcript (249 bp) than WT cells (378 bp), corresponding to the skipping of exon-v6. All PCR products were analyzed in the same 2% agarose gel (**Figure S3**). (D) Representative examples of the sequencing analysis: Panel 1—WT 378 bp transcript, which corresponds to the WT cDNA sequence; Panel 2—249 bp transcript of the transfected cells, which corresponds to the skipping of exon-v6. The sequencing data was generated from cDNA extracted from CD44v6_END transfected cells (E) Immunofluorescence analysis depicting total CD44 (green) and CD44v6 expression (green) of GP202 WT and transfected cells. To facilitate visualization, membranous CD44v6 staining is shown in greater detail in **Figure S4**. Nuclei were stained with DAPI (blue) and white scale bars represent a distance of 50 µm. WT—wild-type; cDNA—complementary DNA.

DISCUSSION

Aiming to unveil whether exon-v6 is, by itself, responsible for increased chemotherapy resistance in CD44v6 expressing GC cells and cell self-renewal, we generated permanent exon-v6 skipping models. We successfully obtained four edited cell lines with specific exon-v6 skipping, where in-frame splicing between exons v5 and v7 occurred, and where CD44v6 protein expression was

knocked-out, while total CD44 expression was maintained. This work confirms that CRISPR/Cas9 editing is efficient at generating CD44v isoforms lacking exon-v6. Moreover, our work indicates that this can be useful to develop in vitro models to study the function of specific exons, by inducing targeted exon skipping without destroying the protein of interest.

In order to mimic patients' treatments, we performed short-term drug assays, mimicking the acute phase after chemotherapy, and long-term drug assays, mimicking the response after a chemotherapy cycle. Short-term analysis suggests that, with CD44v6 KD, cells survive better to cisplatin and 5-FU, compared to CD44v6 expressing cells. CD44v6 KD is associated with several changes in tumor matrix organization [23], and, therefore, we hypothesize that, upon CD44v6 KD, cells activate specific signaling pathways to compensate CD44v6 loss, leading to increased survival right after treatment. However, further studies would be necessary to confirm this hypothesis. Nevertheless, this effect seems transient, since, in the more biologically relevant long-term analysis, CD44v6 KD decreases cell survival, compared to CD44v6 expressing cells. Regarding exon-v6 depleted cells, two edited clones, MKN45_O26 and GP202_O25, present a similar behavior to that of CD44v6 KD in short-term cisplatin treatment. MKN45_O26 also presents the same behavior in response to 5-FU. Importantly, in long-term assays, GP202 edited cell lines respond similarly to the KD of CD44v6 isoforms in response to cisplatin. Therefore, we believe that specific exon-v6 depletion can sensitize GC cells to cisplatin, which may be dependent on the cellular context. Although all edited cell lines present specific exon-v6 skipping, they do not all show similar results in chemotherapy response. This might be caused by the different DNA breakpoints. For instance, MKN45_O15 still possesses half of the exon-v6 sequence while MKN45_O26 has complete deletion. Although both GP202 edited clones present complete exon-v6 deletion, GP202_O26 retains more 46 bp than GP202_O25 in the adjacent intron. Therefore, it is possible that certain regulatory elements are altered within the edited cell lines [24-26], which will ultimately result in differences in gene/protein expression.

Considering that CD44/CD44v overexpression has been described in CSCs [23, 27-31], a subpopulation of tumor cells believed to be responsible for the initiation/maintenance of tumor growth and metastization [32-34], we also investigated whether CD44v6 affects one of the most important CSC feature, their self-renewal capacity. Interestingly, we show that either complete CD44v6 KD or specific exon-v6 skipping from GC cells leads to the severe loss of self-renewal capacity. Many cancer patients experience tumor recurrence because chemotherapy often fails to eliminate CSCs, which are frequently more chemoresistant than the remaining tumor bulk [27]. Being a marker of stemness, CD44v6 is associated with CSC features, like therapy resistance and metastization [23, 27]. Therefore, we demonstrate that removing only exon-v6 results in CSC feature attenuation, by decreasing the self-renewal capacity and resistance to cisplatin of GC cells. In addition to v6 being relevant to drug resistance and cell self-renewal in gastric cancer cells, it is possible that other CD44v exons, such as v9, also play a role in these capabilities. For instance, CD44v9 has been described to confer drug resistance to gastric cancer cells [35, 36], and we have observed that cells that lack exon v9 in addition to exon v6 (both MKN45 and GP202 CD44v6 KD cells) consistently present increased sensitivity to cisplatin and 5-FU, in the long term cell survival (clonogenic) assay.

Therefore, in this context, it is possible that the v9 exon is also contributing (in addition to v6) to the drug resistance phenotype. To clarify this issue, which falls beyond the scope of the present study, similar experiments (to the ones we undertook here) could be performed, using CRISPR/Cas9 cells lacking specifically CD44v9.

Although we successfully used CRISPR/Cas9 to generate our study models, this strategy is far from being used as a therapeutic strategy, due to the technical and ethical issues that it raises [37, 38]. Therefore, we tested another strategy to replicate the CRISPR/Cas9 results (albeit in a transient manner). We used PMOs, a strategy that has been implemented in the clinics. Indeed, PMOs are the only therapeutic option for Duchenne muscular dystrophy patients, where skipping the mutated exon can restore the reading frame, restoring protein production [39]. Hence, CRISPR/Cas9 and PMOs are good techniques for splicing modulation, being the former the best to create stable disease study models and the later the best to apply into pre-clinical studies. We strongly believe that using a PMO approach in CD44v6 expressing tumors would lead to tumor sensitization to chemotherapy and delayed tumor progression. Nevertheless, more studies using this approach, particularly using in vivo models, would be important to support our findings.

CONCLUSIONS

In summary, we demonstrated that depletion of exon-v6 from GC cells can increase sensitivity to cisplatin and impair self-renewal, which could prove to be an extremely valuable approach to improve GC patient therapy and survival. Moreover, we are confident that a PMO approach triggering exon-v6 skipping may be a feasible therapeutic option to sensitize tumor cells and possibly condition the progression of CD44v6 positive gastric cancers.

Author Contributions

G.M.A. and C.O. conceived and supervised the study; G.M.A., C.O., C.P. and S.L. designed the study; S.L. and C.P. acquired the data; S.L., C.P., G.M.A. and C.O. analyzed and interpreted the data; S.L., C.P. and G.M.A. drafted the manuscript; G.M.A. and C.O. edited and reviewed the manuscript; C.O. and G.M.A. 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 was 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 and NORTE-01-0145-FEDER-000029- 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, funded by ERDF, POCI and FCT, and project PTDC/CTM-NAN/120958/2010, from FCT. Salary support to S.L. from an IPATIMUP internal project to C. Oliveira funded by the IPATIMUP Board of Directors; C.P. was supported by the grant SFRH/BD/113031/2015; G.M.A. was supported by the Investigator FCT Program 2013 (IF/00615/2013), POPH-QREN Type 4.2, European Social Fund and Portuguese Ministry of Science and Technology (MCTES).

Acknowledgments

The authors acknowledge the support of the i3S Scientific Platforms “Advanced Light Microscopy”, member of the PPBI (PPBI-POCI-01-0145-FEDER-022122), “Translational Cytometry Unit”, “GenCore” member of the National Laboratory for Genome Sequencing and Analysis-GenomePT (POCI-01-0145-FEDER-022184). The authors would like to thank Celso Reis and Diana Campos from Glycobiology in Cancer at i3S, for kindly lending CD44v9 antibody. S.L. acknowledges the support of the Institute for Biomedical Sciences Abel Salazar from the University of Porto, specifically by the Master’s Degree in Oncology and Carmen Jerónimo. C.P. acknowledges the support of the Faculty of Medicine from University of Porto, specifically by the Doctoral Programme in Biomedicine.

Conflicts of Interest

The authors declare no conflict of interest.

REFERENCES

1. Smyth, E.C.;Verheij, M.;Allum, W.;Cunningham, D.;Cervantes, A.;Arnold, D.;Committee, E.G. Gastric cancer: ESMO Clinical Practice Guidelines for diagnosis, treatment and follow-up. *Ann Oncol.* **2016**, 27, 38-49, DOI: 10.1093/annonc/mdw350
2. Goodison, S.;Urquidi, V.;Tarin, D. CD44 cell adhesion molecules. *Mol Pathol.* **1999**, 52, 189-196, DOI: 10.1136/mp.52.4.189
3. Orian-Rousseau, V. CD44, a therapeutic target for metastasising tumours. *Eur J Cancer.* **2010**, 46, 1271-1277, DOI: 10.1016/j.ejca.2010.02.024
4. Zoller, M. CD44, Hyaluronan, the Hematopoietic Stem Cell, and Leukemia-Initiating Cells. *Front Immunol.* **2015**, 6, 235, DOI: 10.3389/fimmu.2015.00235
5. Senbanjo, L.T.;Chellaiah, M.A. CD44: A Multifunctional Cell Surface Adhesion Receptor Is a Regulator of Progression and Metastasis of Cancer Cells. *Front Cell Dev Biol.* **2017**, 5, 18, DOI: 10.3389/fcell.2017.00018
6. Hasenauer, S.;Malinger, D.;Koschut, D.;Pace, G.;Matzke, A.;von Au, A.;Orian-Rousseau, V. Internalization of Met requires the co-receptor CD44v6 and its link to ERM proteins. *PLoS One.* **2013**, 8, 62357, DOI: 10.1371/journal.pone.0062357
7. Xie, J.W.;Chen, P.C.;Zheng, C.H.;Li, P.;Wang, J.B.;Lin, J.X.;Lu, J.;Chen, Q.Y.;Cao, L.L.;Lin, M., et al. Evaluation of the prognostic value and functional roles of CD44v6 in gastric cancer. *J Cancer Res Clin Oncol.* **2015**, 141, 1809-1817, DOI: 10.1007/s00432-015-1964-8

8. Xu, Y.Y.;Guo, M.;Yang, L.Q.;Zhou, F.;Yu, C.;Wang, A.;Pang, T.H.;Wu, H.Y.;Zou, X.P.;Zhang, W.J., et al. Regulation of CD44v6 expression in gastric carcinoma by the IL-6/STAT3 signaling pathway and its clinical significance. *Oncotarget.* **2017**, 8, 45848-45861, DOI: 10.18632/oncotarget.17435
9. Hu, B.;Luo, W.;Hu, R.T.;Zhou, Y.;Qin, S.Y.;Jiang, H.X. Meta-Analysis of Prognostic and Clinical Significance of CD44v6 in Esophageal Cancer. *Medicine (Baltimore).* **2015**, 94, 1238, DOI: 10.1097/MD.0000000000001238
10. Ma, L.;Dong, L.;Chang, P. CD44v6 engages in colorectal cancer progression. *Cell Death Dis.* **2019**, 10, 30, DOI: 10.1038/s41419-018-1265-7
11. Jiang, H.;Patel, P.H.;Kohlmaier, A.;Grenley, M.O.;McEwen, D.G.;Edgar, B.A. Cytokine/Jak/Stat signaling mediates regeneration and homeostasis in the Drosophila midgut. *Cell.* **2009**, 137, 1343-1355, DOI: 10.1016/j.cell.2009.05.014
12. Qiao, G.L.;Song, L.N.;Deng, Z.F.;Chen, Y.;Ma, L.J. Prognostic value of CD44v6 expression in breast cancer: a meta-analysis. *Onco Targets Ther.* **2018**, 11, 5451-5457, DOI: 10.2147/OTT.S156101
13. Matzke-Ogi, A.;Jannasch, K.;Shatirishvili, M.;Fuchs, B.;Chiblak, S.;Morton, J.;Tawk, B.;Lindner, T.;Sansom, O.;Alves, F., et al. Inhibition of Tumor Growth and Metastasis in Pancreatic Cancer Models by Interference With CD44v6 Signaling. *Gastroenterology.* **2016**, 150, 513-525, DOI: 10.1053/j.gastro.2015.10.020
14. Reeder, J.A.;Gotley, D.C.;Walsh, M.D.;Fawcett, J.;Antalis, T.M. Expression of antisense CD44 variant 6 inhibits colorectal tumor metastasis and tumor growth in a wound environment. *Cancer Res.* **1998**, 58, 3719-3726, DOI: Published August 1998
15. Stroomer, J.W.;Roos, J.C.;Sproll, M.;Quak, J.J.;Heider, K.H.;Wilhelm, B.J.;Castelijns, J.A.;Meyer, R.;Kwakkelstein, M.O.;Snow, G.B., et al. Safety and biodistribution of 99mTechnetium-labeled anti-CD44v6 monoclonal antibody BIWA 1 in head and neck cancer patients. *Clin Cancer Res.* **2000**, 6, 3046-3055, DOI: Published August 2000
16. Tijink, B.M.;Buter, J.;de Bree, R.;Giaccone, G.;Lang, M.S.;Staab, A.;Leemans, C.R.;van Dongen, G.A. A phase I dose escalation study with anti-CD44v6 bivatuzumab mertansine in patients with incurable squamous cell carcinoma of the head and neck or esophagus. *Clin Cancer Res.* **2006**, 12, 6064-6072, DOI: 10.1158/1078-0432.CCR-06-0910
17. da Cunha, C.B.;Oliveira, C.;Wen, X.;Gomes, B.;Sousa, S.;Suriano, G.;Grellier, M.;Huntsman, D.G.;Carneiro, F.;Granja, P.L., et al. De novo expression of CD44 variants in sporadic and hereditary gastric cancer. *Lab Invest.* **2010**, 90, 1604-1614, DOI: 10.1038/labinvest.2010.155
18. Pereira, C.;Ferreira, D.;Mendes, N.;Granja, P.L.;Almeida, G.M.;Oliveira, C. Expression of CD44v6-Containing Isoforms Influences Cisplatin Response in Gastric Cancer Cells. *Cancers (Basel).* **2020**, 12, 858, DOI: 10.3390/cancers12040858
19. Gartner, F.;David, L.;Seruca, R.;Machado, J.C.;Sobrinho-Simoes, M. Establishment and characterization of two cell lines derived from human diffuse gastric carcinomas xenografted in nude mice. *Virchows Arch.* **1996**, 428, 91-98, DOI: 10.1007/bf00193936

20. Ran, F.A.;Hsu, P.D.;Wright, J.;Agarwala, V.;Scott, D.A.;Zhang, F. Genome engineering using the CRISPR-Cas9 system. *Nat Protoc.* **2013**, 8, 2281-2308, DOI: 10.1038/nprot.2013.143
21. Morrisey Lab Protocol: Generating Large (>1kb) Genomic Deletions Using CRISPRs. Available online: https://www.med.upenn.edu/genetics/tcmf/protocols.shtml. (acessed on July 2, 2020)
22. Desmet, F.O.;Hamroun, D.;Lalande, M.;Collod-Béroud, G.;Claustres, M.;Béroud, C. Human Splicing Finder: an online bioinformatics tool to predict splicing signals. **2009**, 37, 67, DOI: 10.1093/nar/gkp215
23. Wang, Z.;Zhao, K.;Hackert, T.;Zoller, M. CD44/CD44v6 a Reliable Companion in Cancer-Initiating Cell Maintenance and Tumor Progression. *Front Cell Dev Biol.* **2018**, 6, 97, DOI: 10.3389/fcell.2018.00097
24. Riethoven, J.J. Regulatory regions in DNA: promoters, enhancers, silencers, and insulators. *Methods Mol Biol.* **2010**, 674, 33-42, DOI: 10.1007/978-1-60761-854-6_3
25. Perri, F.;Longo, F.;Giuliano, M.;Sabbatino, F.;Favia, G.;Ionna, F.;Addeo, R.;Della Vittoria Scarpati, G.;Di Lorenzo, G.;Pisconti, S. Epigenetic control of gene expression: Potential implications for cancer treatment. *Crit Rev Oncol Hematol.* **2017**, 111, 166-172, DOI: 10.1016/j.critrevonc.2017.01.020
26. DeAngelis, J.T.;Farrington, W.J.;Tollefsbol, T.O. An overview of epigenetic assays. *Mol Biotechnol.* **2008**, 38, 179-183, DOI: 10.1007/s12033-007-9010-y
27. Zavros, Y. Initiation and Maintenance of Gastric Cancer: A Focus on CD44 Variant Isoforms and Cancer Stem Cells. *Cell Mol Gastroenterol Hepatol.* **2017**, 4, 55-63, DOI: 10.1016/j.jcmgh.2017.03.003
28. Bourguignon, L.Y.;Shiina, M.;Li, J.J. Hyaluronan-CD44 interaction promotes oncogenic signaling, microRNA functions, chemoresistance, and radiation resistance in cancer stem cells leading to tumor progression. *Adv Cancer Res.* **2014**, 123, 255-275, DOI: 10.1016/B978-0-12-800092-2.00010-1
29. Ishiwata, T.;Matsuda, Y.;Yoshimura, H.;Sasaki, N.;Ishiwata, S.;Ishikawa, N.;Takubo, K.;Arai, T.;Aida, J. Pancreatic cancer stem cells: features and detection methods. *Pathol Oncol Res.* **2018**, 24, 797-805, DOI: 10.1007/s12253-018-0420-x
30. Kozovska, Z.;Gabrisova, V.;Kucerova, L. Colon cancer: cancer stem cells markers, drug resistance and treatment. *Biomed Pharmacother.* **2014**, 68, 911-916, DOI: 10.1016/j.biopha.2014.10.019
31. Yan, Y.;Zuo, X.;Wei, D. Concise Review: Emerging Role of CD44 in Cancer Stem Cells: A Promising Biomarker and Therapeutic Target. *Stem Cells Transl Med.* **2015**, 4, 1033-1043, DOI: 10.5966/sctm.2015-0048
32. Yu, Z.;Pestell, T.G.;Lisanti, M.P.;Pestell, R.G. Cancer stem cells. *Int J Biochem Cell Biol.* **2012**, 44, 2144-2151, DOI: 10.1016/j.biocel.2012.08.022
33. Nassar, D.;Blanpain, C. Cancer Stem Cells: Basic Concepts and Therapeutic Implications. *Annu Rev Pathol.* **2016**, 11, 47-76, DOI: 10.1146/annurev-pathol-012615-044438

34. Clarke, M.F. Clinical and Therapeutic Implications of Cancer Stem Cells. *N Engl J Med.* **2019**, 380, 2237-2245, DOI: 10.1056/NEJMra1804280
35. Miyoshi, S.;Tsugawa, H.;Matsuzaki, J.;Hirata, K.;Mori, H.;Saya, H.;Kanai, T.;Suzuki, H. Inhibiting xCT Improves 5-Fluorouracil Resistance of Gastric Cancer Induced by CD44 Variant 9 Expression. *Anticancer Res.* **2018**, 38, 6163-6170, DOI: 10.21873/anticanres.12969
36. Mashima, T.;Iwasaki, R.;Kawata, N.;Kawakami, R.;Kumagai, K.;Migita, T.;Sano, T.;Yamaguchi, K.;Seimiya, H. In silico chemical screening identifies epidermal growth factor receptor as a therapeutic target of drug-tolerant CD44v9-positive gastric cancer cells. *Br J Cancer.* **2019**, 121, 846-856, DOI: 10.1038/s41416-019-0600-9
37. Huang, J.;Wang, Y.;Zhao, J. CRISPR editing in biological and biomedical investigation. *J Cell Physiol.* **2018**, 233, 3875-3891, DOI: 10.1002/jcp.26141
38. Brokowski, C.;Adli, M. CRISPR Ethics: Moral Considerations for Applications of a Powerful Tool. *J Mol Biol.* **2019**, 431, 88-101, DOI: 10.1016/j.jmb.2018.05.044
39. Lim, K.R.;Maruyama, R.;Yokota, T. Eteplirsen in the treatment of Duchenne muscular dystrophy. *Drug Des Devel Ther.* **2017**, 11, 533-545, DOI: 10.2147/DDDT.S97635

APPENDIX

This appendix includes supplementary figures and tables.

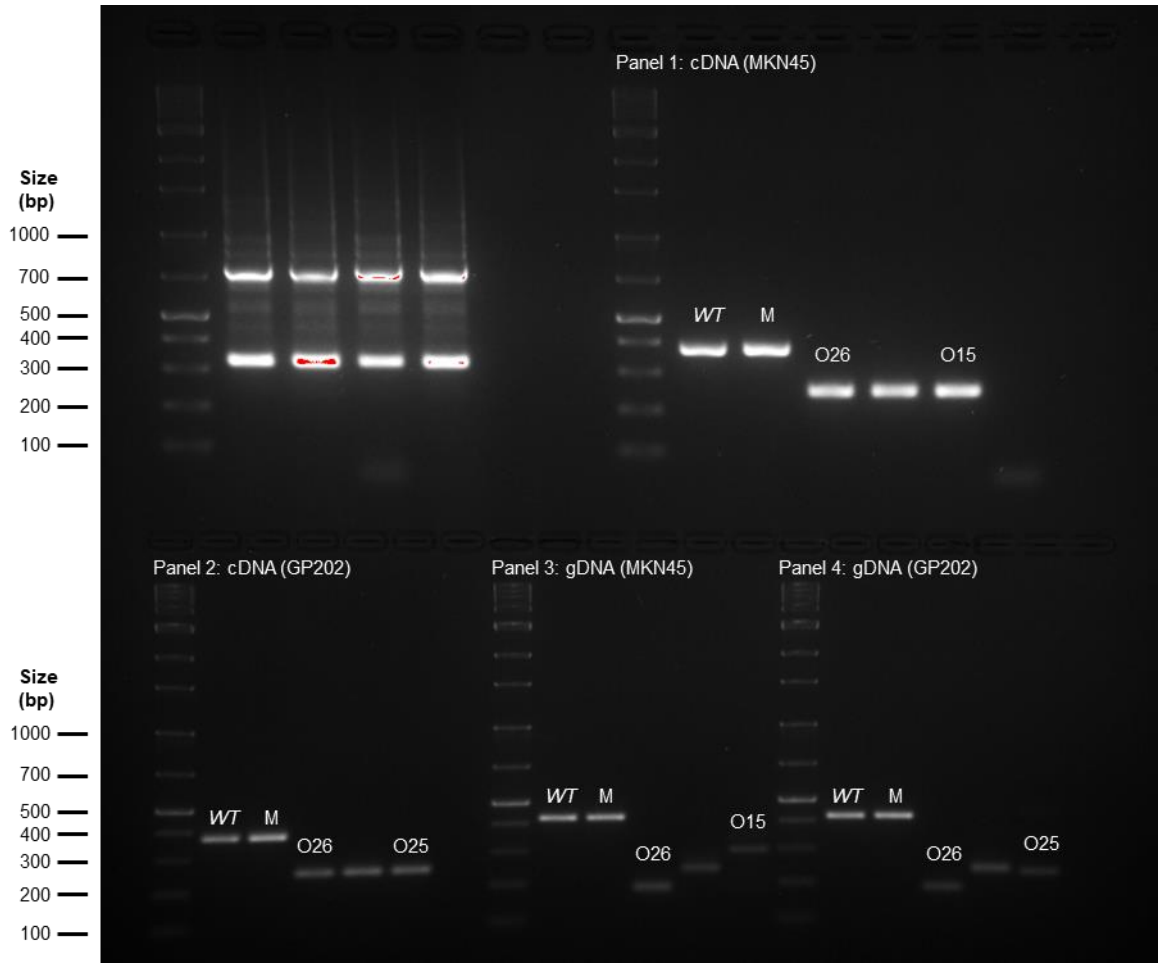


Figure S1. Uncropped image of the 2% agarose gel from Figure 2A.

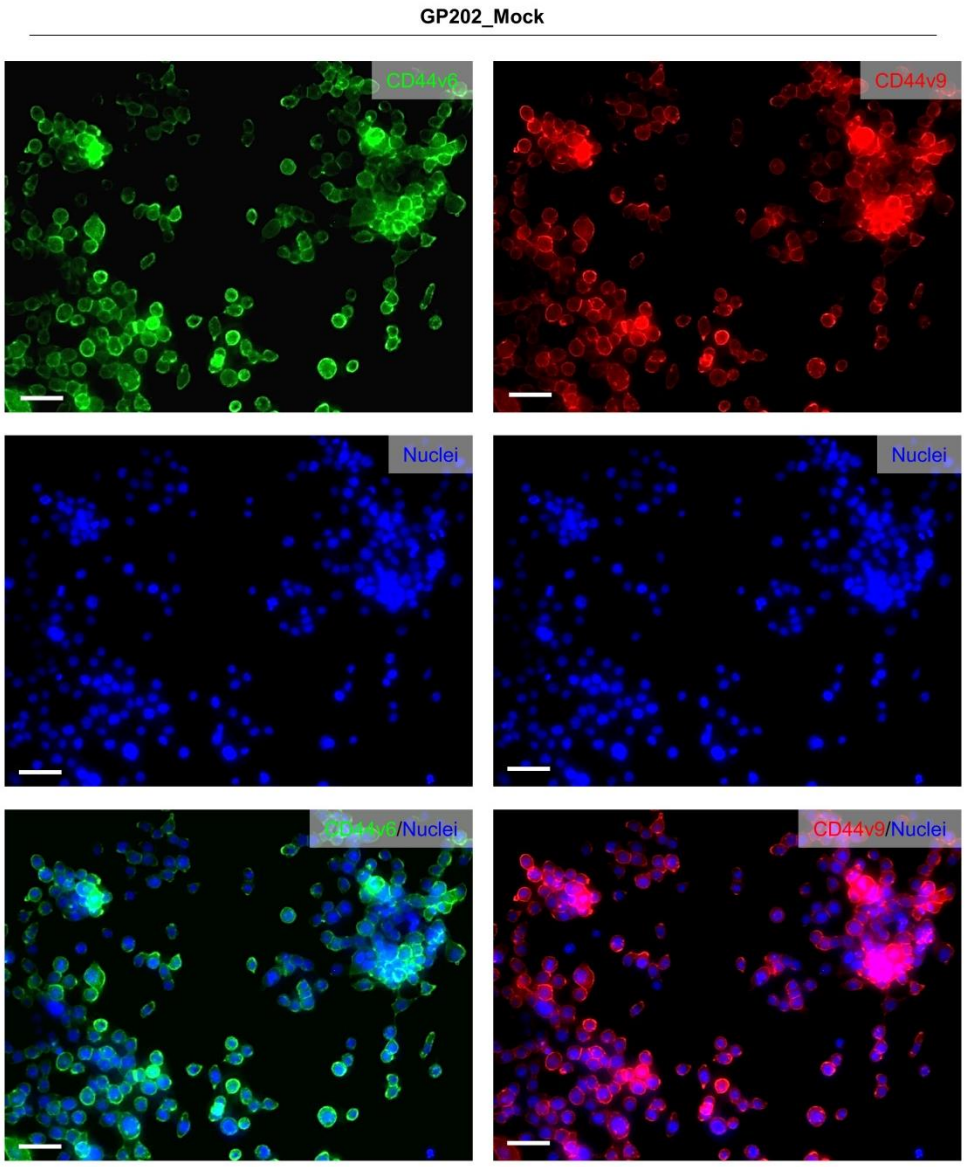


Figure S2. Illustrative immunofluorescence images highlighting CD44v6 (green) and CD44v9 (red) membranous staining observed in Figure 3C. Nuclei were stained with DAPI (blue) and white scale bars represent a distance of 50 µm.

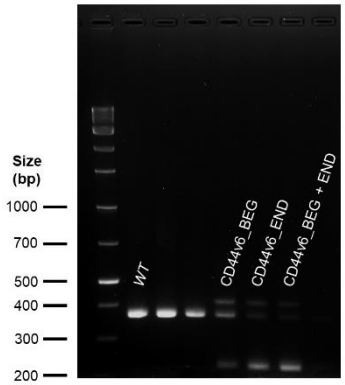


Figure S3. Uncropped image of the 2% agarose gel from Figure 6C.

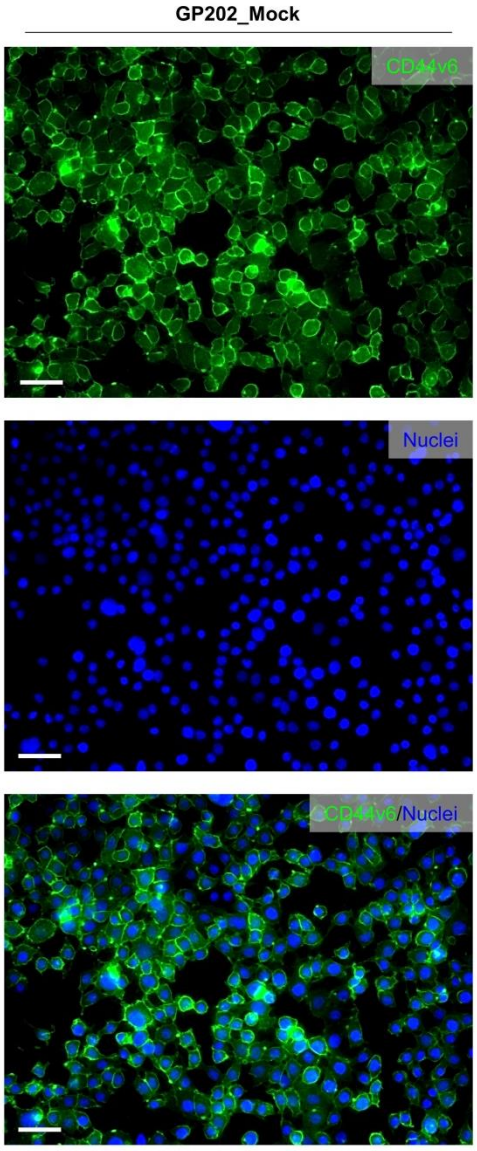
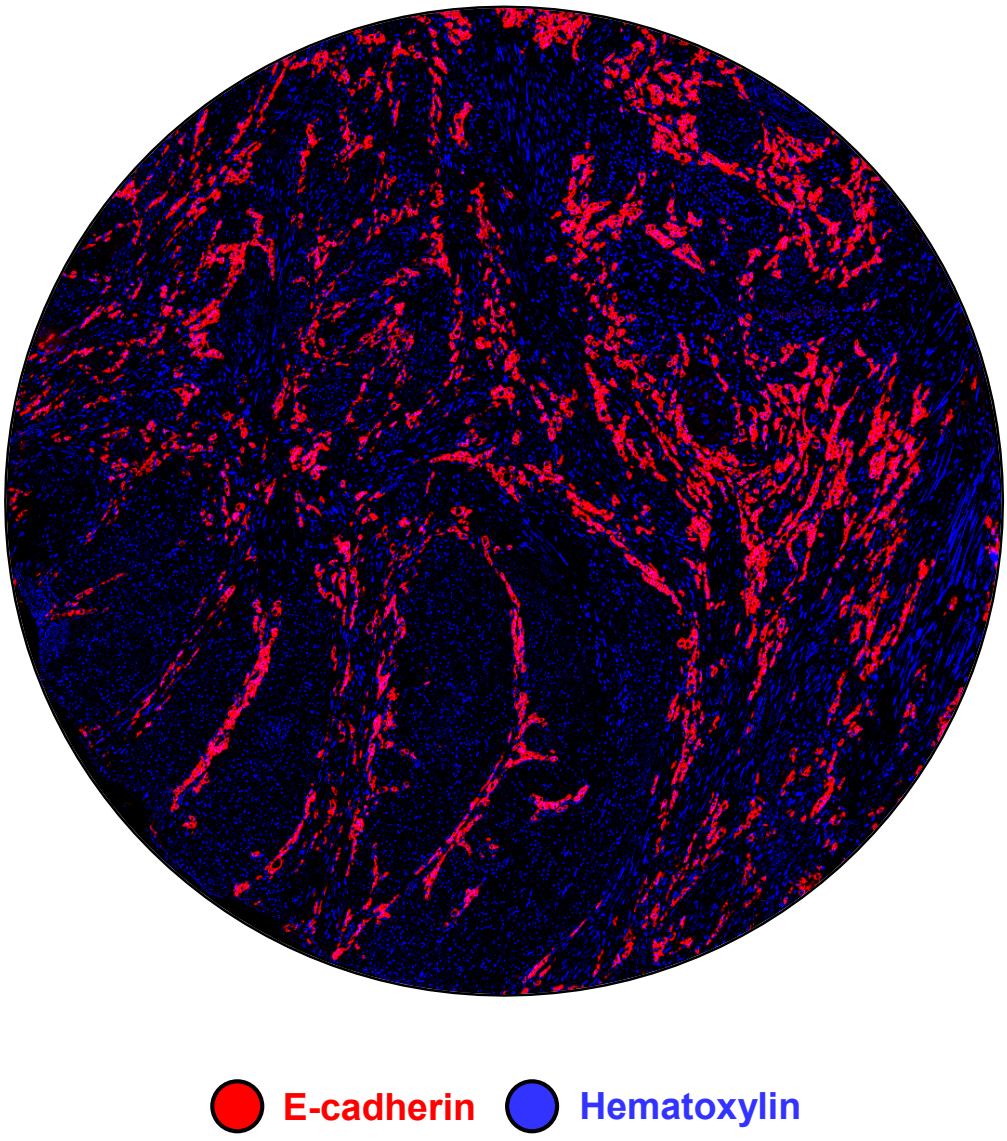


Figure S4. Illustrative immunofluorescence images highlighting CD44v6 (green) membranous staining observed in Figure 6E. Nuclei were stained with DAPI (blue) and white scale bars represent a distance of 50 µm.

Chapter 5

Surrogate Markers for E-cadherin loss-of-function (LoF)



Original Paper 5

The cast of characters behind E-cadherin dysfunction in gastric cancer

Carla Pereira, Marta Ferreira, Diana Martins, Patricia Oliveira, Hugo Osório, Fátima Carneiro, Gabriela M. Almeida, Carla Oliveira

Manuscript in preparation

ABSTRACT

E-cadherin/*CDH1* dysfunction, leading to absent or aberrant protein expression, is the most well-established event in gastric cancer (GC) initiation and progression, independently of the histological type. Although aberrant E-cadherin expression has been correlated with poor outcome in GC, there are conflicting reports about the usefulness of E-cadherin expression as an independent prognostic marker. Hence, we aimed at identifying overexpressed proteins as surrogate markers for E-cadherin LoF in both diffuse- and intestinal-type setting and explored its clinical potential.

To address this, permanent (CRISPR/Cas9) and transient (RNAi) E-cadherin depleted cell models representative of DGC and IGC were established and characterized by label-free quantitative proteomics profiling followed by bioinformatics analysis. Proteomics profiling of E-cadherin depleted cell models revealed that E-cadherin permanent loss in DGC generates an acute phenotype of cell-cell (adherens, tight junctions and desmosomes) and cell-matrix adhesion loss. In opposition, transient E-cadherin LoF in IGC triggers a less prominent cell-cell adhesion loss phenotype due to *de novo* expression of alternative cadherins, namely P-cadherin. We also identified E-cadherin LoF associated markers, specific for DGC and IGC, that were validated by ICC and will be validated in the clinical series of patients.

Our data supports the idea that early and permanent E-cadherin LoF, characteristic of DGC, irreversibly affects cell-cell and matrix adhesion, and likely explains the local dissemination, while the late and transient E-cadherin LoF associated to the cadherin-switch observed in IGC, likely explains dissemination to distant organs. Further, the identification of E-cadherin LoF associated markers, if validated will be fundamental for detection and management in GC patients.

Cadherin switch; CRISPR/Cas9; P-cadherin; Proteomics analysis; siRNA; Surrogate markers

INTRODUCTION

E-cadherin dysfunction, leading to absent or aberrant protein expression, is the most well-established event in GC initiation and progression in both histological types [1]. While E-cadherin permanent loss is the trigger for diffuse-type GC (DGC), aberrant E-cadherin expression is common along the progression of intestinal-type GC (IGC).

IGC evolves in a step-wise manner through the sequential development of histopathological changes (atrophic gastritis, intestinal metaplasia, dysplasia and eventually carcinoma) accumulating a series of genetic and epigenetic alterations together with environmental factors [2]. Along its progression, one somatic hit often triggers the loss of one allele of E-cadherin coding gene (CDH1). It is believed that, for self-benefit, IGC cells are able of transiently modulate E-cadherin expression, migrate to distant sites and reinstall E-cadherin in those organs [3]. The initiation and scattered growth of DGC is often associated with complete loss of E-cadherin function, which limits its migration and invasion capacity for distant sites. Therefore, DGC often evolve through the serosa triggering local dissemination to near lymph nodes and the peritoneum adjacent to the stomach [4]. Remarkably, DGC has poorer prognosis than IGC, likely because it spreads to the peritoneum, while IGC metastasizes to distant organs [5, 6].

Aberrant E-cadherin expression has been extensively correlated with poor outcome in GC [7-9], presenting a strong correlation with tumor grade, depth of invasion and regional lymph node involvement [10, 11]. However, some studies do not support this evidence [12], bringing into question the usefulness of E-cadherin expression as an independent prognostic marker. One of the factors contributing to this conflicting results may arise from the difficulty on quantifying E-cadherin imunohistochemical expression, which can assume a variety of patterns (null, dotted, cytoplasmic, incomplete membrane, complete membrane) [13], thus obscuring the analytical path and contributing to the inconsistencies that lead to erroneous prognostic indications.

Furthermore, the optimal thresholds defining the extent of E-cadherin expression is arbitrary, which might produce heterogeneity. Previously, we developed an image analysis pipeline to automatically quantify E-cadherin protein expression in GC tumors. Indeed, our pipeline performed well quantifying absent E-cadherin protein expression, but was not able to accurately quantify aberrant E-cadherin expression, frequently puzzling functional and dysfunctional patterns [14].

Given the challenge in the analysis of defective patterns of E-cadherin protein expression in GC clinical samples, both by pathologists and computational tools, we aimed to identify surrogate markers for E-cadherin LoF with potential clinical utility.

MATERIALS AND METHODS

All reagents from ThermoFisher Scientific unless otherwise stated.

1. Cell lines and culture conditions

Human GC cell lines MKN-74 and MKN45 were purchased from the JCRB Cell Bank (Japanese Collection of Research Bioresources Cell Bank. Briefly, both cell lines were cultured in RPMI 1640 medium with 10% heat inactivated foetal bovine serum (FBS) (Biowest, Nuaillé, France). Cell lines were maintained at 37 ºC and 5% CO2 in a high humidity atmosphere and sub-cultured every 3 to 4 days. Cells were never continuously cultured for more than 4 months. Cell identification was confirmed by STR analysis and cells were confirmed to be free of mycoplasma contamination.

2. Generation of permanent E-cadherin LoF models by CRISPR/Cas9

The establishment of E-cadherin LoF models by CRISPR-Cas9 technology was conducted through a step-wise process: i) *in silico* design of sgRNAs; ii) construction of an expression plasmid including sgRNA and Cas9 nuclease; iii) plasmid transfection and clonal selection; iv) functional genotyping of genomic deletion and v) clonal expansion.

To impair E-cadherin function, single guided RNA (sgRNA) were designed to target exon 3. In CDH1 locus, exon 3 is in frame, which means that the triplets of nucleotides do not finish or start evenly at the boundary of exon 3 - 4. Hence, insertions or deletions will alter the reading frame and translate different amino acids, originating a nonfunctional protein with a premature stop codon. For exon 3, we introduced a 250bp portion of 5´ and 3´ ends into Zheng lab guide RNAs design tool at http://CRISPR.mit.edu. This tool returns several 21bp targeted sequences located near a 3bp PAM sequence (NGG) for each DNA strand, ordered by the lowest off target effects. The top three target sequences were chosen for each 5' and 3´ ends. Six sgRNAs were purchased from Sigma-Aldrich (Poole, UK) and individually cloned into pSpCas9(BB)-2A-Puro (PX459) V2.0 plasmid (Addgene plasmid #62988, Watertown, MA, USA), in the Bbs I restriction site, according to the "Morrisey Lab Protocol" [15], with few adaptations. Each Sg plasmid was transformed into *Escherichia Coli* one shot Match1TM-T1R (Invitrogen, Carlsbad, CA, USA), according the manufacturer's instructions. A colony PCR was performed to screen the colonies and the positive colonies were amplified and sequenced using the primers A and C from Table S1. The sgRNA were transfected in pairs to MKN74 and MKN45 GC cell lines. Briefly, cells were plated in 12 well plates and allowed to grow for 24 h, until approximately 70% confluency. Afterwards, Sgs' plasmid was complexed with 1.75 µL of Lipofectamine 3000 transfection reagent, according to the manufacturer's instructions. After 48 h, puromycin (Merck, Darmstadt,

Germany) was added to the cells to select vector positive cells. Puromycin was renewed every 72 h, until the non-transfected cells, which do not have the antibiotic selection gene to puromycin carried by the plasmid, were dead.

3. Genotyping Analysis of CRISPR/Cas9 Knockout Models

gDNA was extracted using NZY Tissue gDNA Isolation kit (NZYTech, Lisbon, Portugal) according to the manufacturers' protocol. NanoDrop® ND-1000 UV-Vis Spectrophotometer (ThermoFisher Scientific, Waltham, MA, USA) was used to determine gDNA quality and concentration. gDNA was amplified with multiplex PCR kit (Qiagen, Venlo, The Netherlands) and primers designed to flank exon 3 at *CDH1* locus (primers B/D, Table S1). PCR products were analyzed through 2% agarose gel electrophoresis and sanger sequencing.

4. E-cadherin expression inhibition by siRNA

MKN74 and MKN45 cells were transfected with siRNAs for E-cadherin using Lipofectamine® RNAiMax Transfection Reagent (ThermoFisher Scientific, Waltham, MA, USA), according to the manufacturers' instructions. Briefly, cells were seeded in 6 well-plates ($1x10^5$ cells/well and 1.5x105 cells/well, respectively MKN74 and MKN45 cell lines). After 24 h, lipid based conjugates were prepared by mixing 10 nM of Non targeting siRNA (negative control DS NC1; iDT, Leuven, Belgium), or human CDH1 siRNA (Sense strand: 5'- GCGUCAGGUUCCAUAGGAAUCCUTT - 3' and Antisense strand: 5'- AAAGGAUUCCUAUGGAACCUGACGCAG - 3', custom made from iDT, Leuven, Belgium), to diluted Lipofectamine® RNAiMax Reagent in 1:1 ratio. The conjugates were incubated for 5 min at room temperature and added dropwise to the cells. Upon 48 h of incubation, protein extraction and Western blots were performed to evaluate the efficacy of E-cadherin silencing. In addition, protein lysates and RNA samples were collected and label-free proteomics analysis and RNA-sequencing was performed (as described above).

5. Immunofluorescence analysis

All immunofluorescence reagents were from ThermoFisher Scientific (Waltham, MA, USA), and all antibodies were from Cell Signaling Technology (Beverly, MA, USA) unless stated otherwise.

Cells were grown on glass cover slips for the desired amount of time under normal growth conditions or with specific treatments before being washed twice in PBS and fixed with 4% paraformaldehyde (Merck, Darmstadt, Germany) for 20 min. Post fixation, cover slips were washed with PBS (pH 7.4), blocked with 5% BSA solution for 30 min, and subsequently incubated with the desired primary antibodies.

Primary antibodies, diluted in 5% BSA were: i) mouse monoclonal antibodies – anti-β-catenin (1:50; ON; 4°C); anti-E-cadherin antibody (HECD1) (ab1416; 1:100; ON; 4°C Abcam); ii) rabbit

monoclonal antibodies – anti-E-cadherin antibody (24E10; 1:100; ON; 4°C); anti-P-cadherin (#2130; 1:50; ON; 4°C); PEG10 (1:100; ON, 4°C, Sigma Aldrich). Cells were then washed twice with PBS and incubated in the dark with secondary antibodies: Alexa Fluor 488/Alexa Fluor 594 Donkey Anti-Mouse secondary antibody (1:500; 60 min; Life Technologies) or Alexa Fluor 488/Alexa Fluor 594 anti-rabbit (1:500, 60 min; Life Technologies), after which a final PBS wash was performed. In all cases, the cell nuclei were stained with DAPI (1µg/mL in PBS; 5 min incubation; Sigma). All coverslips were mounted using Vectashield mounting media (Vector Laboratories, Burlingame, CA, USA) and cells were analyzed by fluorescence microscopy (Imager.Z1, AxioCam fluorescence microscope or Eclipse TE-2000, both from Zeiss, Gottingen, Germany) using AxioVision software (Rockville, MD, USA).

6. Protein Extraction and Western Blot

Cultured cells were washed twice with PBS and lysed with cold catenin lysis buffer [1% (v/v) Triton X-100, 1% (v/v) IGEPAL in PBS], supplemented with phosphatase (Sigma-Aldrich, Poole, UK) and protease (Roche, Basel, Switzerland) inhibitor cocktails. Total protein content of the samples was quantified by a modified Bradford Assay (Bio-Rad, Hercules, CA, USA). Thirty-five µg of total protein lysate were resolved by electrophoresis in a 7.5% SDS polyacrylamide gel. The protein was transferred onto a Hybond nitrocellulose membrane (GE Healthcare, Chicago, IL, USA) and stained with Ponceau S solution (Sigma-Aldrich, Poole, UK) to confirm the efficacy of the transfer. Membrane blocking was performed by incubating membranes for 30 min with 5% non-fat milk diluted in 0.5% (v/v) PBS-Tween-20. Membranes were incubated overnight with the for anti-rabbit E-cadherin monoclonal primary antibody (clone 24A10; 1:1000; 4 °C) from Cell Signaling Technology (Beverly, MA, USA). α-tubulin (1:10,000; Sigma-Aldrich, Poole, UK) was used as loading controls. Membranes were further incubated with horseradish peroxidase-conjugated secondary antibodies (GE Healthcare and Santa Cruz) and labelled protein specific signal was detected by ECL (GE Healthcare, Chicago, IL, USA). Bands were quantified with the Quantity One software (Bio-Rad, Hercules, CA, USA). Protein for western blotting was extracted from at least three independent biological replicates.

7. *Label free proteomic quantification*

Proteins were processed for proteomics analysis following the solid-phase-enhanced sample-preparation (SP3) protocol as described in Krijgsveld J. *et al.* [16]. Proteins were extracted and solubilized with 100 mM Tris pH 8.5, 1% sodium deoxycholate, 10 mM tris (2-carboxyethyl) phosphine (TCEP) and 40 mM chloroacetamide for 10 minutes at 95ºC at 1000 rpm (Thermomixer, Eppendorf). Enzymatic digestion was performed with Trypsin/LysC (2 micrograms) overnight at 37ºC at 1000 rpm. The resulting peptides were cleaned-up and desalted with C18 micro columns and further quantified.

8. Protein Identification and Quantification

Raw data were loaded on Proteome Discoverer 2.1 software for identification and/or quantification of peptides and proteins. Identification was performed in the UniProt/ SwissProt human database using Mascot (Version 2.5.1, Matrix Sciences, London, UK). Carbamidomethylation of cysteines, amino terminus, and lysine were set as fixed modifications and methionine oxidation as variable modifications. Trypsin was selected as the enzyme. Only high confidence master proteins with at least two unique peptides were considered.

9. E-cadherin, P-cadherin, PEG10 and CK8/18 immunohistochemistry (IHC) of GC samples

Immunostaining targeted for E-cadherin (clone 24E10, 1:350 dilution for 32 min; Cell Signaling, Beverly, MA, USA), P-cadherin (clone 56, 1:100 dilution for 60 min; BD Transduction Biosciences, Lexington, KY, USA), PEG10 (1:100 dilution for 60 min; Sigma Aldrich, Poole, UK) and CK8/18 (Ready to use, 16 min, Roche Tissue Diagnostics) was performed in consecutive tissue sections, distanced a few micrometers apart (3-5µm). The staining was carried out on the automated Ventana BenchMark ULTRAStaining System, using OptiView DAB IHC Detection Kit (Roche/Ventana Medical Systems, Tucson, AZ, USA), according to the manufacturers' instructions. H staining was used to reveal all tissue. Positive staining controls were placed next to the paraffin sections. Normal gastric mucosa, skin, placenta and appendix were used as positive controls for E-cadherin, P-cadherin, PEG10 and CK8/18 expression, respectively.

10. Statistical Analysis

Statistical analysis was performed using GraphPad Prism version 7.00 software (GraphPad Software Inc., San Diego, CA, USA). A normality test was performed using a Shapiro-Wilk test. Two-way ANOVA with Tukey's post hoc test for multiple comparison analysis was used, with 95% confidence interval. Significant differences were considered significant when $p < 0.05$.

RESULTS AND DISCUSSION

In order to identify surrogate markers for E-cadherin LoF, we established permanent (CRISPR/Cas9) and transient (RNAi) E-cadherin deficient cell models, representative of DGC and IGC, respectively.

1. Establishment of Intestinal and Diffuse-Like In Vitro Cell Models

To achieve the stable *CDH1* knockout, we used MKN45 GC cell lines, a poorly differentiated adenocarcinoma cell line [17]. We designed a strategy to target the adjacent areas of exon 3 in

CDH1 locus (Figure 1A) by designing single guide RNAs (sgRNAs) (g1 and g2) in the vicinity of this exon (Figure 1B). After genome editing, we obtained one edited cell line with a deletion of 201 bp (Figure 1C and Table S1 (Primers B and D)), also confirmed by Sanger sequencing (Figure 1D). By immunofluorescence analysis and western blot, we verified that our editing strategy was efficient, as no E-cadherin expression was detected in edited cells (Figure 1E and F).

In the intestinal setting, we used MKN-74, a well-differentiated GC cell line [17], and modeled the expression of E-cadherin by siRNA. Specific siRNAs were used to inhibit E-cadherin expression. As observed by immunofluorescence analysis, E-cadherin expression was absent in MKN-74 depleted cells (Figure 2A). By western blot, we confirmed the expression inhibition levels that, were, on average, ~85%, in three independent experiments (Figure 2B).

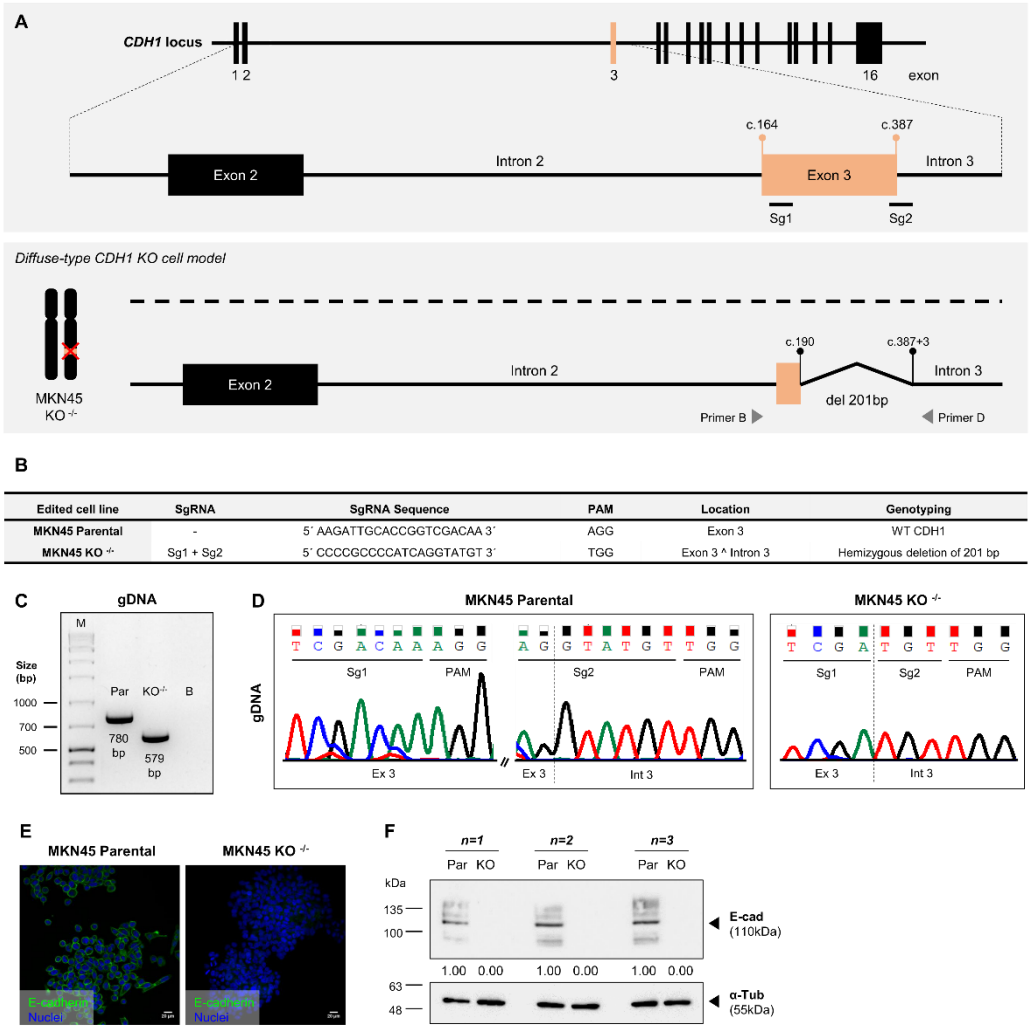


Figure 1. Genotyping analysis and characterization of E-cadherin knockout model in the diffuse-like setting. **(A)** Human CDH1 locus and CRISPR/Cas9 single guide RNA (sgRNA) design strategy. **(B)** Specification of the MKN45 edited cell line by CRISPR/Cas9 and the combination of sgRNAs transfected to create each edited cell line. **(C)** Analysis of gDNA from parental (1) and edited GC cell line (2). MKN45 KO$^{-/-}$ edited cell line presents a transcript with smaller size (579 bp). **(D)** Representative examples of sequencing analysis of MKN45 parental cells and MKN45 KO$^{-/-}$ cells. **(E)** Immunofluorescence analysis for E-cadherin (green)

expression in MKN45 parental and KO$^{-/-}$ cells. Nuclei were stained with DAPI (blue) and white scale bars represent a distance of 20 µm. **(F)** Western Blot depicting the E-cadherin protein levels in MKN45 parental and KO$^{-/-}$ cells in three replicates. E-cadherin and α-Tubulin were run in the same gel. (M) Marker, (B) Blank.

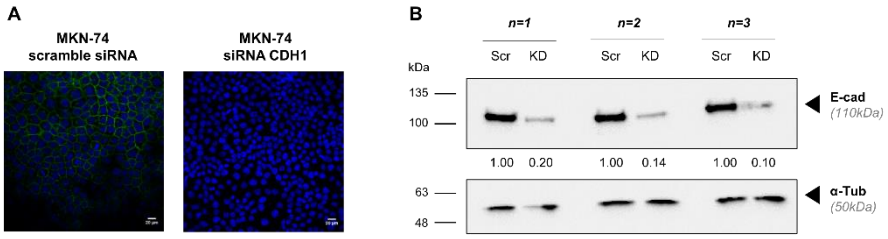


Figure 2. Characterization of E-cadherin knockdown model in the intestinal-like setting. **(A)** Immunofluorescence analysis of E-cadherin expression (green) in MKN-74 cell line upon incubation with *CDH1* specific siRNAs, compared with incubation with scramble siRNA. Nuclei were stained with DAPI (blue) and white scale bars represent a distance of 20 µm. **(B)** Western blot depicting the E-cadherin expression levels. E-cadherin and α-Tubulin were run in the same gel.

2. E-cadherin Is a Master Regulator of α- and β-catenin Protein Expression

We performed an integrative analysis of their proteomic content of E-cadherin deficient cell models. In average, 4800 proteins were detected in both GC cell lines models, from which roughly 100 were differentially expressed when compared to the control (scramble or parental cell line) (Figure S1B).

To start, we analyzed the pool of differentially expressed genes (DEGs) mutually altered in both models, transient intestinal model *versus* diffuse permanent model (Figure 3A). In both models, E-cadherin and its α- and β-catenin partners (encoded by CTNNA1 and CTNNB1 genes, respectively) were downregulated in all models, while p120 (encoded by CTNND1) remained unaltered (Figure 3B). Cadherin-catenin complex constitutes the core of a specialized type of adhesion junction named adherens junctions [18]. The classical E-cadherin is constituted by five extracellular domains, a single transmembrane domain and a cytoplasmic domain with highly conserved sites for the binding of armadillo proteins p120 and β-catenin [19]. While p120 associates directly with the membrane proximal region of the cadherin cytoplasmic tail, β-catenin binds to the so-called catenin-binding domain (CBD) serving as scaffold to anchor α-catenin that subsequently binds, directly or indirectly, to actin filaments. Overall, our results showed that E-cadherin expression is a master regulator of α- and β-catenin, but not of p120.

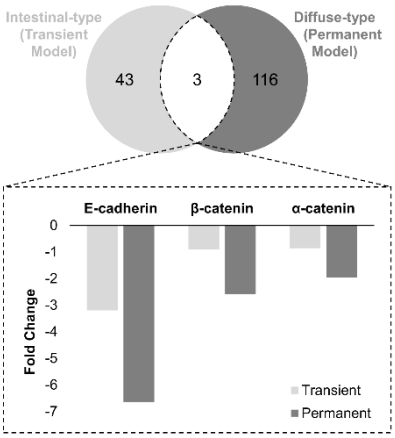


Figure 3. (A) Venn diagram depicting the pool of proteins differentially expressed in IGC transient model *versus* DGC permanent model. This diagram was generated using InteractiVenn online tool. **(B)** Protein expression levels of the three genes mutually altered in transient IGC and permanent DGC cell models.

3. Permanent E-cadherin LoF Triggers PEG10 De Novo Expression in the Diffuse-Like Setting

Afterwards, we analyzed the pool of genes specifically altered in the diffuse-like setting (Table S2 and Figure S2). In general, E-cadherin knockout (MKN45 parental *versus* MKN45 KO$^{-/-}$) led to the alteration 116 protein coding-genes (n=80 downregulated, n=38 upregulated, Figure 4A). Moreover, permanent E-cadherin LoF generated an acute behavior of cell-cell (adherens, tight junctions and desmosomes) and cell-matrix adhesion loss (Figure 4B).

Given that our aim was to identify surrogate markers of E-cadherin LoF, we next explored exclusively the pool of overexpressed proteins. As depicted in Figure 4A, permament E-cadherin LoF led to *de novo* expression of three protein-coding genes (KLK6, ADAM28 and PEG10), that are not expressed in the MKN45 parental cells. Based on literature research, we got particularly interested in PEG10 (Paternally Expressed Gene 10), which has been extensively implicated in the proliferation, apoptosis and metastasis of tumors. PEG10 is strongly expressed in placenta, ovary, testis and brain, but presents low expression levels in other normal tissues, such as stomach. In the cancer context, PEG10 was found to be overexpressed in a variety of cancers, including GC [20]. In this sense, we confirmed the expression of PEG10 in MKN45 parental and KO$^{-/-}$ *in vitro* model (Figure 4C) and performed a small pilot study with 40 GC tumors from diffuse- and mixed-type GC to analyze the expression of PEG10. Most of the cases did not expressed PEG10, although it was possible to observe PEG10 pontual staining in some cells (Figure 4D). However, few cases presented a massive PEG10 expression in regions of muscular invasion (Figure 4D, bottom panel). We confirmed by CK8/18 staining that these were not indeed tumor cells. Therefore, PEG10 has not yet been confirmed as a surrogate marker of E-cadherin LoF using patients tumours.

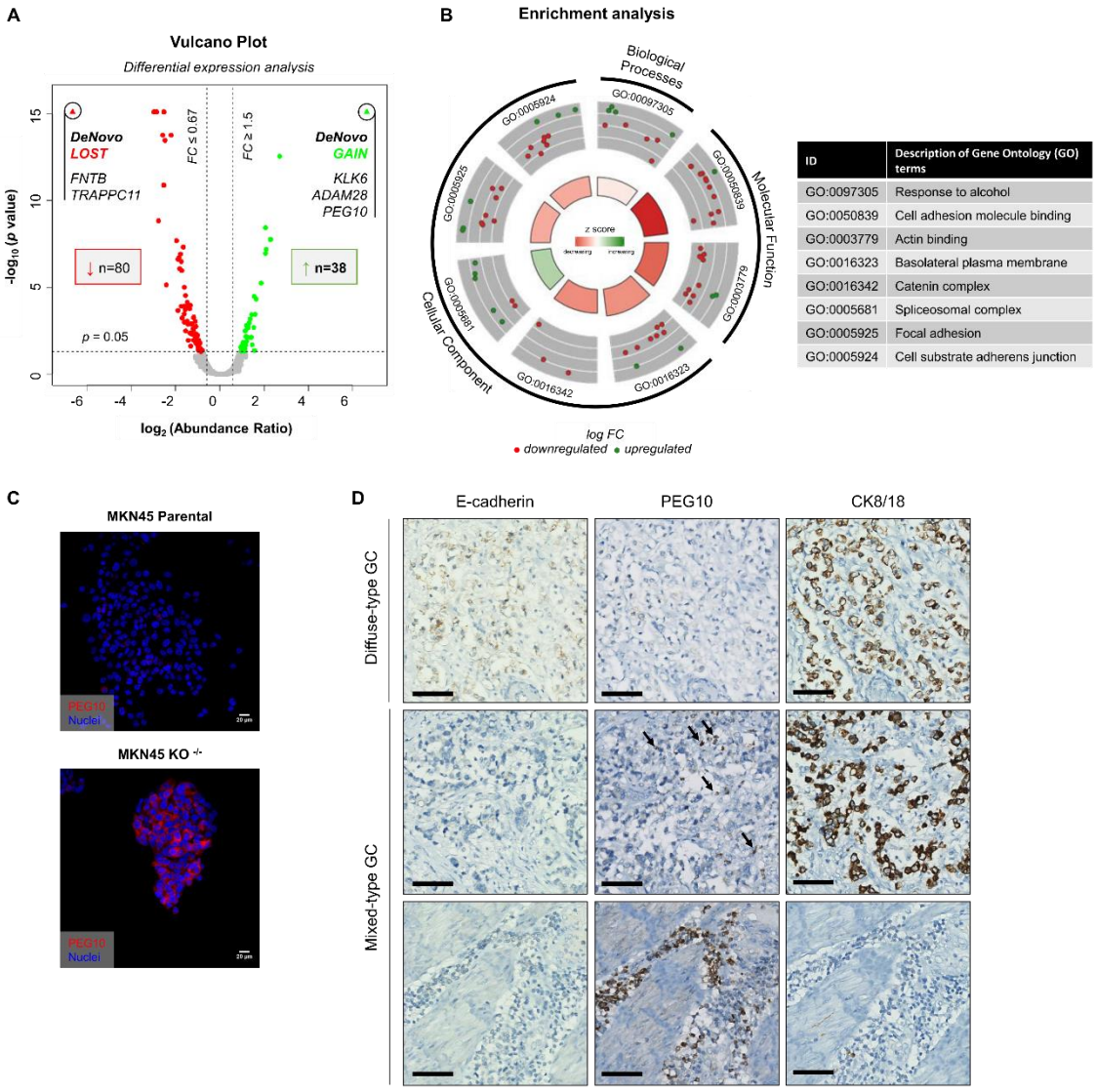
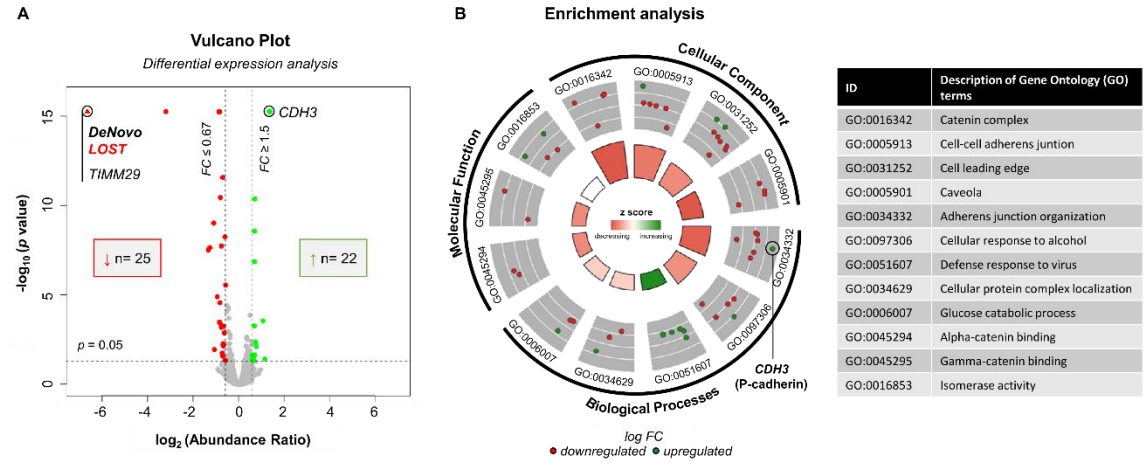


Figure 4. (A) Vulcano plot illustrating the differentially expressed proteins in permanent DGC model. Proteins with fold change (FC) greater than 1.5 or less than 0.67 and $p<0.05$ were considered up- (green) and down-regulated (red), respectively. **(B)** Circular plot illustrating the enriched gene ontology (GO) terms regarding biological processes, molecular function and cellular components of up-regulated and downregulated proteins. **(C)** Immunofluorescence analysis of PEG10 (red) expression in MKN-45 GC cells with preserved and deficient E-cadherin expression. Nuclei were stained with DAPI (blue) and white scale bars represent a distance of 20 µm. **(D)** Preliminary immunohistochemistry analysis of PEG10, E-cadherin and CK8/18 expression in consecutive tumors sections of GC tumors from diffuse and mixed histotype. Black arrows spot PEG10 staining, and black bars represent a distance of 70 µm.

4. Transient E-cadherin LoF Triggers Cadherin Switch from E- to P-cadherin in the Intestinal-like Setting

In the intestinal-like setting, a total of 47 protein-coding genes were found differentially expressed (n=25 downregulated; n=22 upregulated) upon E-cadherin knockdown (Table S3 and Figure S3). In addition, transient E-cadherin LoF triggers a less prominent cell-cell adhesion loss phenotype (Figure 5B), likely due to expression of alternative cadherins, namely P-cadherin. P-cadherin overexpression is linked to significant tumour promoting effects in many cancer types, including GC [21]. Its expression occurs in numerous epithelial tumour types and is often associated with more invasive lesions [22, 23]. By immunofluorescence, we confirmed that P-cadherin is overexpressed in MKN74 siRNA CDH1 and co-localizes with the remaining β-catenin that is present in E-cadherin deficient cells (Figure 5C). We performed a small pilot study to validate the P-cadherin expression in GC tumors from intestinal-type. We observed that regions with aberrant E-cadherin expression presented P-cadherin expression (Figure 5D). We will further validate the clinical significance of P-cadherin as a surrogate marker of E-cadherin LoF in a large cohort of GC patients. Recently, Kim M. and colleagues reported that P-cadherin expression was associated with Laurén intestinal histotype and patients whose tumors presented P-cadherin positive expression had a tendency towards a favorable outcome [23].

For the best of our knowledge, this is the first report showing that E-cadherin LoF triggers P-cadherin overexpression. In the future, it is important to explore E-cadherin and P-cadherin cooperation in GC and assess whether the concomitant expression of both molecules impact tumor aggressiveness and GC patients' outcome.



ID	Description of Gene Ontology (GO) terms
GO:0016342	Catenin complex
GO:0005913	Cell-cell adherens juntion
GO:0031252	Cell leading edge
GO:0005901	Caveola
GO:0034332	Adherens junction organization
GO:0097306	Cellular response to alcohol
GO:0051607	Defense response to virus
GO:0034629	Cellular protein complex localization
GO:0006007	Glucose catabolic process
GO:0045294	Alpha-catenin binding
GO:0045295	Gamma-catenin binding
GO:0016853	Isomerase activity

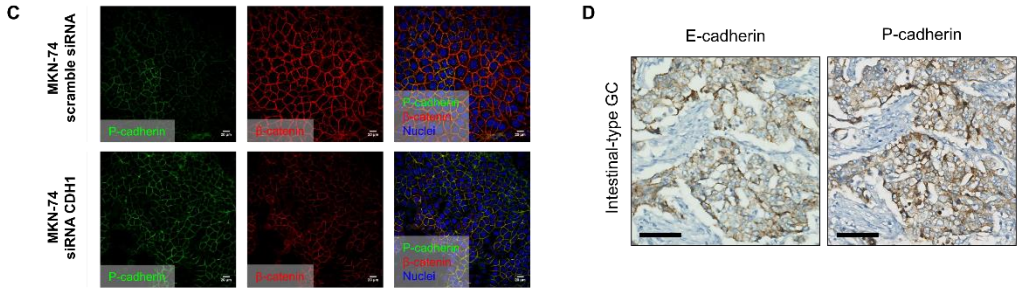


Figure 5. (A) Vulcano plot illustrating the differentially expressed proteins in the transient intestinal-like cell model. Proteins with fold change (FC) greater than 1.5 or less than 0.67 and $p<0.05$ were considered up- (green) and down-regulated (red), respectively. **(B)** Circular plot illustrating the enriched gene ontology terms related with the biological processes, molecular function and cellular components of up-regulated and downregulated proteins. **(C)** Immunofluorescence expression analysis of P-cadherin (green) and β-catenin (red) in MKN-74 GC cells with preserved and depleted E-cadherin expression. Nuclei were stained with DAPI (blue) and white scale bars represent a distance of 20 µm. **(D)** Preliminary immunohistochemistry analysis of P-cadherin and E-cadherin expression in consecutive tumors sections of GC tumors from intestinal-type. Black bars represent a distance of 70 µm.

CONCLUSION

Our data supports the idea that permanent E-cadherin LoF, characteristic of DGC, irreversibly affects cell-cell and matrix adhesion, while transient E-cadherin LoF is associated to the cadherin-switch observed in IGC. If validated, the identification of Ecadherin LoF-associated markers will be fundamental for detection and management in GC patients.

Funding

This work was supported 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 – 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 and NORTE-01-0145-FEDER-000029 - 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, funded by ERDF, POCI and FCT, and project PTDC/CTM-NAN/120958/2010, from FCT. CP was supported by the grant SFRH/BD/113031/2015. GMA was supported by the Investigator FCT Program 2013 (IF/00615/2013), POPH - QREN Type 4.2, European Social Fund and Portuguese Ministry of Science and Technology (MCTES). Ipatimup integrates the i3S Research Unit, which is partially supported by FCT, the Portuguese Foundation for Science and Technology.

Acknowledgements

The authors acknowledge the support of the i3S Scientific Platforms “Advanced Light Microscopy”, member of the PPBI (PPBI-POCI-01-0145-FEDER-022122) and “Translational Cytometry Unit” and “GenCore” member of the National Laboratory for Genome Sequencing. The authors would like to thank Joana Paredes and Ana Sofia Ribeiro from Epithelial Interaction in

Cancer Group at i3S, for kindly lending P-cadherin antibody, and Gilza Gonçalves and Dina Leitão for kindly lending CK8/18 antibody and helping to optimize the antibody conditions.
C.P. acknowledges the support of the Doctoral Programme in Biomedicine at the Faculty of Medicine from University of Porto (FMUP).

Contributors

C.P., G.M.A. and C.O. were involved in the project conceptualization, writing and edition and formal analysis of this paper. H.O. prepared the samples for label free proteomic quantification. M.F., H.O. and P.O. performed the bioinformatics analysis behind proteomics and transcriptomics data and C.L. in the statistical analysis. The authors thank Ana Mafalda Rocha for the technical support with RNA quality.

REFERENCES

1. Corso, G.;Carvalho, J.;Marrelli, D.;Vindigni, C.;Carvalho, B.;Seruca, R.;Roviello, F.;Oliveira, C. Somatic mutations and deletions of the E-cadherin gene predict poor survival of patients with gastric cancer. *J Clin Oncol.* **2013**, 31, 868-875, DOI: 10.1200/jco.2012.44.4612
2. Peleteiro, B.;Lopes, C.;Figueiredo, C.;Lunet, N. Salt intake and gastric cancer risk according to Helicobacter pylori infection, smoking, tumour site and histological type. *British journal of cancer.* **2011**, 104, 198-207, DOI: 10.1038/sj.bjc.6605993
3. Figueiredo, C.;Garcia-Gonzalez, M.A.;Machado, J.C. Molecular pathogenesis of gastric cancer. *Helicobacter.* **2013**, 18 Suppl 1, 28-33, DOI: 10.1111/hel.12083
4. Cisło, M.;Filip, A.A.;Arnold Offerhaus, G.J.;Ciseł, B.;Rawicz-Pruszyński, K.;Skierucha, M.;Polkowski, W.P. Distinct molecular subtypes of gastric cancer: from Laurén to molecular pathology. *Oncotarget.* **2018**, 9, 19427-19442, DOI: 10.18632/oncotarget.24827
5. Ribeiro, M.M.;Sarmento, J.A.;Sobrinho Simões, M.A.;Bastos, J. Prognostic significance of Lauren and Ming classifications and other pathologic parameters in gastric carcinoma. *Cancer.* **1981**, 47, 780-784, DOI: 10.1002/1097-0142(19810215)47:4<780::aid-cncr2820470424>3.0.co;2-g
6. Adachi, Y.;Yasuda, K.;Inomata, M.;Sato, K.;Shiraishi, N.;Kitano, S. Pathology and prognosis of gastric carcinoma: well versus poorly differentiated type. *Cancer.* **2000**, 89, 1418-1424, DOI:
7. Xing, X.;Tang, Y.B.;Yuan, G.;Wang, Y.;Wang, J.;Yang, Y.;Chen, M. The prognostic value of E-cadherin in gastric cancer: a meta-analysis. *Int J Cancer.* **2013**, 132, 2589-2596, DOI: 10.1002/ijc.27947
8. Gabbert, H.E.;Mueller, W.;Schneiders, A.;Meier, S.;Moll, R.;Birchmeier, W.;Hommel, G. Prognostic value of E-cadherin expression in 413 gastric carcinomas. *Int J Cancer.* **1996**, 69, 184-189, DOI: 10.1002/(sici)1097-0215(19960621)69:3<184::Aid-ijc6>3.0.Co;2-w
9. Lazăr, D.;Tăban, S.;Ardeleanu, C.;Dema, A.;Sporea, I.;Cornianu, M.;Lazăr, E.;Vernic, C. The immunohistochemical expression of E-cadherin in gastric cancer; correlations with

clinicopathological factors and patients' survival. *Rom J Morphol Embryol.* **2008**, 49, 459-467, DOI:

10. Torabizadeh, Z.;Nosrati, A.;Sajadi Saravi, S.N.;Yazdani Charati, J.;Janbabai, G. Evaluation of E-cadherin Expression in Gastric Cancer and Its Correlation with Clinicopathologic Parameters. *International journal of hematology-oncology and stem cell research.* **2017**, 11, 158-164, DOI:
11. Anbiaee, R.;Mojir Sheibani, K.;Torbati, P.;Jaam, H. Abnormal expression of e-cadherin in gastric adenocarcinoma, and its correlation with tumor histopathology and helicobacter pylori infection. *Iran Red Crescent Med J.* **2013**, 15, 218-222, DOI: 10.5812/ircmj.4032
12. Schizas, D.;Moris, D.;Michalinos, A.;Kanavidis, P.;Oikonomou, D.;Papalampros, A.;Machairas, A.;Liakakos, T. E-cadherin in gastric carcinomas: Relations with histological parameters and its prognostic value. *J buon.* **2017**, 22, 383-389, DOI:
13. Oka, H.;Shiozaki, H.;Kobayasi, K.;Tahara, H.;Tamura, S.;Yamazaki, K.;Mori, T. [Immunoreactive expression of E-cadherin in human gastric cancer: preliminary report]. *Nihon Geka Gakkai zasshi.* **1990**, 91, 1814, DOI:
14. Solorzano, L.;Pereira, C.;Martins, D.;Almeida, R.;Carneiro, F.;Almeida, G.;Oliveira, C.;Wahlby, C. Towards automatic protein co-expression quantification in immunohistochemical TMA slides. *IEEE Journal of Biomedical and Health Informatics.* **2020**, 1-1, DOI: 10.1109/JBHI.2020.3008821
15. Morrisey Lab Protocol: Generating Large (>1kb) Genomic Deletions Using CRISPRs. Available online: https://www.med.upenn.edu/genetics/tcmf/protocols.shtml. (acessed on July 2, 2020)
16. Hughes, C.S.;Moggridge, S.;Müller, T.;Sorensen, P.H.;Morin, G.B.;Krijgsveld, J. Single-pot, solid-phase-enhanced sample preparation for proteomics experiments. *Nature Protocols.* **2019**, 14, 68-85, DOI: 10.1038/s41596-018-0082-x
17. Yokozaki, H. Molecular characteristics of eight gastric cancer cell lines established in Japan. *Pathology International.* **2000**, 50, 767-777, DOI: 10.1046/j.1440-1827.2000.01117.x
18. Tian, X.;Liu, Z.;Niu, B.;Zhang, J.;Tan, T.K.;Lee, S.R.;Zhao, Y.;Harris, D.C.H.;Zheng, G. E-cadherin/β-catenin complex and the epithelial barrier. *Journal of biomedicine & biotechnology.* **2011**, 2011, 567305-567305, DOI: 10.1155/2011/567305
19. Ratheesh, A.;Yap, A.S. A bigger picture: classical cadherins and the dynamic actin cytoskeleton. *Nat Rev Mol Cell Biol.* **2012**, 13, 673-679, DOI: 10.1038/nrm3431
20. Xie, T.;Pan, S.;Zheng, H.;Luo, Z.;Tembo, K.M.;Jamal, M.;Yu, Z.;Yu, Y.;Xia, J.;Yin, Q., et al. PEG10 as an oncogene: expression regulatory mechanisms and role in tumor progression. *Cancer Cell International.* **2018**, 18, 112, DOI: 10.1186/s12935-018-0610-3
21. Vieira, A.F.;Paredes, J. P-cadherin and the journey to cancer metastasis. *Mol Cancer.* **2015**, 14, 178, DOI: 10.1186/s12943-015-0448-4
22. van Roy, F. Beyond E-cadherin: roles of other cadherin superfamily members in cancer. *Nature Reviews Cancer.* **2014**, 14, 121-134, DOI: 10.1038/nrc3647

23. Kim, M.A.;Jung, E.J.;Lee, H.S.;Lee, H.E.;Yang, H.K.;Oh, D.Y.;Bang, Y.J.;Kim, W.H. P-cadherin expression in gastric carcinoma: its regulation mechanism and prognostic significance. *Hum Pathol.* **2010**, 41, 877-885, DOI: 10.1016/j.humpath.2009.04.031

APPENDIX

This appendix includes supplementary figures and tables.

Figure S1. Bioinformatics analysis of proteomic data. **(A)** Raw data was filtered based on the four criteria: (1) number of unique peptides higher or equal to two and contaminants is false; (2) Adjusted *p* value equal or lower than 0.05; (3) Abundance ratio bigger than 1.5 (upregulated) or lower than 0.67 (downregulated); and (4) Concordance between replicates. **(B)** Data filtering in diffuse and intestinal-like *in vitro* cell models.

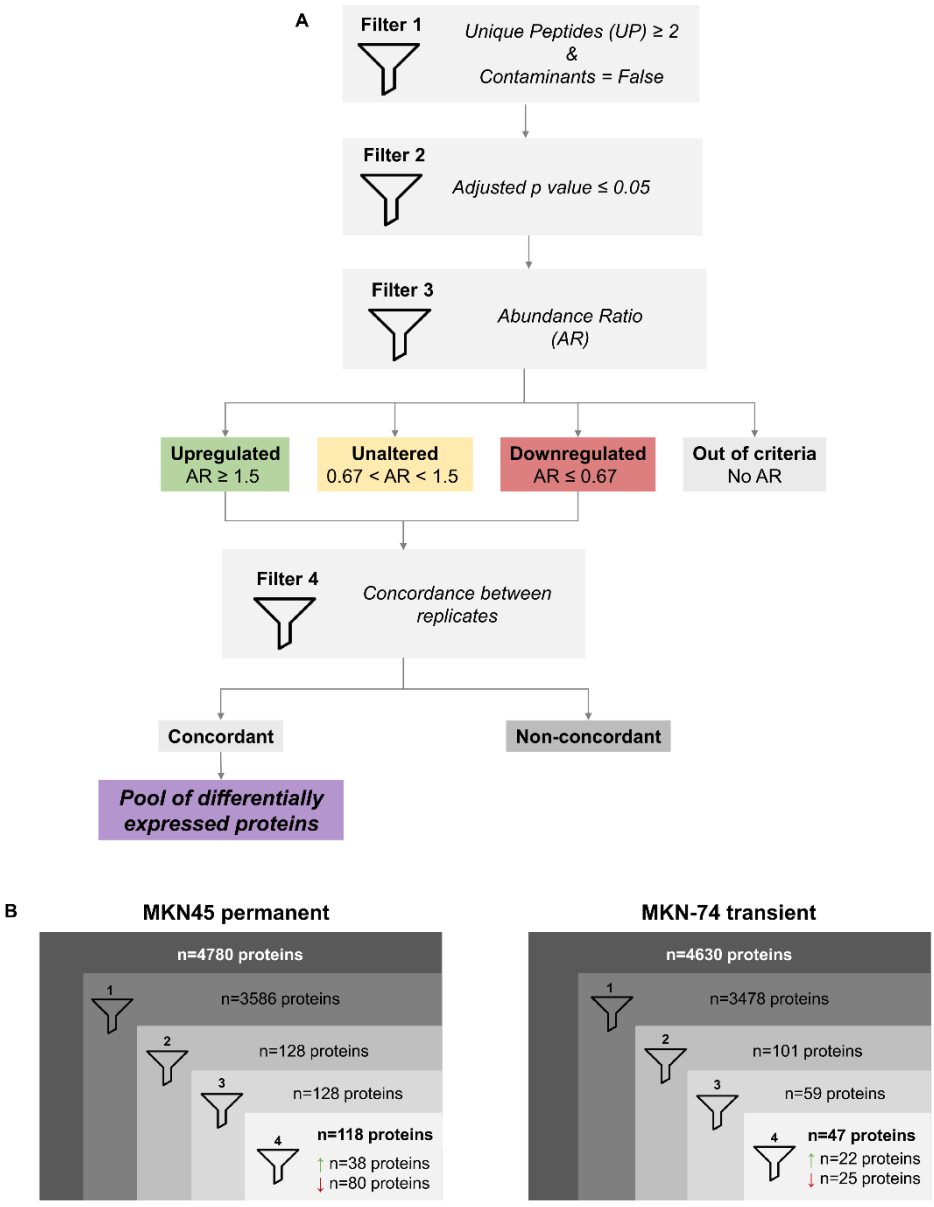


Figure S2. Functional protein association network in the diffuse-like setting (MKN45 permanent model). Red or green nodes represent downregulated or upregulated protein-coding coding, respectively. Disconnected protein coding genes were not included. This analysis was performed using STRING online tool.

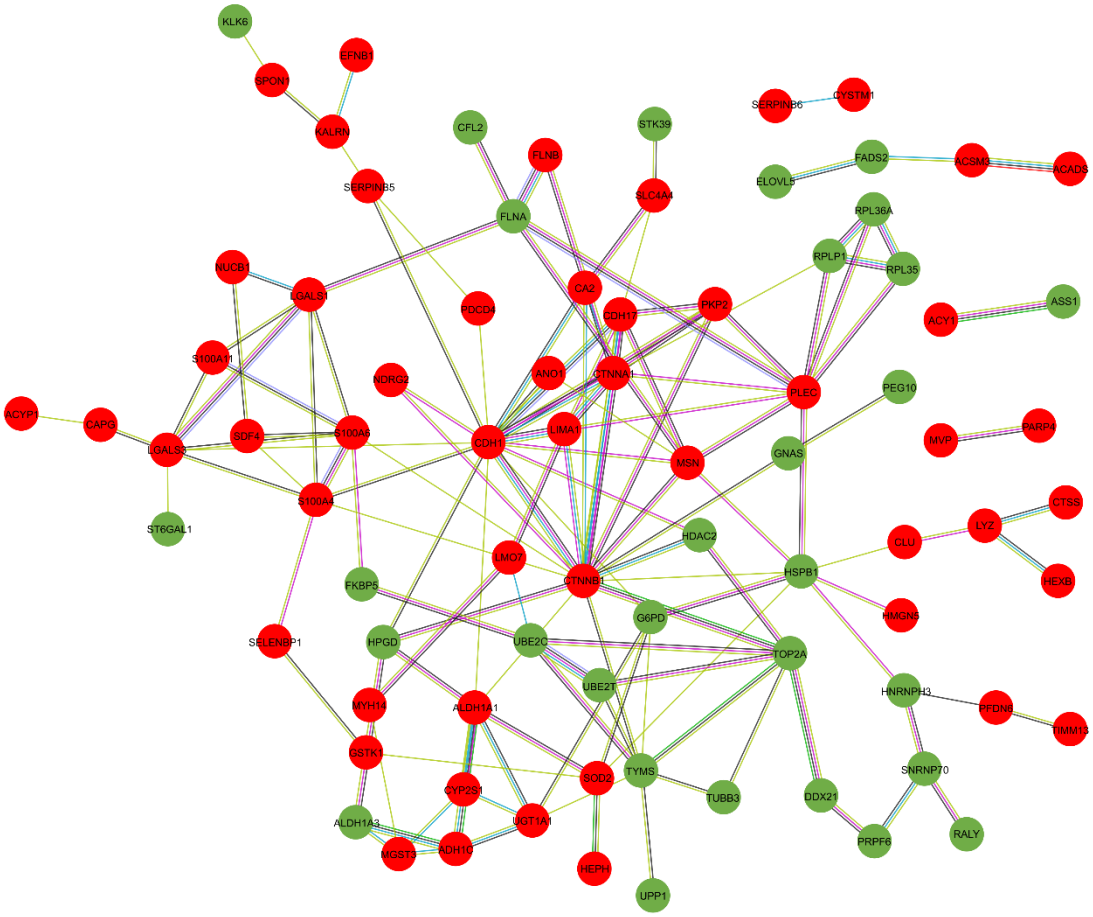


Figure S3. Functional protein association network in the intestinal-like setting (MKN74 transient model). Red or green nodes represent downregulated or upregulated protein-coding coding, respectively. Disconnected protein coding genes were not included. This analysis was performed using STRING online tool.

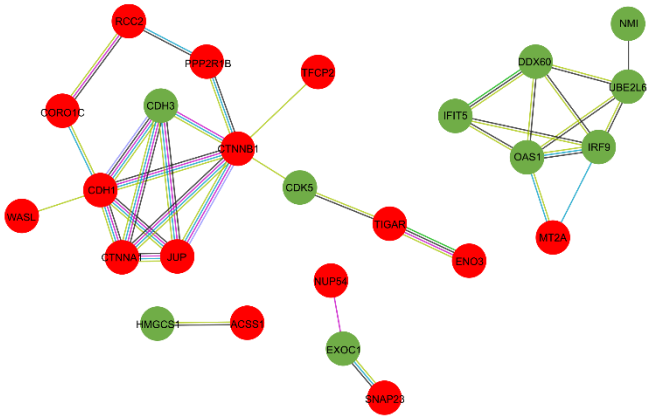


Table S1. Details of the primers designed to validate CRISPR/Cas9 constructs (primer A and C) and genotype the MKN-45-derived knockout clones obtained by CRISPR/Cas9 (primer B and D).

Direction	In-text name	Binding Site	Sequence	Melting temperature
Forward	Primer A	pSpCas9(BB)_Bbs I	5' GGGCCTATTTCCCATGATTCCTT 3'	Tm = 68.5°C
	Primer B	*CDH1* Intron 2	5' TCAGAGCACAAGGAAGTCATC 3'	Tm = 64°C
Reverse	Primer C	pSpCas9(BB)_Bbs I	5' GACTCGGTGCCACTTTTTCAA 3'	Tm = 66.2°C
	Primer D	*CDH1* Intron 3	5' CTGAACACCGTCTGGGACTG 3'	Tm = 68°C

Table S2. List of differentially expressed proteins in the diffuse-like setting (MKN45 permanent model).

Uniprot ID	Gene	Abundance Ratio	Post-translational modification
Q9UG63	ABCF2	2.361	
P16219	ACADS	0.441	
Q53FZ2	ACSM3	0.392	
Q03154	ACY1	0.513	
G3V2U7	ACYP1	0.465	
Q9UKQ2	ADAM28	100	
P00326	ADH1C	0.458	
Q8TD06	AGR3	0.179	
A0A0J9YWL0	AIM1	0.371	Phospho [S1301(97.8); S1308(97.8)]
O60218	AKR1B10	0.303	
P00352	ALDH1A1	0.431	
P47895	ALDH1A3	3.102	
Q5XXA6	ANO1	0.138	
P00966	ASS1	2.167	
P00918	CA2	0.341	
P40121	CAPG	0.514	Acetyl [N-Term]
Q8N163	CCAR2	0.496	
P12830	CDH1	0.01	
Q12864	CDH17	0.173	
Q9Y281	CFL2	2.958	
P12277	CKB	0.406	
P12532	CKMT1B	0.418	
P10909	CLU	0.474	
P35221	CTNNA1	0.257	Phospho [S641(100)]
A0A2R8YCH5	CTNNB1	0.167	Acetyl [N-Term]; Phospho [S552(99.6)]
P25774	CTSS	0.463	
Q96SQ9	CYP2S1	0.388	
Q9H1C7	CYSTM1	0.327	
O95865	DDAH2	0.359	
Q9NR30	DDX21	2.035	
O00115	DNASE2	0.576	
Q8IYU8	EFHA1	0.146	
P98172	EFNB1	0.511	
F8W9E7	ELMO3	0.336	
Q9NYP7	ELOVL5	3.639	Acetyl [N-Term]
Q6P179	ERAP2	0.268	
Q76MJ5	ERN2	0.327	
O95864	FADS2	2.999	
E9PJX4	FGD4	0.399	
Q13451	FKBP5	4.25	
P21333	FLNA	2.299	
O75369	FLNB	0.495	
P49356	FNTB	0.01	
P11413	G6PD	2.222	
Q5JWF2	GNAS	1.966	

Q9Y2Q3	GSTK1	0.317	
Q92769	HDAC2	1.914	Phospho [S394(100)]
A0A0C4DG76	HEPH	0.473	
P07686	HEXB	0.521	
P16402	HIST1H1D	2.355	
P54868	HMGCS2	0.508	
Q5JSL0	HMGN5	0.187	
P31942	HNRNPH3	2.921	
P15428	HPGD	4.157	
P04792	HSPB1	2.384	
H7BXZ5	KALRN	0.508	
Q92876	KLK6	100	
O00515	LAD1	0.55	
Q8IVL5	LEPREL1	0.368	
P09382	LGALS1	0.297	
P17931	LGALS3	0.515	
Q9H008	LHPP	0.416	
Q9UHB6	LIMA1	0.275	Phospho [S362(100); S604(97.3)]
E9PMS6	LMO7	0.437	
Q9BS40	LXN	0.439	
P61626	LYZ	0.127	
O14880	MGST3	0.367	
P26038	MSN	0.395	
Q14764	MVP	0.467	
Q969H8	MYDGF	0.488	
Q7Z406	MYH14	0.284	
Q9UBC5	MYO1A	0.326	
A0A0A0MS87	NDRG2	0.514	
Q02818	NUCB1	0.512	
Q6DKJ4	NXN	4.851	
O95340	PAPSS2	2.732	
Q9UKK3	PARP4	0.507	
Q53EL6	PDCD4	0.29	
A0A087WXK2	PEG10	100	
O15212	PFDN6	0.536	
Q99959	PKP2	0.476	
Q6P4A8	PLBD1	0.547	
Q15149	PLEC	0.516	
O94906	PRPF6	2.296	
Q15274	QPRT	6.502	Acetyl [N-Term]
Q9UKM9	RALY	2.693	
Q15382	RHEB	2.247	
P42766	RPL35	2.023	
J3KQN4	RPL36A	2.665	
P05386	RPLP1	2.02	
H0YER1	RSF1	2.628	Acetyl [N-Term]
P31949	S100A11	0.471	
P26447	S100A4	0.279	
P06703	S100A6	0.503	
Q9H190	SDCBP2	0.305	
H0Y3T6	SDF4	0.437	
Q13228	SELENBP1	0.172	
P36952	SERPINB5	0.358	Acetyl [N-Term]
P35237	SERPINB6	0.518	Acetyl [N-Term]
O75368	SH3BGRL	0.217	
Q9Y6R1	SLC4A4	0.413	
P08621	SNRNP70	2.247	
P04179	SOD2	0.549	
Q9HCB6	SPON1	0.268	
P15907	ST6GAL1	4.154	
Q9UEW8	STK39	2.486	
P0DMN0	SULT1A4	0.327	Acetyl [N-Term]
Q9H2K8	TAOK3	0.342	
Q9Y5L4	TIMM13	0.506	

P11388	TOP2A	2.88	
P55327	TPD52	0.551	
Q7Z392	TRAPPC11	0.01	
P19075	TSPAN8	0.491	
Q13509	TUBB3	2.584	Acetyl [K336(100)]
P04818	TYMS	2.048	
O00762	UBE2C	2.178	
Q9NPD8	UBE2T	2.41	
P22309	UGT1A1	0.298	
Q16880	UGT8	2.057	
Q16831	UPP1	2.197	

Table S3. List of differentially expressed proteins in the intestinal-like setting (MKN-74 transient model).

Uniprot ID	Gene	Abundance Ratio	Post-translational modification
Q9NUB1	ACSS1	0.631	
F5H6I7	ATL3	0.589	
Q6DT37	CDC42BPG	0.588	
P12830	CDH1	0.109	
P22223	CDH3	2.565	
Q00535	CDK5	1.623	
Q9ULV4	CORO1C	0.468	
P35221	CTNNA1	0.532	
A0A2R8YCH5	CTNNB1	0.551	Phospho [S675(100)]
O94760	DDAH1	0.571	
Q8IY21	DDX60	1.519	
P13929	ENO3	0.623	Acetyl [K358(100)]
Q9NV70	EXOC1	1.713	
Q9BZQ8	FAM129A	1.569	
Q9BRX5	GINS3	1.668	
Q01581	HMGCS1	1.61	
Q13325	IFIT5	1.554	
Q00978	IRF9	1.559	
P14923	JUP	0.613	
J3KT75	MPDU1	1.709	
P02795	MT2A	0.666	Acetyl [N-Term]
A0A0R4J2G3	NCEH1	1.627	
C9JE98	NCOR2	0.01	
Q9UMS0	NFU1	2.205	
Q13287	NMI	1.542	Acetyl [N-Term]
Q86WQ0	NR2C2AP	1.537	
Q7Z3B4	NUP54	0.557	
P00973	OAS1	2.098	
P68402	PAFAH1B2	1.618	
Q6PCE3	PGM2L1	1.603	
Q9Y3C6	PPIL1	0.477	
Q9ULR3	PPM1H	1.504	
P30154	PPP2R1B	0.608	
Q9P258	RCC2	0.658	
O00161	SNAP23	0.573	
Q9H2K8	TAOK3	0.624	
Q92609	TBC1D5	0.396	
Q12800	TFCP2	0.621	
Q9NQ88	TIGAR	0.646	
Q9BSF4	TIMM29	0.01	
Q86WV6	TMEM173	1.554	
Q8WWH5	TRUB1	0.564	
Q8NBS9	TXNDC5	0.669	
Q16222	UAP1	0.616	
O14933	UBE2L6	1.558	
O00401	WASL	0.416	
Q969T9	WBP2	0.519	

Chapter 6

General Discussion and Concluding Remarks

A detailed analysis of the results is presented along each publication (Chapter 3, 4 and 5), therefore we prepared a general and integrative discussion in light of the recent advances in GC.

GC molecular complexity and heterogeneous character nourished the interest of exploring multiple biomarkers to improve the translational value of biomarkers in the clinical setting. This thesis was dedicated to explore the role and clinical implications of two cell adhesion molecules recurrently altered throughout GC development: E-cadherin and CD44v6. While E-cadherin has a firm reputation in the GC setting, the role of CD44v6 is unexplored territory in the GC. Along this thesis, these molecules were analyzed from the individual and co-operative standpoint.

Historically, the benefit of using computational tools for validation of tissue biomarkers was weakened by the shortcomings of image analysis to automatically recognize clinically relevant areas and to the difficulties associated with the management of gigapixel images [1]. Although pathologists are more versatile across different clinical and pathological settings, computer algorithms have the advantage of objectively assess, quantify and relate a large number of features in a systematic and high-throughput way, when trained for specific applications. In this thesis, we established an image analysis pipeline to define tumor regions, align immunohistochemical slides, and quantify E-cadherin and CD44v6 co-expression. This tool was developed based on the pathologists' methodology, who prior to the development of this pipeline, visually inspected E-cadherin and CD44v6 expression profiles and score them. By comparing the scoring performance of pathologists' manual *versus* computer assessment of CD44v6 and E-cadherin individually, we observed that our method performed better quantifying *de novo* CD44v6 expression than E-cadherin LoF, mostly due to the difficulty of distinguishing and estimating the multiple of patterns assumed by E-cadherin. Notwithstanding, this result evidenced that computer-aided classification of biomarker expression can facilitate the identification of equivocal cases with high risk of misdiagnosis. Through cross-slide image registration, we observed that the co-occurrence of E-cadherin LoF and *de novo* CD44v6 expression is very frequent, however the degree of co-expression of both proteins varies widely.

Given the likelihood of both events occurring in the same tumor and the myriad of events in which E-cadherin and CD44v6 are involved, namely tumor dissemination and invasion, we move forward to explore the concomitant impact of E-cadherin and CD44v6 alterations in tumor aggressiveness and patients' outcome. To do so, we used two contemporary GC cohorts from distinct world regions - Portugal, representing Western Europe cohort (WE-C) and South Korea, representing the East Asia cohort (EA-C) - and correlated the expression profiles of E-cadherin and CD44v6 with the survival and clinicopathological features of these GC patients. Our analysis showed that ~12% of chemo-naive WE-C and EA-C GC patients presenting concomitant abnormal E-cadherin and very high CD44v6 expression had particular aggressive tumors. However, survival rates were the opposite in the two cohorts, very poor in the WE-C and high in the EA-C. Early stage EA-C patients from this particular subset were diagnosed 8-years younger and their tumors permeated more the lymphatic vessels, promoting lymph node (LN) metastasis, usually triggered in

early stages of GC. In contrast, tumors from this WE-C patient subset invaded deeper into the gastric wall, more often permeating the vasculature and nerves. Indeed, the potential consequences of nodal metastization were counteracted by an extensive LN dissection in the EA-C, explaining the impressive outcomes, while in the WE-C, it would not be simple to neutralize tumor aggressiveness surgically as cancer appears to have spread for more than one route. For the best of our knowledge, this is the first report describing the mutual involvement of E-cadherin and CD44v6 in GC aggressiveness in two independent cohorts. Further studies with a single-cell resolution and at a molecular level are essential to clarify E-cadherin and CD44v6 interplay. Notwithstanding, by using our tool for image registration, we observed that it is very likely that E-cadherin LoF and *de novo* CD44v6 occur concomitantly in the same cancer cells' population, which supports patients' stratification based on both molecular features. Based on the evidences collected from both cohorts, we suggest that the assessment of concomitant E-cadherin and CD44v6 expression in tumor biopsies can help stratify GC patients predicted to be more aggressive features and with a higher likelihood of disseminate. This information is particularly important in WE countries, where routine screening strategies are far from being implemented.

Survival outcomes are not only dependent on how early the tumor is detected and whether it can be completely surgically removed, but also whether it responds to chemotherapy (CT). CD44v6-containing isoforms have been widely associated to drug resistance in several cancers, including GC. In this sense, we then dedicated to unveil the role of *de novo* CD44v6 expression in the transformative process and/or progression of GC, if it conditions response to therapy, or if it is only a bystander marker in all or some of these events. To do so, we generated isogenic cell lines by stably expressing a CD44 standard isoform (CD44s) and CD44v6-containing isoform highly expressed in GC cells, and tested them for different cancer cell hallmarks and to the conventional chemotherapeutic drug, cisplatin. Expression of CD44v6 or CD44s were not differently associated to any *cancer-related* features like doubling-time, invasion, cell-adhesion, cell motility or in vivo tumorigenic potential. Our *in vitro* results, somehow, corroborate with our previous observations in patients' samples. Indeed, d*e novo* CD44v6 expression was not restricted to advanced GCs, but was observed throughout the process of malignant transformation of gastric mucosa including in precursor lesions (intestinal metaplasia, dysplasia) [2]. Therefore, it is not surprising that CD44v6 is not a driver of GC development. However, it remains to be elucidated whether CD44v6 expression in premalignant lesions indicates early malignant changes or, are likely to have a normal or physiological role in the gastric mucosa, whose expression is extended into pathological contexts.

Regarding the role of CD44v6 in response to therapy, our data suggests that CD44v6 influences the response to cisplatin chemotherapeutic agents, leading to decreased cisplatin-induced apoptosis compared to CD44v6 negative cells (either Mock or CD44s). Likewise, downregulation of CD44v6 containing isoforms, in GC cell lines that endogenously express it, led to increased sensitivity to cisplatin-induced apoptosis. Moreover, we dissect the mechanism by which CD44v6 expressing cells have increased cisplatin survival in GC mediated by Stat3 and P38 signalling pathways. Stat3 or P38 signalling pathways have been described as mediators of survival in response to cisplatin [3, 4]. None of the three cell lines present phosphorylated Stat3

(pStat3) at basal level. Following 24 hours of cisplatin treatment, pStat3 levels increased in CD44+ cells. Although Stat3 activation becomes a general survival response upon 48 hours of cisplatin treatment, CD44v6+ cells seem to more rapidly induce and maintain Stat3 activation. In ovarian cancer, high expression of activated Stat3 has been shown to contribute to cisplatin resistance [5]. Moreover, abrogation of Stat3 activity has been shown to circumvent cisplatin resistance in ovarian cancer cells [6], making it a promising target to reverse cisplatin resistance in ovarian and other cancer types [3].

Considering the distinct response of CD44v6+ and CD44v6- cells to cisplatin treatment, it is likely that the proportion of CD44v6+ and CD44v6- cells in heterogeneous tumors influences response to chemotherapy. Our experiments using co-cultures of CD44v6+ and CD44v6- (in the same proportion) showed that CD44v6+ cells have indeed a selective advantage in a heterogeneous setting. At tumor level, the overgrowth of CD44v6+ cells after cisplatin treatment may indicate that this cell population can get enriched after chemotherapy and ultimately arise in GC relapse. This is plausible as CD44/CD44v are considered CSC markers in several tumor types, including GC [7, 8].

Driven by these findings, we then investigated whether exon-v6 itself is responsible for conferring these properties to CD44v6-containing isoforms. In this study, we used CRISPR/Cas9 to remove exon-v6 from the CD44 locus, by targeting splice-sites and/or adjacent regions, to generate specific exon-v6 skipping from CD44v6-containing isoforms in two GC cell lines. Our results show that specific exon-v6 skipping induces increased cell sensitivity to cisplatin and decreased self-renewal capacity, to the same extent as the full depletion of CD44v6-containing isoforms. In sum, our data highlight a feasible alternative based on PMO's based approach to full CD44v6-containing isoforms depletion, raising the possibility of using exon-v6 skipping as targeted therapy to delay GC cells self-renewal and sensitize gastric tumors to chemotherapy.

Additionally, we are currently exploring whether CD44v6 expression in tumors predicts the benefit from patients in receiving conventional chemotherapy in addition to surgery (**Appendix A, Original Paper 6**). Moreover, it is important to note whether this benefit is proportional to the fraction of CD44v6+ cells in tumors. In colorectum cancer, CD44v6 was proven a useful marker of therapy response, where patients with moderate or strong CD44v6 tumor expression respond better to chemotherapy with irinotecan [9].

E-cadherin LoF leads to aberrant or absent protein expression. Although dysfunctional, aberrant E-cadherin expression is still present and can assume a variety of patterns, which obscures the analytical path. During the development of our image analysis pipeline (*Original Paper 1*), we realized that the evaluation of E-cadherin expression is also challenging for the computer. Indeed, our pipeline performed well quantifying absent E-cadherin protein expression, but for the computer, aberrant (dysfunctional) or normal (functional) E-cadherin expression were ambiguous. Given that E-cadherin itself is not helpful to study E-cadherin LoF, we set out to identify overexpressed protein as surrogate markers of E-cadherin LoF with potential clinical. While E-cadherin permanent loss is the trigger for diffuse GC (DGC), aberrant E-cadherin expression is common along the progression of intestinal GC (IGC). Therefore, E-cadherin depleted cell models, permanent (CRISPR/Cas9) to

mimic the diffuse GC type; and transient (RNAi) to mimic intestinal GC were established. By label-free quantitative proteomics profiling we found that permanent E-cadherin LoF dramatically affects cell-cell and cell-matrix adhesion and results in *de novo* expression of three protein-coding genes: *PEG10*, *KLK6* and *ADAM28* by E-cadherin deficient cells. Based on literature research, we got particularly interested in PEG10, which presents low expression levels in stomach normal tissue, but it is overexpressed in a variety of cancers, including GC [10]. We validated the expression of PEG10 in our *in vitro* cell model and performed a small pilot study to analyze the expression of PEG10 in GC patients with diffuse and mixed-type GC. Most of the cases did not expressed PEG10, but some of them presented some staining in cells present in the middle of tumor cells. However, we confirmed by cytokeratin 8/18 marker that this staining was not at the tumor cells. In sum, PEG10 has not yet been confirmed as a surrogate marker of E-cadherin LoF using patients tumours. The transient E-cadherin LoF model was associated with a cadherin-switch, from E- to P-cadherin, which is related with adherens junction organization, but is also linked to significant tumour promoting effects, including in GC [11]. As before, we performed a small pilot study in GC tumor samples from intestinal-type and we observed that regions with aberrant E-cadherin expression present P-cadherin expression. We will further validate the clinical significance of P-cadherin as a surrogate marker of E-cadherin LoF in a large cohort of GC patients.

Table 1. New insights and practical implications disclosed in this thesis.

CONTEXT	PRACTICAL IMPLICATIONS	REFERENCE
The Co-Operative Role of E-cadherin and CD44v6 in GC *(Chapter 3)*	Automated image registration of TMA imunohistochemical slides highlights E-cadherin and CD44v6 co-occurrence and allows patients' stratification based on the combined expression of these markers;	Solorzano L, Pereira C *et al.* (2020) IEEE J Biomed Health
	Concomitant E-cadherin dysfunction and CD44v6 overexpression determines disease aggressiveness in early GC, which may lead to dismal outcomes if appropriate surgical approaches are not performed;	Pereira C *et al.* Ready for submission
Mechanisms of Therapy Response Mediated by CD44v6 *(Chapter 4)*	CD44v6 influences response to cisplatin treatment in GC cells likely through likely pSTAT3 or pP38 and is involved in GC cell overgrowth after cisplatin treatment;	Pereira C and Ferreira D *et al.* (2020) Cancers, 12 (4), 858
	Removing only exon-v6 from CD44v6-containg isoforms results in cancer stem cells feature attenuation, by decreasing the self-renewal capacity and resistance to cisplatin of GC cells;	Lobo S and Pereira C *et al.* (2020) Cancers, 12 (9), 2378
Surrogate Markers for E-cadherin LoF *(Chapter 5)*	E-cadherin LoF triggers a cadherin-switch, from E- to P-cadherin, which may aid the analysis of E-cadherin dysfunction in GC tumor samples.	Pereira C *et al.* Manuscript in preparation

REFERENCES

1. Louis, D.N., et al., *Computational Pathology: A Path Ahead.* Arch Pathol Lab Med, 2016. **140**(1): p. 41-50.
2. da Cunha, C.B., et al., *De novo expression of CD44 variants in sporadic and hereditary gastric cancer.* Lab Invest, 2010. **90**(11): p. 1604-14.
3. Sun, C.-Y., et al., *Targeting STAT3 inhibition to reverse cisplatin resistance.* Biomedicine & Pharmacotherapy, 2019. **117**: p. 109135.
4. Achkar, I.W., et al., *Cisplatin based therapy: the role of the mitogen activated protein kinase signaling pathway.* Journal of Translational Medicine, 2018. **16**(1): p. 96.
5. Wu, C.-J., et al., *Activation of STAT3 and STAT5 Signaling in Epithelial Ovarian Cancer Progression: Mechanism and Therapeutic Opportunity.* Cancers, 2019. **12**(1): p. E24.
6. Ji, T., et al., *Abrogation of constitutive Stat3 activity circumvents cisplatin resistant ovarian cancer.* Cancer letters, 2013. **341**(2): p. 231-239.
7. Zoller, M., *CD44: can a cancer-initiating cell profit from an abundantly expressed molecule?* Nat Rev Cancer, 2011. **11**(4): p. 254-67.
8. Kodama, H., et al., *Prognostic impact of CD44-positive cancer stem-like cells at the invasive front of gastric cancer.* British Journal of Cancer, 2017. **116**(2): p. 186-194.
9. Bendardaf, R., et al., *CD44 variant 6 expression predicts response to treatment in advanced colorectal cancer.* Oncology reports, 2004. **11**(1): p. 41-45.
10. Xie, T., et al., *PEG10 as an oncogene: expression regulatory mechanisms and role in tumor progression.* Cancer Cell International, 2018. **18**(1): p. 112.
11. Vieira, A.F. and J. Paredes, *P-cadherin and the journey to cancer metastasis.* Mol Cancer, 2015. **14**: p. 178.

Chapter 7

Appendices

Appendix A

CD44v6 expression is a novel predictive marker of therapy response and poor prognosis in gastric cancer patients

Carla Pereira, Daniel Ferreira, Carolina Lemos, Diana Martins, Nuno Mendes, Daniela Almeida, Pedro Granja, Fátima Carneiro, Raquel Almeida, Gabriela M. Almeida, Carla Oliveira,

bioRxiv. (2018): 468934

The content of this article was splitted in two articles for publication purposes. The portion encompassing the in vitro data was recently published in Cancers and can be found in Chapter 4 (Original Paper 3) of this thesis.

The patient data was re-analyzed taking into consideration another independent GC patient cohort from South Korea. This manuscript is currently under preparation.

CD44v6 expression is a novel predictive marker of therapy response and poor prognosis in gastric cancer patients

Carla Pereira [1,2*], Daniel Ferreira [1,2,3*], Carolina Lemos [1,4,5], Diana Martins[1,2], Nuno Mendes [1,2], Daniela Almeida [6], Pedro Granja[1,3,5], Fátima Carneiro[1,2, 7,8], Raquel Almeida[1,2,7,9], Gabriela M Almeida[1,2,7#], Carla Oliveira[1,2,7#]

* Both CP and DF contributed equally to the work

Correspondence to:

C. Oliveira, Expression Regulation in Cancer Group, Ipatimup, Instituto de Investigação e Inovação em Saúde (I3S), Rua Alfredo Allen, Nº 208, 4200-135 Porto Portugal, Tel: +351 220 408 800; E-mail: carlaol@ipatimup.pt

or

G.M. Almeida, Expression Regulation in Cancer Group, Ipatimup, Instituto de Investigação e Inovação em Saúde (I3S), Rua Alfredo Allen, Nº 208, 4200-135 Porto Portugal, Tel: +351 220 408 800; E-mail: galmeida@ipatimup.pt

Authors affiliations:

1 - i3S - Instituto de Investigação e Inovação em Saúde, Universidade do Porto, Porto, Portugal;

2 - Ipatimup - Institute of Molecular Pathology and Immunology of the University of Porto, Porto, Portugal;

3 - INEB, Instituto de Engenharia Biomédica, Universidade do Porto, Porto, Portugal.

4 - IBMC - Institute for Molecular and Cell Biology, Porto, Portugal;

5 - ICBAS - Instituto Ciências Biomédicas Abel Salazar, Universidade do Porto, Porto, Portugal;

6 - Medical Oncology Department, Centro Hospitalar de São João, Porto, Portugal;

7 - Faculty of Medicine of the University of Porto, Porto, Portugal;

8 - Department of Pathology, Centro Hospitalar de São João, Porto, Portugal;

9 - Faculty of Sciences of the University of Porto, Porto, Portugal.

Keywords: gastric cancer, CD44v6, prognostic marker, predictive marker, therapy response, gastric cancer survival, tumor heterogeneity

Abstract:

Late diagnosis, modest treatment options and lack of predictive markers of therapy response dictate the poor overall survival (OS) of ~1 year in most gastric cancer (GC) patients. We hypothesized that the level of CD44v6 expression in tumor cells could predict therapy response and prognosis in GC patients.

We analyzed a surgical tumor series of GC patients for the extension of CD44v6 membranous immuno-expression, clinical-pathological features, patient survival, and response to therapy. By integrating this information, we assessed the value of CD44v6 expression to predict benefit from current treatment regimens and prognosis in GC patients. We used GC cell lines and mouse xenografts to assess and/validate the biological impact of CD44v6 expression in GC cells behavior.

We demonstrated that GC patients whose tumors present higher levels of CD44v6 membranous expression benefit from adding chemotherapy to surgery as opposed to those without CD44v6 expression. Moreover, patients bearing CD44v6_high tumors presented worse OS than those bearing CD44v6_absent/low tumors, consolidating the role of CD44v6 expression as an independent factor of poor prognosis in this disease. Finally, our *in vitro* and patients' data pinpoints the CD44v6+ cell population as the driver of tumor recurrence following conventional chemotherapy, in heterogeneous tumors composed by CD44v6- and CD44v6+ cells.

Our study pioneers the identification of CD44v6 as a potential predictive marker of response to conventional chemotherapy, and consolidates CD44v6 as an independent marker of poor prognosis in GC. Overall, our data strongly

supports selection of patients with high CD44v6 expressing tumors for conventional chemotherapy with or without surgery, regardless of the TNM stage.

Introduction:

Gastric cancer (GC) is the 3rd leading cause of cancer related mortality worldwide, with >750,000 deaths estimated to occur every year.[1] Over 70% of GC patients present with locally advanced and/or unresectable disease, for whom conventional chemotherapy (mostly platinum-based) becomes the main treatment option, with a median overall survival (OS) of ~1 year.[2,3] Despite treatment improvements, the use of targeted therapies in GC has proved disappointing,[4] with the only approved therapies (Trastuzumab, against HER2 and Ramucirumab, against VEGFR2) showing limited OS improvement (2 to 3 months).[5,6] Therefore, it is crucial to identify biomarkers that can better define prognosis, but mainly that allow predicting who is likely benefitting more from a given treatment regimen. This would trigger better patient stratification for treatment, likely improving OS.

CD44, the main hyaluronic acid receptor, is a transmembrane glycoprotein involved in key cellular processes, such as lymphocyte activation, recirculation and homing, as well as epithelial cell adhesion and migration.[7,8] The human CD44 gene (NG_008937) encodes a polymorphic group of proteins generated by alternative splicing. The standard CD44 isoform (CD44s) includes only the constitutive exons, while the variant isoforms (CD44v) contain one or more variable exons (in addition to the constitutive ones).[7-9] CD44s is ubiquitously expressed at the surface of most mammalian cells, whereas the expression of CD44v is highly restricted (*e.g.* during lymphocyte activation and homing) and to specific tissue types.[7-10] Aberrant expression of CD44v isoforms may occur in diseased cells, and have been associated particularly with several cancer-associated features like invasion and metastization.[9,11,12] Moreover, CD44 and

its CD44v isoforms are expressed as surface markers of cancer stem cells (CSCs), influencing key CSC-associated properties such as tumor initiation, self-renewal, metastasis and chemoresistance.[9,13-15]

In the stomach, we have shown that CD44s isoform is widely expressed in both normal and diseased gastric epithelial cells.[16] In contrast, we showed that CD44v6-containing isoforms, which are absent from normal gastric epithelial cells, become overexpressed in stomach premalignant and malignant lesions and display high expression in ~70% of all GCs.[16]

Meta-analysis data showed that positive CD44 expression was significantly associated with worse GC patients' prognosis and lower OS, and suggested a similar association for CD44v6 expression.[17] Comparable data exists for other cancer types such as lung, colorectal and breast.[13] Recently, two reports showed a positive association between increased expression of CD44v6 and worse OS in GC patients.[18,19] However, both studies showed a positive correlation between CD44v6 expression and higher TNM staging, leaving reasonable doubt as to whether CD44v6 tumor expression is an independent factor of poor prognosis in GC. In addition, most articles published on CD44v6 in GC cohorts included a limited number of patients, making it difficult to obtain statistically significant results. No reports exist, exploring whether tumor CD44v6 expression influences patient therapy response.

Therefore, we aimed to clarify the role of CD44v6 in GC by using a large GC cohort to investigate, simultaneously, the relationship between CD44v6 expression and clinical-pathological features, patient OS and therapy response.

Materials and Methods:

All reagents were purchased from ThermoFisher Scientific (Waltham, MA, USA) unless otherwise stated.

Patient sample/data collection and tissue microarray (TMA) preparation

A cohort of 334 GC patients surgically treated at Centro Hospitalar de São João (CHSJ), Portugal, and part of the Tumor biobank of CHSJ/Ipatimup[20] was assembled (n=334) (Appendix Fig 1). A TMA was prepared from paraffin-embedded tumor material and patients' clinical-pathological, treatment and follow-up data collected. This study was approved by the Institutional Ethics Committee of CHSJ (Ethics Committee references CES 122/15 and CES 117/18), and informed patient consent obtained. This study is REMARK compliant.

CD44v6 immunohistochemistry of GC samples

Immunohistochemistry (IHC) staining for CD44v6 (clone MA54) was performed in 3-µm sections from TMAs, using the automated Ventana BenchMark ULTRAStaining System, using the OptiView DAB IHC Detection Kit (both from Roche/Ventana Medical Systems, Tucson, AZ, USA) according to manufacturers' instructions. Positive (human skin) and negative staining controls were performed in parallel with paraffin sections. The percentage of tumor cells displaying membranous expression of CD44v6 was assessed, for each sample, by four researchers in two independent evaluations performed in a blind manner. Four categories were defined to classify the extent of CD44v6 tumor expression: "CD44v6_0" – no staining at the cell membrane; "CD44v6_1+" – membranous

staining in less than 10% of tumor cells; “CD44v6_2+” – membranous staining in between 10 and 50% of tumor cells; “CD44v6_3+” – membranous staining in more than 50% of the tumor cells (Fig 1A).

Cell lines and culture conditions

Human GC cell lines MKN74 and MKN45 were purchased from the JCRB Cell Bank (Japanese Collection of Research Bioresources Cell Bank), and the non-commercial cell line GP202 cell line was established at Ipatimup.[21] All cell lines were cultured in RPMI 1640 medium with 10% heat inactivated fetal bovine serum (FBS) (Biowest, Nuaillé, France). Cell lines were maintained at 37 ºC and 5% CO_2 in a high humidity atmosphere and sub-cultured every 3 to 4 days. Cells were grown in the absence of antibiotics except for cell selection in MKN74 cells, where G418 was used. Cells were never continuously cultured for more than 4 months. Cell identification was confirmed by STR analysis and cells were confirmed to be free of mycoplasma contamination.

Generation of an isogenic cell line model of tumor CD44v6 status

Sequences for CD44s and CD44v6 (a v3-v10 transcript present in GP202) were cloned into a pIRES-EGFP2 plasmid.[22] MKN74 cells were transfected with either the empty vector (MKN74_Mock), pIRES-EGFP2_CD44v6 (MKN74_CD44v6) or pIRES-EGFP2_CD44s (MKN74_CD44s) and pure cell populations obtained by selective pressure with 1 mg/mL of G418 and bead sorting with the Magnetic Separation kit CELLection Pan Mouse IgG kit from Dynabeads®, performed according to the manufacturers’ instructions.

Assessment of cell survival upon drug treatments

MKN74_Mock, MKN74_CD44v6 and MKN74_CD44s cells were seeded in 96-well plates ($2.5x10^3$ per well) under normal conditions (5% CO_2 humidified atmosphere at 37 ºC) and allowed to adhere for 24 h. Cells were then incubated with different concentrations of cisplatin or 5-FU for 48 h and processed for the Sulforhodamine B (SRB) assay. Briefly, at the required times, cells were fixed in 10% trichloroacetic acid (TCA) for 1 h on ice, proteins stained with 4% SRB solution (Sigma-Aldrich; Poole, UK) for 30 min, wells washed repeatedly with 1% acetic acid to remove the unbound dye, and the protein-bound stain was solubilized with 10 mM Tris solution. The SRB absorbance was measured at 560 nm and background corrected at 655 nm, using a microplate reader (PowerWave HT Microplate Spectrophotometer; BioTek, Bad Friedrichshall, Germany). At least three independent experiments were performed, each measured in triplicate. Cell survival for each drug treatment was calculated, as a percentage, in relation to the respective vehicle-treated control, for each cell line.

Assessment of cisplatin-induced apoptosis

MKN74_Mock, MKN74_CD44v6 and MKN74_CD44s cells were seeded in 6-well plates ($1.0x10^5$ cells/ well), left to adhere for 24 h, and incubated with 10 μM of cisplatin for 24 and 48 h. Cisplatin-induced apoptosis was assayed by labelling cells with Annexin V-APC antibody (ImmunoTools, Friesoythe, Germany), according to the manufacturer's instructions. Measurement of phosphatidylserine externalization was analyzed using an Accuri C6 flow cytometer and Accuri C6 software (BD Biosciences, Franklin Lakes, NJ, USA), plotting at least 20,000

events per sample. Results represent the average of at least three independent experiments.

CD44v6 expression inhibition by siRNA

GP202 and MKN45 cells were transfected with siRNAs for CD44v6 using Lipofectamine® RNAiMax Transfection Reagent, according to the manufacturers' instructions. Briefly, cells were seeded in 6 well-plates ($2x10^5$ cells/well and $1.5x10^5$ cells/well, respectively GP202 and MKN45 cell lines). After 24 h, lipid based conjugates were prepared by mixing 10 nM of non-targeting siRNA (negative control DS NC1; iDT, Leuven, Belgium), or human CD44v6 siRNA (Sense strand: 5'- GCGUCAGGUUCCAUAGGAAUCCUTT - 3' and Antisense strand: 5'- AAAGGAUUCCUAUGGAACCUGACGCAG - 3', custom made from iDT, Leuven, Belgium), to diluted Lipofectamine® RNAiMax Reagent in 1:1 ratio. The conjugates were incubated for 5 min at room temperature and added dropwise to the cells. Upon 48 h of incubation, protein extraction and Western blots were performed to evaluate the efficacy of CD44v6 silencing. In addition, *in vitro* cisplatin treatments were carried out and apoptosis assessed (as described above).

Percentage of CD44v6+ cells upon cisplatin incubation

MKN74_Mock and MKN74_CD44v6 were mixed and seeded in 6 well plates with an initial density of $5x10^4$ cells/well (50:50 proportion), allowed to grow for 24 h and incubated with 10 μM of cisplatin or vehicle (0.9% (v/v) NaCl) for 6 h. Cells were then washed with PBS, growth medium was replaced with cisplatin-free media and co-cultures were maintained during 15 days. CD44v6 cell

enrichment was assessed at several time points by flow cytometry using a CD44v6 conjugated antibody. Briefly, at each time point, cells were washed with PBS, versinized for 15 min, blocked in PBS - 2% FBS for 30 min and incubated with CD44v6-APC conjugated antibody (2F10; 10 µL/106 cells; 30 min; R&D Systems, McKinley Place, MN, USA). Cells were subsequently washed with PBS - 2% FBS and immediately proceeded for analysis.

Fluorescence was analyzed using a FACS Canto II (BD Biosciences, Franklin Lakes, NJ, USA). The mean fluorescence intensity was measured for at least 20,000 gated events per sample and data analyzed using the software FlowJo, version 10.

Statistical Analysis

Age, gender and patients' clinical-pathological features were compared according to CD44v6 expression and treatment type. Univariate analyses for categorical variables were performed using Chi-square or Fisher-exact test as appropriate. Continuous variables were analyzed using Student's t-test or One-way ANOVA. For the statistical analyses of experiments in Fig 2 a Two-way ANOVA was performed. Post-hoc tests were performed for significant One- or Two-way ANOVA results using Tukey's Post Hoc Test.

Kaplan-Meier estimates of Overall Survival (OS) and Relapse-Free Survival (RFS) were obtained between the groups. Afterwards, Multivariate Cox regression analysis of OS and RFS were performed, and the hazard ratio (HR) and 95% CI estimated, in order to determine factors that were independently associated with RFS and OS. A *P*-value < .05 was considered significant and all

analyses were performed using IBM SPSS Statistics version 24 for Windows. This study is TRIPOD compliant.

Results:

Consolidation of CD44v6 *de novo* expression in tumor cells as a poor prognosis marker in GC patients

Data on clinical-pathological features and treatment type (Appendix Table 1), OS and RFS was collected from a surgical series of 334 GC patients. This series has a good representation of all disease stages and, as expected, patient OS and RFS significantly worsens with increasing pTNM (Pathological Tumor-Node-Metastasis) staging (Appendix Fig 2A and B). Analysis of OS and RFS in chemotherapy treated *vs.* untreated stage II to IV patients, demonstrated that chemotherapy plus surgery increases OS ($P < .005$), although no advantage was observed for RFS (Appendix Fig 3A and B).

The extent of CD44v6 membranous expression was analyzed in all cases and tumors were classified into four sub-categories, from absent to high (0, 1+, 2+ and 3+; Fig 1A): 46% (155/334) were CD44v6_0; 15% (50/334) were CD44v6_1+; 25% (82/334) were CD44v6_2+, and; 14% (47/334) were CD44v6_3+. We found a strong association between high CD44v6 expression and poorer OS (Fig 1B). Indeed, GC patients whose tumors express CD44v6 in >50% of tumor cells (CD44v6_3+), have statistically significant worse OS than those lacking CD44v6 expression or if CD44v6 is present in <50% of tumor cells ($P < .001$) (Fig 1B). Therefore, it seemed reasonable to group patients into two expression categories: “CD44v6_absent/low” (combining 0, 1+ and 2+

cases) and “CD44v6_high” (3+ cases). GC patients whose tumors were classified as “CD44v6_high” had significantly worse OS than those with absent/low CD44v6 expression (median OS ~ 15 months vs 51 months, *P* < .0001; Fig 1C). When comparing the clinical-pathological features of patients bearing CD44v6_absent/low with those bearing CD44v6_high tumors, we found that CD44v6_high patients more often presented vascular invasion (*P* < .005), tumor cells in the surgical margins (*P* < .01) and perineural invasion (*P* < .05) (Table 1). Multivariate analysis further demonstrated that CD44v6_high was an independent factor of poor prognosis in this series (Fig 1D). Additionally, and somehow expected, high pTNM stage (*P* < .0001*)*, older age at onset (*P* < .0001*)* and presence of tumor cells in surgical margins (*P* = .003) were also identified as independent poor prognosis factors in this series.

In summary, we identified a clear cut-off for defining overexpression of CD44v6 in GC (>50% tumor cells overexpressing membranous CD44v6), which demonstrates prognostic value, likely due to greater ability of CD44v6_high cells to access the vasculature and perineural space, and to spread throughout the stomach wall.

Adding chemotherapy to surgery benefits particularly patients presenting CD44v6_high tumors

Given that patients bearing “CD44v6_high” tumors present worse prognosis, we hypothesize that they may also respond worse to cisplatin-based therapy. Indeed, we observed that “CD44v6-high” patients have worse median OS regardless of receiving chemotherapy in addition to surgery or not (Figs 2A and B). However, administration of chemotherapy to CD44v6_high patients resulted

in a 6-fold increase in their median OS, from 4 to 24 months (Figs 2A and B). Although not as striking, the median OS of patients with CD44_absent/low tumors almost doubled (from 20 to 38 months) when treated with conventional chemotherapy in addition to surgery (Figs 2A and B). This benefit was mainly associated with patients with CD44v6_2+ (Fig 2E). Overall, these data support chemotherapy not only in CD44v6-high (CD44v6_3+) patients, but also in CD44v6_2+ patients, and fails to support our initial hypothesis that the poor prognosis of CD44v6-high patients is related to weak response to chemotherapy. Indeed, we further verified that the greater the percentage of tumor cells expressing CD44v6, the greater the treatment benefit reflected in the median OS (Figs 2C-G). For instance, CD44v6_1+, 2+ and 3+ patients had a 2.6 (not statistically significant), 3.8 and 6-fold improvement in OS when treated with chemotherapy in addition to surgery, respectively (P = .004, Figs 2D-F). In contrast, no OS improvement was observed when administering chemotherapy to CD44v6 negative patients (Fig 2C).

Although the extent of CD44v6 expression does not seem to affect RFS (Appendix Figs 4A-G), the mean time to relapse decreases with increasing percentage of CD44v6+ cells in the tumor (~28, 21 and 15 months for CD44v6 1+, 2+ and 3+, respectively; Appendix Fig 4H) in patients treated with surgery and chemotherapy. This is true after chemotherapy, when considering only the sub-set of patients that relapsed. These results suggest that CD44v6+ cells in the tumor are likely involved in triggering a faster tumor relapse following chemotherapy. In patients from the CD44v6_0 group, chemotherapy seems to decrease time to relapse as compared with surgery alone (from 19 to 13 months), unlike what occurs in the CD44v6+ groups (Appendix Fig 4H).

These data raise the importance of considering the CD44v6 sub-categorization when selecting patients for chemotherapy, as: (1) chemotherapy significantly improved OS and time to relapse specifically in CD44v6_2+ and 3+ patients, and; (2) chemotherapy does not influence OS and decreases time to relapse in CD44v6 negative (CD44v6_0) patients.

CD44v6 overexpression in GC cell lines influences response to cisplatin treatment

We next wanted to understand how does CD44v6 overexpression influences response to conventional chemotherapeutic agents used in GC treatment, namely cisplatin and 5-fluorouracil (5-FU). For this, we generated three isogenic GC cell lines: mock cells lacking CD44, CD44s and CD44v6 expressing cells (Appendix Fig 5).

While no differences were observed between the isogenic cells for doubling-time, invasion, cell-adhesion and motility, tumorigenic potential *in vivo* (Appendix Fig 6); CD44v6 overexpressing cells survived better to cisplatin treatment, by decreasing apoptosis levels (Fig 3A and C), but not to 5-FU treatment (Fig 3B). To validate these findings, we depleted CD44v6 expression from cells endogenously expressing CD44v6 (Appendix Fig 7). Indeed, CD44v6 depletion increased cisplatin-induced apoptosis (Fig 3D), validating our results. Since some CD44v6 effectors are also implicated in cisplatin resistance, namely STAT3 and P38, we evaluated their modulation in our experimental models. STAT3 was rapidly activated in CD44v6 overexpressing cells in response to cisplatin (Appendix Fig 8). Although this might explain the increased cell survival in CD44v6 overexpressing cells, this behavior was also observed in

CD44s cells. This data was not further clarified neither when analyzing P38 expression in the isogenic model nor in cell lines endogenously expressing CD44v6 (data not shown).

Altogether, these *in vitro* studies clearly and consistently indicated that CD44v6 overexpression increases survival of GC cells in response to cisplatin treatment.

CD44v6 expressing cells may contribute to patient relapse after chemotherapy

We observed that patients' tumors with greater CD44v6+ cell populations relapse faster, following chemotherapy than those with small fractions of this sub-population (Appendix Fig 4H). In addition, our *in vitro* results show that CD44v6+ cells survive better than CD44v6- cells in response to cisplatin (included in the chemotherapeutic regimen of most patients studied; Appendix Table 2). So, we hypothesized that upon chemotherapy, CD44v6+ cells prevail over the CD44v6- counterpart, triggering faster tumor relapse and dominating recurring lesions. We therefore analyzed the CD44v6 *status* in surgically-resected tumors from patients treated with neo-adjuvant chemotherapy or chemo-naïve. Supporting our hypothesis, 50% of tumors treated prior to surgery were CD44v6_3+, comparing to 25% of chemo-naïve tumors (Fig 4A). Despite the low number of patients, these results suggest that chemotherapy may increase the abundance of CD44v6+ cell population in gastric cancers.

We tested this hypothesis *in vitro* by incubating co-cultures of CD44v6+ and CD44v6- cells with cisplatin and subsequently allowing them to recover for 15 days (as an attempt to mimic what happens in patients that receive chemotherapy and then present tumor relapse). We showed that upon

recovery, CD44v6+ cells represent >80% of the cell culture, suggesting that CD44v6 is involved in GC cell overgrowth after cisplatin treatment (Figs 4B and C).

Discussion:

Herein, we established a protocol to evaluate CD44v6 in GC, according to the percentage of tumor cells expressing membranous CD44v6. The categorization of CD44v6 expression was further refined through correlation with OS of GC patients, and allowed demonstrating that in GC, CD44v6 is an independent marker of poor prognosis with further potential as a predictive marker of response to conventional chemotherapy. Our data also shed light into GC molecular heterogeneity and its relation with therapy, by suggesting that the CD44v6+ population may be driving recurrence following conventional chemotherapy, in tumors composed of CD44v6+ and CD44v6- cell populations. Approximately 14% of GC patients present with CD44v6_high tumors (Fig 5A), which have a significantly worse median OS compared with CD44v6_absent/low patients. Importantly, this occurs independently of TNM staging (Appendix Fig 9A-D), as demonstrated by multivariate analysis. Studies so far reporting an association between CD44v6 expression and worse GC patient survival,[18,19] enclosed low patients' number and failed to demonstrate prognostic value for CD44v6 expression independently of TNM staging. Our study addresses this issue in the largest GC patient cohort tested so far, and demonstrates that CD44v6 is truly an independent marker of poor prognosis, as it has been described for colorectal cancer.[25]

We further demonstrated that CD44v6 expression in tumors predicts benefit from application of conventional chemotherapy in addition to surgery. This benefit is directly proportional to the fraction of CD44v6+ cells in tumors, with patients whose tumors lack CD44v6 expression not benefiting from chemotherapy treatment (Fig 5B-5D). The latter also appear to have worse RFS and to relapse sooner than CD44v6_0 patients not receiving chemotherapy in addition to surgery. To the best of our knowledge, this is the first report demonstrating CD44v6 as a useful marker to predict therapy response in GC, similarly to what was shown in colorectal cancer, where patients with moderate or strong CD44v6 tumor expression respond better to chemotherapy with irinotecan.[26]

Nearly half of the tumors herein studied possess some degree of heterogeneity regarding CD44v6 expression (with CD44v6+ and CD44v6- cell populations co-existing in the same tumor), which may have important consequences in terms of clinical outcome. Our *in vitro* data shows that CD44v6 expression promotes survival of GC cells after cisplatin treatment, as others have shown for several tumor models,[15,23,24] and in co-cultures of CD44v6+ and CD44v6– cells. Our patient data also suggests enrichment in the CD44v6+ cell population after neo-adjuvant treatment and in relapses, but validation in larger and/or independent GC patient series is still needed.

CD44/CD44v are considered cancer stem cell markers in several tumor types, including GC,[13,14] supporting our hypothesis that CD44v6+ tumor cells may be potentiating relapse. If our findings are confirmed, strategies to specifically eliminate CD44v6+ cells could lead to decreased recurrence and improved patient survival. One possible strategy is the targeting of CD44v6 downstream

effectors, as direct targeting of CD44v6 ought to be avoided due to the lethal skin toxicity side effects described in a Phase I clinical trial using a highly potent antimicrotubule agent coupled to a monoclonal antibody against CD44v6.[27] TNM staging has long been the most important tool to assess prognosis in GC. Our data shows that evaluating CD44v6 by IHC may provide oncologists with additional and important information for stratifying GC patient for treatment. Namely, if CD44v6 IHC is performed in GC biopsies and turns out CD44v6_high, this information may be used to recommend surgery combined with chemotherapy. This would be particularly important for patients with CD44v6 2+ or 3+ tumors diagnosed in stage I and stage II, who often do not undergo chemotherapy. Moreover, finding tumors lacking CD44v6 expression could help oncologists selecting patients less likely to benefit from chemotherapy, hence avoiding its related toxicity side effects (Fig 5E).

Conclusion

Our study is the first one to identify CD44v6 as a potential predictive marker of response to conventional chemotherapy. In addition, we provide compelling evidence that consolidates CD44v6 is an independent marker of poor prognosis, and a likely marker of faster relapse. Importantly, our data supports selection of patients with high CD44v6 expressing tumors for conventional chemotherapy in addition to surgery, even in TNM stage I and stage II patients, who are generally treated with surgery alone.

Acknowledgements:

This work was supported 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 – 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-07-0124-FEDER-000029 - supported by Norte Portugal Regional Programme (NORTE 2020), under the PORTUGAL 2020 Partnership Agreement, through the European Regional Development Fund (ERDF) - and project PTDC/CTM-NAN/120958/2010, from FCT.

Carla Pereira was supported by the grant SFRH/BD/113031/2015 and Daniel Ferreira by the grant PD/BD/105976/2014.

Gabriela M. Almeida was supported by the Investigator FCT Program 2013 (IF/00615/2013), POPH - QREN Type 4.2, European Social Fund and Portuguese Ministry of Science and Technology (MCTES).

References

[1] Bray F, Ferlay J, Soerjomataram I, et al: Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. CA Cancer J Clin 10.3322/caac.21492. [Epub ahead of print], 2018

[2] Oba K, Paoletti X, Bang YJ, et al: Role of chemotherapy for advanced/recurrent gastric cancer: an individual-patient-data meta-analysis. Eur J Cancer 49:1565-1577, 2013

[3] Cervantes A, Roda D, Tarazona N, et al: Current questions for the treatment of advanced gastric cancer. Cancer Treat Rev 39:60-67, 2013

[4] Durães C, Almeida GM, Seruca R, et al: Biomarkers for gastric cancer: prognostic, predictive or targets of therapy? Virchows Arch 464:367-378, 2014

[5] Bang YJ, Van Cutsem E, Feyereislova A, et al: 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, 2010

[6] Fuchs CS, Tomasek J, Yong CJ, et al: Ramucirumab monotherapy for previously treated advanced gastric or gastro-oesophageal junction adenocarcinoma (REGARD): an international, randomised, multicentre, placebo-controlled, phase 3 trial. Lancet 383:31-39, 2014

[7] Ponta H, Wainwright D, Herrlich P: The CD44 protein family. Int J Biochem Cell Biol 30: 299-305, 1998

[8] Ponta H, Sherman L, Herrlich PA: CD44: from adhesion molecules to signalling regulators. Nat Rev Mol Cell Biol 4: 33-45, 2003

[9] Prochazka L, Tesarik R, Turanek J: Regulation of alternative splicing of CD44 in cancer. Cell Signal 26:2234-2239, 2014

[10] Ruiz P, Schwarzler C, Gunthert U: CD44 Isoforms during Differentiation and Development. Bioessays 17:17-24, 1995

[11] Naor D, Wallach-Dayan SB, Zahalka MA, et al: Involvement of CD44, a molecule with a thousand faces, in cancer dissemination. Seminars in Cancer Biology 18: 260-267, 2008

[12] Senbanjo LT, Chellaiah MA: CD44: A Multifunctional Cell Surface Adhesion Receptor Is a Regulator of Progression and Metastasis of Cancer Cells. Front Cell Dev Biol 5:18, 2017

[13] Yan Y, Zuo X, Wei D: Concise Review: Emerging Role of CD44 in Cancer Stem Cells: A Promising Biomarker and Therapeutic Target. Stem Cells Transl Med 4:1033-1043, 2015

[14] Todaro M, Gaggianesi M, Catalano V, et al: CD44v6 is a marker of constitutive and reprogrammed cancer stem cells driving colon cancer metastasis. Cell Stem Cell 14:342-356, 2014

[15] Zöller M: CD44, Hyaluronan, the Hematopoietic Stem Cell, and Leukemia-Initiating Cells. Front Immunol 6:235, 2015

[16] da Cunha CB, Oliveira C, Wen X, et al: De novo expression of CD44 variants in sporadic and hereditary gastric cancer. Lab Invest 90:1604-1614, 2010

[17] Chen Y, Fu Z, Xu S, et al: The prognostic value of CD44 expression in gastric cancer: a meta-analysis. Biomed Pharmacother 68:693-697, 2014

[18] Xie JW, Chen PC, Zheng CH, et al. Evaluation of the prognostic value and functional roles of CD44v6 in gastric cancer. J Cancer Res Clin Oncol 141:1809-1817, 2015

[19] Xu YY, Guo M, Yang LQ, et al: Regulation of CD44v6 expression in gastric carcinoma by the IL-6/STAT3 signaling pathway and its clinical significance. Oncotarget 8:45848-45861, 2017

[20] Rodrigues M, Vitó I, Santos R, et al: Establishment of a Tumour Bank: the experience of the Department of Pathology of Hospital S. João (Porto, Portugal). Cell Tissue Bank 10:75-77, 2009

[21] Gärtner F, David L, Seruca R, et al: Establishment and characterization of two cell lines derived from human diffuse gastric carcinomas xenografted in nude mice. Virchows Arch 428:91-98, 1996

[22] Bordeira-Carrico R, Ferreira D, Mateus DD, et al: Rescue of wild-type E-cadherin expression from nonsense-mutated cancer cells by a suppressor-tRNA. Eur J Hum Genet 22:1085-1092, 2014

[23] Lv L, Liu HG, Dong SY, et al: Upregulation of CD44v6 contributes to acquired chemoresistance via the modulation of autophagy in colon cancer SW480 cells. Tumour Biol 37:8811-8824, 2016

[24] Mooney KL, Choy W, Sidhu S, et al: The role of CD44 in glioblastoma multiforme. J Clin Neurosci 34:1-5, 2016

[25] Wang JL, Su WY, Lin YW, et al: CD44v6 overexpression related to metastasis and poor prognosis of colorectal cancer: A meta-analysis. Oncotarget 8:12866-12876, 2017

[26] Bendardaf R, Lamlum H, Ristamäki R, et al: CD44 variant 6 expression predicts response to treatment in advanced colorectal cancer. Oncol Rep 11:41-45, 2004

[27] Riechelmann H, Sauter A, Golze W, et al: Phase I trial with the CD44v6-targeting immunoconjugate bivatuzumab mertansine in head and neck squamous cell carcinoma. Oral Oncol 44:823-829, 2008

[28] Lordick F, Allum W, Carneiro F, et al: Unmet needs and challenges in gastric cancer: the way forward. Cancer Treat Rev 40:692-700, 2014

Figure Legends:

Fig 1: (A) Immunohistochemistry characterization of samples regarding the presence and extent of CD44v6 expression. Images are representative of the four categories defined: CD44v6_0 (no staining at the cell membrane); CD44v6_1+ (membranous staining in < 10% of tumor cells); CD44v6_2+ (membranous staining in ≥ 10% and < 50% of tumor cells); CD44v6_3+ (membranous staining in ≥ 50% of tumor cells); **(B)** Kaplan-Meier estimates showing OS of GC patients according to CD44v6 sub-categories and corresponding median OS - CD44v6_3+ patients have worse OS compared to other sub-categories ($P < .001$); **(C)** Kaplan-Meier estimates showing OS of GC patients according to absent/low and high expression of CD44v6 and corresponding median OS – "CD44v6_High" patients have the worst prognosis ($P < .0001$); **(D)** Forest plot of the multivariate analysis. Expression of CD44v6 in more than 50% of tumor cells was identified as an independent factor of poor prognosis in GC.

Fig 2: (A) Kaplan-Meier estimates showing OS of patients treated with surgery alone or **(B)** surgery plus chemotherapy, according to CD44v6 status; **(C)** Kaplan-Meier estimates showing OS of CD44v6_0 **(D)** CD44v6_1+, **(E)** CD44v6_2+ and **(F)** CD44v6_3+ GC patients, according to whether they were treated with surgery alone or with surgery plus chemotherapy; **(G)** Forest plot showing the benefit of treating patients with surgery plus chemotherapy according to CD44v6 classification.

Fig 3: Assessing the response of the isogenic MKN74 cells to conventionally used chemotherapeutic agents: **(A)** % Cell survival upon incubation with cisplatin or **(B)** 5-FU for 48 h (compared to vehicle control) in MKN74 cells; **(C)** % Apoptotic cells in MKN74 cells incubated with 10 µM cisplatin or vehicle (0.9% NaCl) for 48 h; **(D)** % Apoptotic cells in GP202 and MKN45 cell lines in response to 48 h treatment with cisplatin or vehicle, following a 24 h incubation with CD44v6 or scramble siRNAs. The cisplatin concentrations used are considered to be clinically relevant. Results are expressed as the average + SD of at least three independent experiments. Statistically significant results was determined by Two-way ANOVA with Tukey's multiple comparisons test (* $P <$ .05; ** $P <$ .001; **** $P <$ .0001).

Fig 4: In heterogeneous tumors, CD44v6+ cells may be enriched following treatment with chemotherapeutic agents **(A)** frequency of CD44v6_low (comprising CD44v6 sub-categories 1+ and 2+) and CD44v6_high in chemo-naïve (comprising both tumors from patients that did not receive chemotherapy in addition to surgery and also patients that only received chemotherapy following tumor resection); **(B)** % of CD44v6+ cells in co-cultures of CD44v6+ and CD44v6- MKN74 cells (MKN74_CD44v6 and MKN74_Mock cells, respectively) following a 6 hour incubation with cisplatin (or vehicle control) and up to 15 days of recovery. At the end of the experiment > 80% of the cell culture is composed of CD44v6+ cells; **(C)** Scheme depicting the cell enrichment experiment.

Fig 5: (A) Classification of GC in sub-categories of CD44v6 expression and % of cases per sub-category; **(B)** 5-year survival of patients treated with surgery alone and with surgery plus chemotherapy, per CD44v6 sub-category; **(C)** median OS of patients treated with surgery alone and with surgery plus chemotherapy, per CD44v6 sub-category; **(D)** GC patient survival benefit of adding chemotherapy to surgery, expressed as fold change (FC) between the OS of patients treated with surgery + chemotherapy with the OS of patients treated with surgery alone; **(E)** proposed scheme for improving GC patient stratification using CD44v6 as a marker for treatment selection (adapted from [28]). The obtained data show that patients with operable >T1N0 and CD44v6+ tumors benefit from receiving chemotherapy in addition to surgery. In addition, our data suggests patients with T1N0 and CD44v6+ tumors may benefit from additional chemotherapy, and that patients with inoperable or metastatic GC with CD44v6+ tumors may benefit from more aggressive chemotherapeutic strategies, highlighting the importance of performing additional studies on this issue.

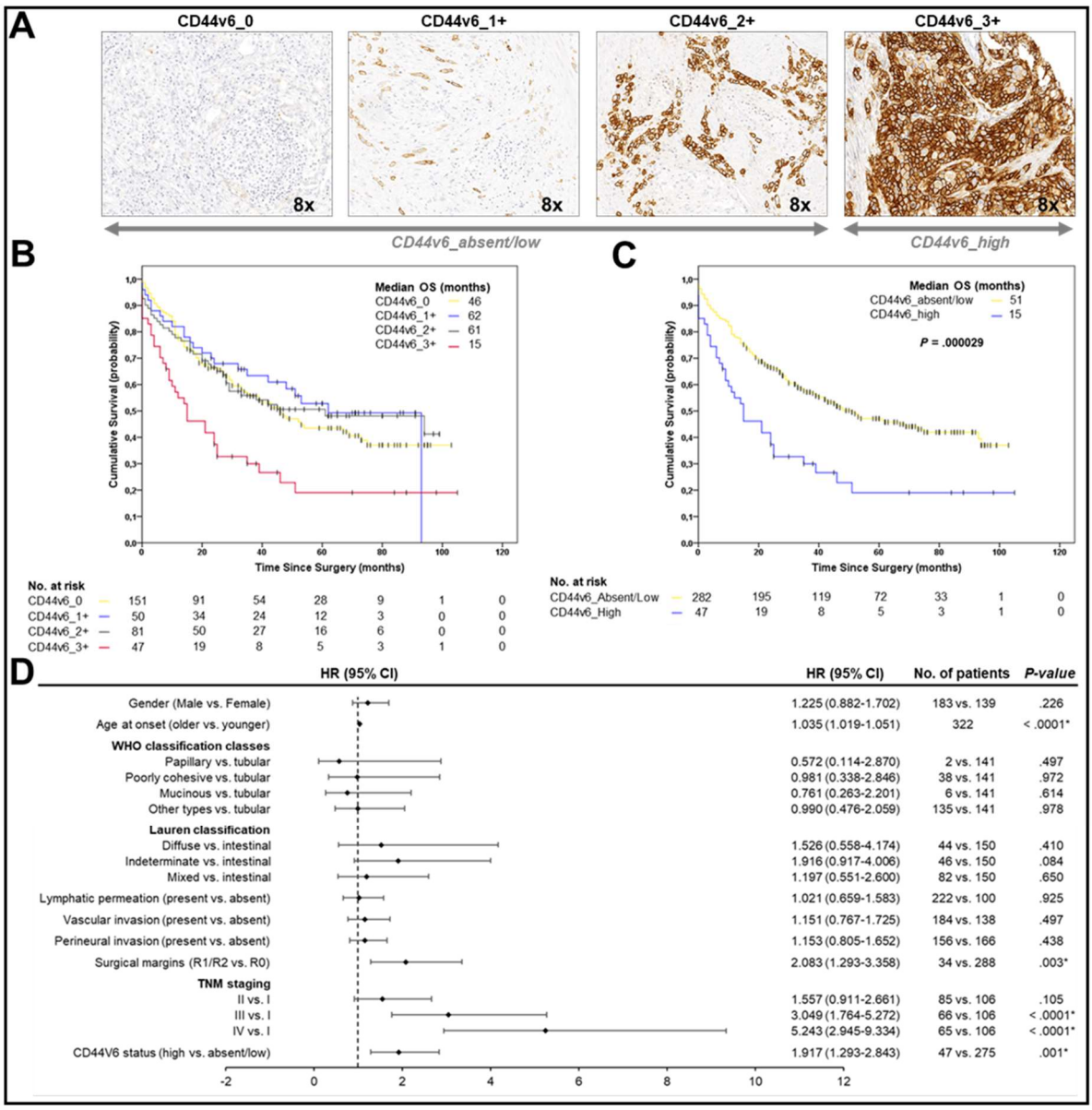


Fig 1.

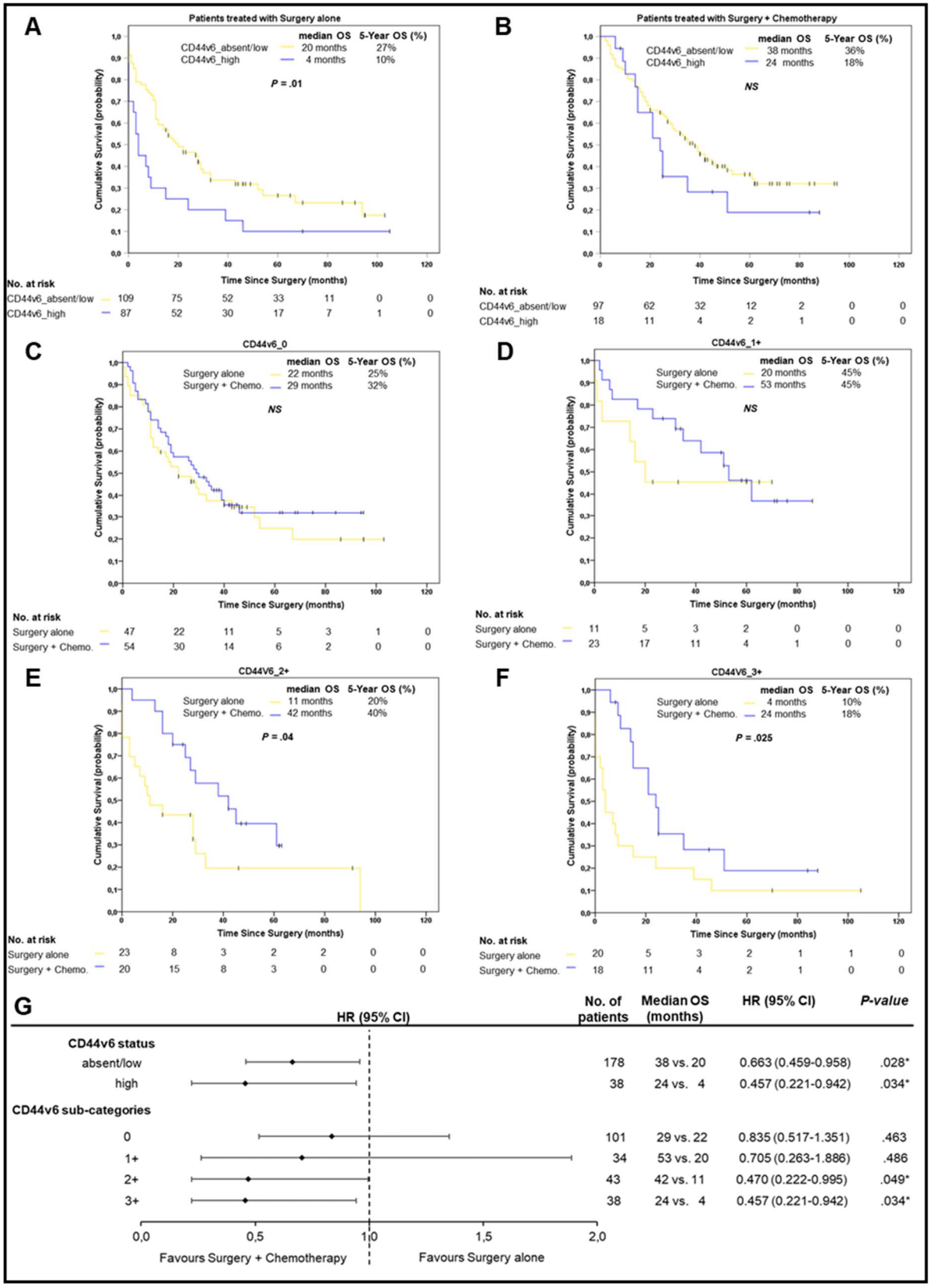


Fig 2.

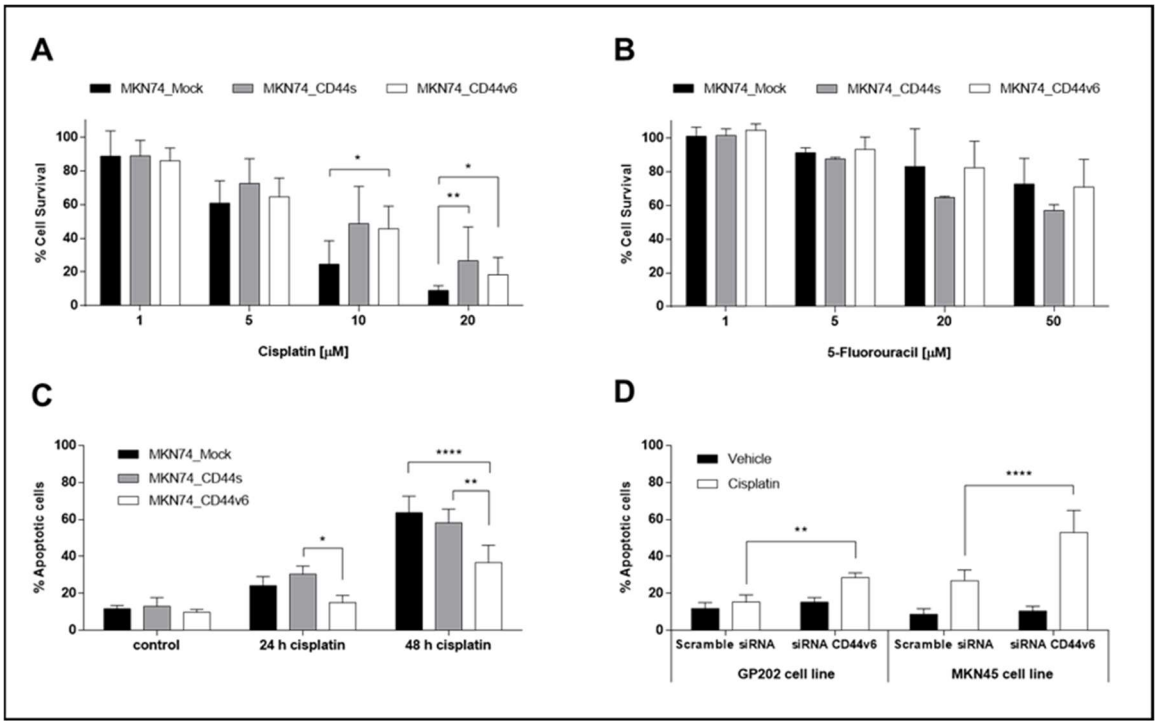


Fig 3.

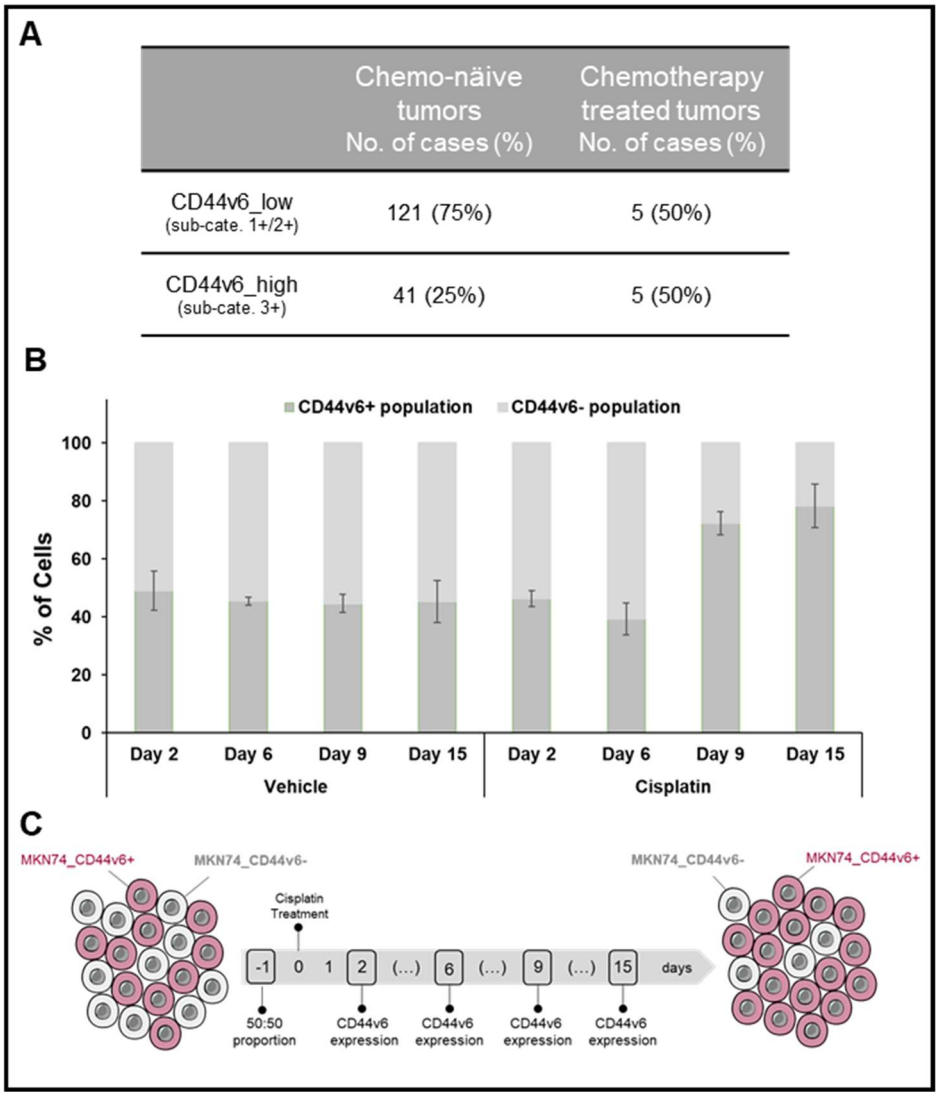


Fig 4.

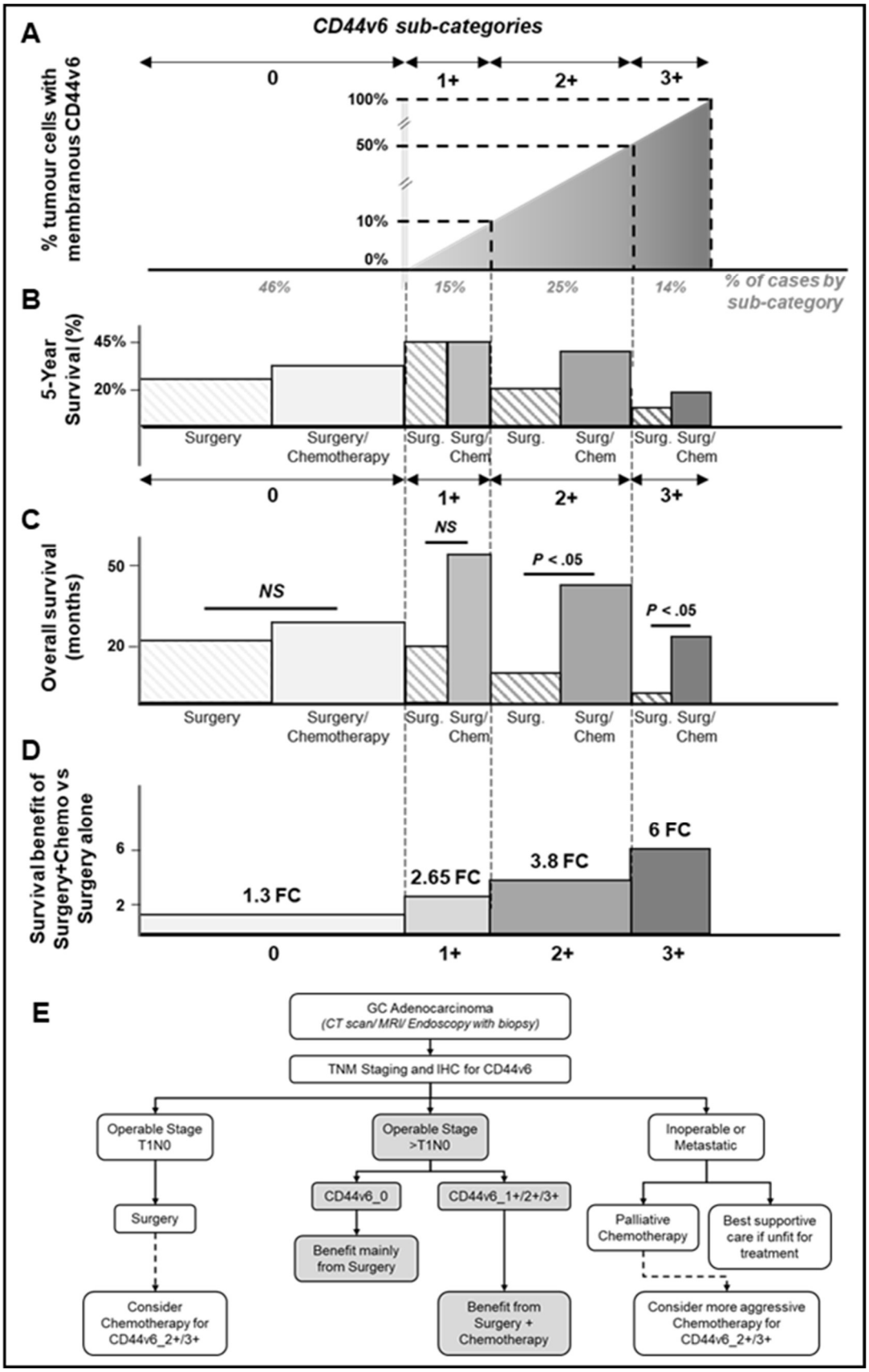


Fig 5.

Table 1: Clinical-pathological associations with extent of CD44v6 expression (absent/low *vs.* high) in gastric tumors.

Variables	Total No. patients (n= 334)	CD44v6 absent/low (n= 287)	CD44v6 high (n= 47)	*P-value*
Age (years)				> .05
Mean	67.6	67.3	69.3	
Standard deviation	*11.9*	*11.9*	*11.9*	
Gender				> .05
Male	188 *(56.3%)*	167 *(58.2%)*	21 *(44.7%)*	
Female	146 *(43.7%)*	120 *(41.8%)*	26 *(55.3%)*	
Male:Female ratio	1.3:1	1.4:1	0.8:1	
Lauren classification				> .05
Intestinal	156 *(46.7%)*	139 *(48.4%)*	17 *(36.2%)*	
Diffuse	44 *(13.2%)*	39 *(13.6%)*	5 *(10.6%)*	
Mixed	86 *(25.7%)*	68 *(23.7%)*	18 *(38.3%)*	
Indeterminate	48 *(14.4%)*	41 *(14.3%)*	7 *(14.9%)*	
WHO classification				> .05
Tubular	146 *(43.7%)*	129 *(44.9%)*	17 *(36.2%)*	
Papillary	2 *(0.6%)*	2 *(0.7%)*	0 *(0.0%)*	
Mucinous	7 *(2.1%)*	7 *(2.4%)*	0 *(0.0%)*	
Poorly cohesive	38 *(11.4%)*	32 *(11.1%)*	6 *(12.8%)*	
Other types	141 *(42.2%)*	117 *(40.8%)*	24 *(51.1%)*	
Growth pattern				> .05
Expansive	61 *(18.3%)*	55 *(19.2%)*	6 *(12.8%)*	
Infiltrative	259 *(77.5%)*	220 *(76.7%)*	39 *(83.0%)*	
Unclassified	14 *(4.2%)*	12 *(4.2%)*	2 *(4.3%)*	
Wall invasion				> .05
Mucosa + Submucosa	79 *(23.7%)*	74 *(25.8%)*	5 *(10.6%)*	
Muscular	42 *(12.6%)*	36 *(12.5%)*	6 *(12.8%)*	
Subserosa + Serosa	201 *(60.2%)*	167 *(58.2%)*	34 *(72.3%)*	
Other organs	12 *(3.6%)*	10 *(3.5%)*	2 *(4.3%)*	
Lymphatic permeation	332*1			> .05
Absent	103 *(31.0%)*	94 *(33.0%)*	9 *(19.1%)*	
Present	229 *(69.0%)*	191 *(67.0%)*	38 *(80.9%)*	
Perineural invasion	333*1			**.043**
Absent	173 *(52.0%)*	155 *(54.2%)*	18 *(38.3%)*	
Present	160 *(48.0%)*	131 *(45.8%)*	29 *(61.7%)*	
Vascular invasion	331*1			**.002**
Absent	138 *(41.7%)*	128 *(45.1%)*	10 *(21.3%)*	
Present	193 *(58.3%)*	156 *(54.9%)*	37 *(78.7%)*	
Surgical margins	333*1			**.009**
R0	298 *(89.5%)*	261 *(91.3%)*	37 *(78.7%)*	
R1/R2	35 *(10.5%)*	25 *(8.7%)*	10 *(21.3%)*	
Depth of invasion (T)				> .05
pT1	79 *(23.7%)*	74 *(25.8%)*	5 *(10.6%)*	
pT2	79 *(23.7%)*	68 *(23.7%)*	11 *(23.4%)*	
pT3-T4	176 *(52.7%)*	145 *(50.5%)*	31 *(66.0%)*	
Lymph node metastases (N)	333*1			> .05
Absent (pN0)	129 *(38.7%)*	116 *(40.6%)*	13 *(27.7%)*	
Present (pN+)	204 *(61.3%)*	170 *(59.4%)*	34 *(72.3%)*	
Distant metastases (M)				> .05
Absent	265 *(79.3%)*	229 *(79.8%)*	36 *(76.6%)*	
Present	69 *(20.7%)*	58 *(20.2%)*	11 *(23.4%)*	
pTNM Staging				> .05
I	109 *(32.6%)*	100 *(34.8%)*	9 *(19.1%)*	
II	87 *(26.0%)*	72 *(25.1%)*	15 *(31.9%)*	
III	69 *(20.7%)*	57 *(19.9%)*	12 *(25.5%)*	
IV	69 *(20.7%)*	58 *(20.2%)*	11 *(23.4%)*	
Disease Recurrence	265*2			> .05
Absent	192 *(72.5%)*	168 *(73.4%)*	24 *(66.6%)*	
Present	73 *(27.5%)*	61 *(26.6%)*	12 *(33.3%)*	
Treatment Regimen	326*1			> .05
Surgery alone	203 *(62.3%)*	175 *(62.7%)*	28 *(59.6%)*	
Pt-based therapy	113 *(34.7%)*	94 *(33.7%)*	19 *(40.4%)*	
Non-Pt based therapy	10 *(3.1%)*	10 *(3.6%)*	0 *(0.0%)*	

*[1] Remaining data not available; *[2] Stage IV patients were excluded. pTNM (pathological tumor-node-metastasis).

Appendix B

The cast of characters behind E-cadherin dysfunction in gastric cancer

Carla Pereira, Marta Ferreira, Diana Martins, Patricia Oliveira, Fátima Carneiro, Gabriela M. Almeida, Carla Oliveira

This work was presented in two scientific meetings:

8th i3S Annual Meeting, Póvoa do Varzim (Portugal)

23rd Annual Meeting of Portuguese Society of Human Genetics, Coimbra (Portugal)

Poster illustrating the preliminary work presented in Original Paper 5