AVALIAÇÃO DO MICROAMBIENTE TUMORAL EM ADENOCARCINOMAS GÁSTRICOS A PARTIR DE DADOS DE EXPRESSÃO GÊNICA

JAQUELINE RAMALHO BUTTURA

Tese apresentada à Fundação Antônio Prudente para a obtenção do título de Doutor em Ciências.

Área de concentração: Oncologia

Orientador: Dr. Israel Tojal da Silva, PhD
Coorientador: Dr. Kenneth John Gollob, PhD

São Paulo

2021

FICHA CATALOGRÁFICA

Buttura, Jaqueline Ramalho

Avaliação do microambiente tumoral em adenocarcinomas gástricos a partir de dados de expressão gênica/ Jaqueline Ramalho Buttura - São Paulo, 2021.

137p.

Tese (Doutorado)-Fundação Antônio Prudente.

Curso de Pós-Graduação em Ciências - Área de concentração: Oncologia.

Orientador: Israel Tojal da Silva

Descritores: 1. Adenocarcinoma gástrico/Gastric cancer. 2. Microambiente tumoral/Tumoral microenvironment. 3. Infiltrado inflamatório/Inflammatory infiltrate 4. Assinaturas mutacionais/Mutational signatures. 5.Imunoterapia/Immunotherapy. 6. Prognóstico/Prognostic.

Talvez não tenha conseguido fazer o melhor, mas lutei para que o melhor fosse feito. Não sou o que deveria ser, mas Graças a Deus, não sou o que era antes.

Marthin Luther King.

DEDICATÓRIA

A meus pais, Waldir e Roselei

e ao meu marido Renato.

AGRADECIMENTOS

Primeiramente, agradeço a Deus por essa oportunidade, pelas pessoas que colocou no meu caminho e por me guiar sempre. A experiência desses quatro anos resultou em muito mais além deste trabalho, pois foi um crescimento pessoal enorme.

Agradeço pela minha família, meus pais Waldir e Roselei, meus irmãos Gisele e Júlio, e meu marido Renato e meus sogros, Rita e José Carlos, que todos os dias me apoiam e me incentivam, e torcem para que tudo se encaminhe e dê certo. A família com certeza é a base: muita gratidão!

Agradeço ao Dr. Israel da Silva Tojal pela orientação nesses quatro anos, que também foram muito além do projeto em si: pela paciência, pelos *brainstorms*, pelos questionamentos e olhar crítico, e enfim, por não desistir. Também ao Dr. Kenneth Gollob pelas correções e apoio na área de imunologia.

Agradeço aos membros do Laboratório de Biologia Computacional e Bioinformática: Alexandre, Rodrigo, Renan, Monize, Daniel, Bruno, Felipe, Gabriela, por toda ajuda e contribuição ao longo desses anos, e por serem exemplo de bons profissionais. Em especial pela Monize, com quem pude publicar o trabalho relacionado a esta tese e sua dissertação e minha companheira de equipe até hoje.

Por todo o time de pesquisa no CIPE, pelos trabalhos em conjunto, pela discussão de ideias; com certeza uma das principais coisas que levo são todas as pessoas que conheci nesse caminho.

Agradeço, por fim, a banca avaliadora por aceitar avaliar e contribuir com este trabalho.

RESUMO

Buttura JR. **Avaliação do microambiente tumoral em adenocarcinomas gástricos a partir de dados de expressão gênica**. [Tese]. São Paulo; Fundação Antônio Prudente; 2021.

O câncer gástrico (CG) tem grande importância clínica devido a suas altas taxas de incidência e mortalidade estando entre os cinco tipos de câncer mais frequentes no mundo. O CG encontra-se associado ao diagnóstico tardio, que, somado a agressividade da doença, resulta em sobrevida mediana curta. A principal forma de tratamento é a cirurgia em conjunto com a quimioterapia. Mais recentemente, o tratamento com imunoterpia foi aprovado para casos de CG avançado metastático. O CG é usualmente classificado nos subtipos histológicos difuso e intestinal, porém, devido à sua alta heterogeneidade genética, pode ser também classificado em quatro subtipos moleculares (CIN, MSI, GS e EBV). Ambos aspectos de classificação estão associados a diferentes prognósticos e taxas de sobrevida, mas ainda não são suficientes para direcionar ou predizer quais pacientes respondem melhor a uma ou outra terapia. Neste contexto, outro fator relacionado à patogênese da doença, resposta a tratamento e prognóstico é o contexto imunológico do microambiente tumoral, como a composição e orientação funcional das células do infiltrado inflamatório associado ao tumor. Por meio de um método analítico de deconvolução aplicado a dados de expressão gênica é possível estimar a proporção de diferentes tipos celulares, além de seus aspectos funcionais. Neste trabalho, foi feito um levantamento das ferramentas e listas de genes marcadores de células imunes a fim de definir qual teria a melhor performance na estimação da proporção de células imunes em amostras de CG. Não observamos diferenças na composição do infiltrado inflamatório entre os subtipos histológios; já entre os subtipos moleculares de CG as proporções de linfóctios T CD8 se destacaram

como maiores nos subtipos EBV e MSI, por exemplo. Foi então proposta uma classificação baseada em assinaturas mutacionais que conseguiu resgatar pacientes que apresentaram melhor prognóstico. O contexto imunológico nos pacientes com alta atividade mutacional (S4) apresentou elementos de que ser mais responsivo com maior expressão de genes associados à resposta imune citotóxica e do mecanismo de escape pelos *checkpoints* imunes. Além disso, há indícios de que há formação das estruturas terciárias de linfócitos pela correlação entre as quantidades estimadas de linfócitos T e B. Já o grupo com baixa atividade da assinatura (S4), associado a um pior prognóstico, apresentou maiores quantidades de linfócitos Treg. A avaliação global do microambiente tumoral, seus aspectos moleculares e imunológicos, proporciona dados mais infomativos para estratificação dos pacientes quanto ao prognóstico e podem ser utilizadas em breve na prática clínica para direcionamento terapêutico.

Descritores: Adenocarcinoma gástrico. Microambiente tumoral. Infiltrado inflamatório. Assinaturas mutacionais. Imunoterapia. Prognóstico.

ABSTRACT

Buttura JR. [**Evaluation of the tumor microenvironment in gastric adenocarcinomas using gene expression data**]. [tese]. São Paulo; Fundação Antonio Prudente; 2021.

Gastric cancer (GC) has great clinical importance due to its high incidence and mortality rates, being among the five most frequent types of cancer in the world. GC is associated with late diagnosis which, added to the aggressiveness of the disease, results in a short median survival. The main form of treatment is surgery in conjunction with chemotherapy. Recently, immunotherapy treatment has been approved for advanced metastatic GC. GC is usually classified into diffuse and intestinal histological subtypes, however, due to its high genetic heterogeneity, it can also be classified into four molecular subtypes (CIN, MSI, GS and EBV). Both classification are associated with different prognosis and survival rates, but they are still not sufficient to direct or predict which patients respond better to one or another therapy. In this context, another factor related to the pathogenesis of the disease, response to treatment and prognosis is the immunological context of the tumor microenvironment, such as the composition and functional orientation of the cells of the inflammatory infiltrate associated with the tumor. Through an analytical method of deconvolution applied to gene expression data, it is possible to estimate the proportion of different cell types, in addition to their functional aspects. In this work, the tools and lists of immune cell marker genes were surveyed in order to define which one would perform better in estimating the proportion of immune cells in GC samples. We did not observe differences in the composition of the inflammatory infiltrate between the histological subtypes; among the CG molecular subtypes, the proportions of T CD8 lymphocytes stood out as higher in the EBV and MSI subtypes, for example. A classification based

on mutational signatures was then proposed, which managed to rescue patients with a better prognosis. The immunological context in these patients with high mutational activity (S4) showed elements of being more responsive with greater expression of genes associated with the cytotoxic immune response and the escape mechanism by immune *checkpoints*. In addition, there is evidence that there is formation of tertiary lymphocyte structures by the correlation between the estimated amounts of T and B lymphocytes. The group with low mutational signature activity (S4), on the other hand, associated with a worse prognosis, had higher amounts of Treg lymphocytes. The global assessment of the tumor microenvironment, its molecular and immunological aspects, provides more informative data for stratifying patients regarding prognosis and may soon be used in clinical practice for therapeutic guidance.

Keywords: Gastric cancer. Tumoral microenvironment. Inflammatory infiltrate. Mutational signatures. Immunotherapy. Prognostic.

LISTA DE FIGURAS

Figura 1 – **Principais subtipos de células imunológicas.** 2

Figura 2 – ***Hallmarks* do câncer** 3

Figura 3 – *Immunophenoscore* 9

Figura 4 – **Subtipos imunes (C1-6)** 9

Figura 5 – **Composição do infiltrado de células imunes nos subtipos imunes (C1-6).** 10

Figura 6 – **Ferramenta *xCell*** 11

Figura 7 – **Assinaturas mutacionais como junção de vários processos mutacionais** 15

Figura 8 – **Primeiras assinaturas mutacionais identificadas em câncer de mama** 16

Figura 9 – **Distribuição das 30 assinaturas mutacionais por subtipo tumoral** 18

Figura 10 – **Subtipos moleculares de câncer gástrico de acordo com *The Cancer Genome Atlas* (TCGA)** 26

Figura 11 – **Sobrevida global e livre de doença entre os subtipos histológicos de Lauren em coorte chinesa** 29

Figura 12 – **Sobrevida global associada a tumores gástricos localizados na cárdia e não cárdia** 30

Figura 13 – **Sobrevida global entre os subtipos histológicos de Lauren em coorte chinesa por estadiamento** 31

Figura 14 – **Sobrevida global associada aos quatro subtipos moleculares em duas coortes independentes.** 33

Figura 15 – **Similaridade entre diferentes listas de genes marcadores de células imunes** 43

Figura 16 – **Diagrama de Venn ilustrando a sobreposição dos genes.** 44

Figura 18 – ***Heatmaps*** **de expressão de cada lista de genes considerando 8 tipos celulares imunes básicos e os dados de expressão de referência do HPCA (Mabbott et al. 2013)** 55
Figura 19 – **Ferramenta CellMix (Gaujoux e Seoighe 2013).** 56
Figura 20 – **Esquema de análises utilizando os *13 subsets* para comparar os métodos de deconvolução parcial e completa.** 58
Figura 21 – **Esquema da análise de infiltrado inflamatório utilizando método TIMER (Li et al. 2017) para construção das assinaturas de refêrencia.** 60
Figura 22 – **Esquema das análises de *benchmarking*** 64
Figura 23 – **Distribuição do MSE para cada lista de assinaturas de genes considerando 3 células analisadas (linfócitos (B, T CD4+ e T CD8+), macrófagos e DC), para cada um dos métodos** 66
Figura 24 – **Distribuição do MSE para cada lista de assinaturas de genes considerando 4 células analisadas (linfócitos (B, T CD4+, T CD8+ e células NK), neutrófilos, macrófagos e DC), para cada um dos métodos** 67
Figura 25 – **Correlação de Pearson entre as proporções estimadas de 7 células pelo CIBERSORT-Newman2015 e a expressão de genes marcadores principais característicos de cada célula.** 69
Figura 26 – **Correlação de Pearson das proporções estimadas de células CIBERSORT-Newman2015 com *scores* de *xCell* (Aran et al. 2017)**. 70
Figura 27 – ***Heatmap*** **de correlação de Pearson entre as quantidades absolutas estimadas com método *CIBERSORT* (linhas) e proporções de células por *deep learning* (colunas) (Saltz et al. 2018).** 72

Figura 28 – **Correlograma de Pearson das quantidades absolutas estimadas com método *CIBERSORT* (linhas) e expressão de genes principais marcadores (colunas).** . . . 74
Figura 29 – ***Heatmap* de correlação de Pearson entre as quantidades absolutas de células imunes estimadas com método *CIBERSORT* (linhas) e *scores* estimados pelo método *xCell* (colunas) a partir do FPKM de amostras TCGA-STAD.** . . . 75
Figura 30 – **Proposta e considerações para análise da composição do infiltrado inflamatório utilizando métodos de deconvolução.** . . . 78
Figura 31 – ***Boxplots* das proporções de células estimadas pelo CIBERSORT-Newman2015 nos subtipos histológicos de AdG.** . . . 80
Figura 32 – ***Boxplots* das proporções de células estimadas pelo CIBERSORT-Newman2015 nos subtipos moleculares de AdG.** . . . 82
Figura 33 – Sobrevida global entre os subtipos moleculares de AdG. . . . 83
Figura 34 – ***Heatmaps* da proporção estimada das 7 células nos subtipos intestinal juntamente com a classificação molecular.** . . . 84
Figura 35 – **Modelo simples de Cox de sobrevida global em amostras de AdG (TCGA-STAD).** . . . 85
Figura 36 – **Modelo de regressão simples de Cox para sobrevida global considerando variáveis clínicas.** . . . 87
Figura 37 – **Modelo de regressão múltiplo considerando as variáveis clínicas significativas e exposição a assinatura S4** . 87
Figura 38 – **Sobrevida global entre os grupos S4-high e S4-low (Buttura et al. 2021)** . . . 88

Figura 39 – ***Heatmap*** **com valores normalizados para comparação entre os grupos S4-T e S4-O mostrando a exposição da assinatura mutational S4 (MMR), carga mutacional (TMB), status MSI, genes MMR com mutação somática (SRG) e heterogeneidade tumoral, bem como anotação de variáveis clínicas (sítio anatômico, classificação histológica de Lauren e subtipo molecular).** . . . 89
Figura 40 – **Comparação das quantidades de células imunes estimadas pelo método CIBERSORT entre os grupos S4-high e S4-low.** 90
Figura 41 – **Expressão gênica de genes marcadores de resposta imune com diferença significativa entre os grupos S4-high e S4-low**. 91
Figura 42 – **Modelo simples de Cox para sobrevida global considerando as variáveis clínicas, moleculares e imunes nas amostras de AdG (TCGA-STAD).** 92
Figura 43 – **Modelo simples de Cox para sobrevida global considerando expressão de genes efetores da resposta imune em amostras de AdG (TCGA-STAD).** 93
Figura 44 – **Modelo simples de Cox para sobrevida global considerando quantidade estimada de células imunes em amostras de AdG (TCGA-STAD).** 94
Figura 45 – **Correlograma de Pearson entre as quantidades de células estimadas por grupo S4-high e S4-low.** 96
Figura 46 – **Correlograma de Pearson da expressão de genes efetores da resposta imune por grupo S4-high e S4-low.** . . 99
Figura 47 – **Correlograma de Pearson da expressão de genes do repertório de receptores de linfócitos T (TCR) por grupo S4-high e S4-low.** 101

Figura 48 – **Correlograma de Pearson da expressão de genes do repertório de receptores de células B (Ig) por grupo S4-high e S4-low.** 103

LISTA DE TABELAS

Tabela 1 – **Mecanismos de reparo do DNA** 19
Tabela 2 – **Genes marcadores utilizados para avaliação da estimação de células imunes** 38
Tabela 3 – **Revisão de listas de genes marcadores e matrizes de assinatura de métodos de deconvolução pré existentes.** 42
Tabela 4 – **Revisão de métodos de avaliação da composição de infiltrado inflamatório utilizando dados de expressão gênica.** 57
Tabela 5 – **Revisão de métodos de avaliação da composição de infiltrado inflamatório utilizando dados de expressão gênica.** 93

SUMÁRIO

1 INTRODUÇÃO 1

1.1 Aspectos imunológicos do câncer 1

1.1.1 Bases da imuno oncologia 1

1.1.2 Avaliação do perfil imunológico 7

1.2 Contexto mutacional em câncer 13

1.2.1 Assinaturas mutacionais 13

1.2.2 Mecanismos de reparo 19

1.3 Câncer gástrico 22

1.3.1 Epidemiologia 22

1.3.2 Aspectos clínicos e moleculares 23

1.3.3 Prognóstico e tratamento 28

2 JUSTIFICATIVA 34

3 OBJETIVOS 35

3.1 Objetivo geral 35

3.2 Objetivos específicos 35

4 METODOLOGIA 36

4.1 Dados clínicos e moleculares 36

4.2 Estimação do microambiente tumoral e contexto imunológico 36

4.2.1 Genes marcadores e construção da matriz de assinatura 36

4.2.2 Avaliação de desempenho dos métodos de deconvolução 37

4.3 Análises estatísticas 39

5 RESULTADOS E DISCUSSÃO 40

5.1 *Benchmarking* da avaliação da composição do infiltrado inflamatório 40

5.1.1 Revisão de genes marcadores de células imunes 40
5.1.2 Revisão dos métodos de deconvolução 55
5.1.3 Desempenho dos algoritmos de deconvolução de células imunes . . . 63
5.1.4 Proposta de *pipeline* para avaliação do contexto imunológico 76
5.2 Contexto imunológico em câncer gástrico e seus aspectos prognósticos 79
5.2.1 Infiltrado inflamatório nos subtipos de câncer gástrico 79
5.2.2 Características clínico-patológicas, assinaturas mutacionais e interação com contexto imunológico 86

6 CONCLUSÕES 104

7 REFERÊNCIAS 105

APÊNDICES 115

Apêndice A – PUBLICAÇÃO: MUTATIONAL SIGNATURES DRIVEN BY EPIGENETIC DETERMINANTS ENABLE THE STRATIFICATION OF PATIENTS WITH GASTRIC CANCER FOR THERAPEUTIC INTERVENTION 116

1 INTRODUÇÃO

1.1 ASPECTOS IMUNOLÓGICOS DO CÂNCER

1.1.1 Bases da imuno oncologia

O sistema imunológico é composto por células e moléculas cujo principal objetivo é reconhecer, conter e eliminar organismos, incluindo células estranhas, que ameaçam o equilíbrio dinâmico entre a morte e a proliferação celular, de acordo com a demanda funcional. Qualquer estímulo que quebre este equilíbrio promove adaptações celulares que podem ser reversíveis ou progredir para irreversibilidade levando à morte celular exacerbada (lesão tecidual) ou transformação celular (malignidade).

Didaticamente, o sistema imunológico pode ser entendido como resposta do tipo inata e do tipo adaptativa de acordo com uma resposta mais rápida, pronta e inespecífica, no primeiro caso, ou mais lenta e específica, no segundo caso. Estes tipos de resposta estão associadas a diferentes tipos celulares que podem ser também divididos basicamente nas linhagens mielóide e linfóide (Figura 1).

Em geral, as células da linhagem mielóide participam da resposta imune inata, que está de certa forma associada também a processos da inflamação aguda. Os granulócitos (neutrófilos, basófilos, mastócitos, eosinófilos), principalmente os neutrófilos, e os fagócitos, macrófagos teciduais e monócitos recrutados da circulação, são as primeiras células que agem contra um agente injuriante, neutralizando e promovendo a fagocitose e digestão de microorganismos ou células alteradas. A resposta inata reconhece padrões e inicialmente contribui para aumentar a circulação sanguínea no local, atrair mais células inflamatórias e capturar antígenos para formação da resposta adaptativa.

Na resposta adaptativa ou adquirida, os linfócitos T e B são os responsáveis pela resposta específica contra antígenos do tipo celular e humoral (anticorpos),

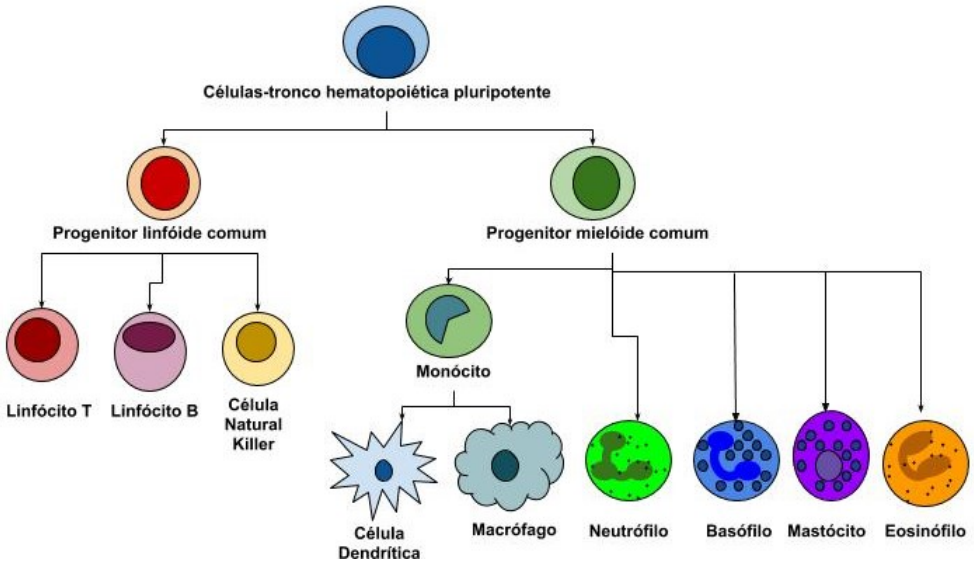


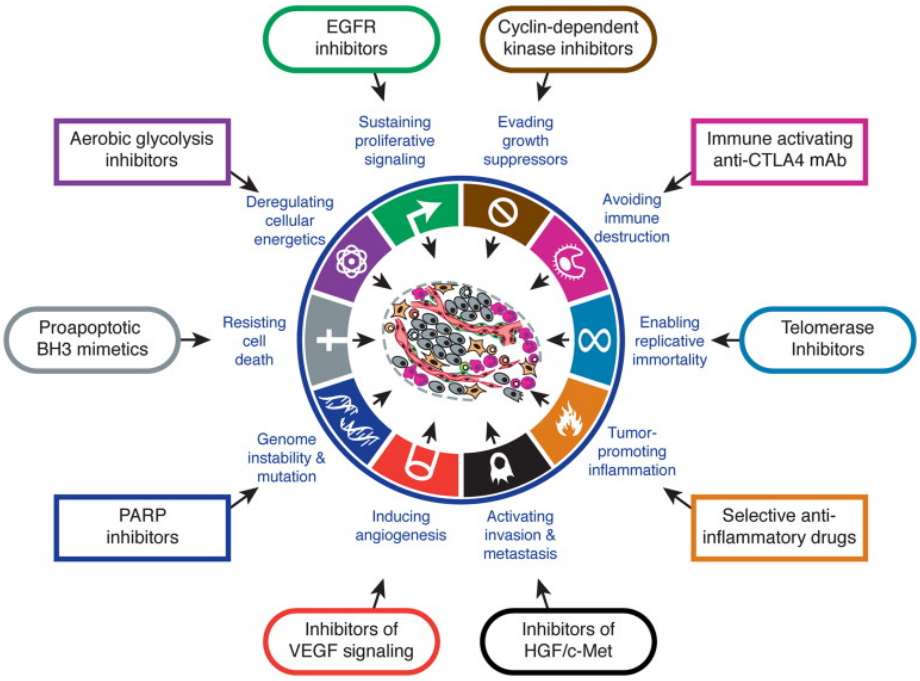
Figura 1 – Principais subtipos de células imunológicas. Os tipos básicos de células imunológicas da linhagem mielóide são os monócitos presentes na circulação e que nos tecidos se diferenciam em macrófagos; células dendríticas, e granulócitos: neutrófilos, basófilos, mastócitos e eosinófilos. Os subtipos básicos de células da linhagem linfóide são os linfócitos B e T e as células *Natural Killer* (NK)). Imagem própria.

respectivamente. Quando, na inflamação aguda, as células da imunidade inata não são suficientes para eliminar o agente agressor, forma-se um ambiente inflamatório crônico no qual tem-se mais infiltrado de característica linfóide, além da ativação de vias de reparo tecidual. Os mecanismos de reparo são importantes na resolução da inflamação, com participação de células estromais, como fibroblastos, que auxiliam a cicatrização do tecido e restabelecimento da homeostasia.

Segundo Rudolph Virchow, "câncer são feridas que não cicatrizam" (Whiteside 2008) e estão associados a um microambiente típico de inflamação crônica (Giraldo et al. 2015; Whiteside 2008). O estabelecimento de um tumor acontece quando há um desequilíbrio entre a taxa de proliferação celular e morte celular. No processo de carcinogênese, diversos fatores injuriantes (biológicos, físicos e/ou químicos) causam lesão celular e contribuem para a instalação de um microambiente inflamatório. Somado a isso, há o acúmulo de mutações nas células que auxiliam na transformação celular. As principais marcas do câncer então são células tumorais que diferem de células do tecido normal adjacente pela sua maior capacidade de proliferação celular, resistência à morte celular e capacidade de

invasão de tecidos adjacentes (Hanahan e Weinberg 2000).

Em 2011, esta lista de marcas do câncer foi atualizada incluindo o contexto imunológico como importante fator no processo de carcinogênese e prognóstico do câncer (Figura 2). A inflamação crônica associada ao câncer pode ter uma função pró tumoral na medida que contribui com moléculas bioativas no microambiente tumoral, incluindo fatores de crescimento que sustentam sinais proliferativos, fatores de sobrevida que diminuem a morte celular, fatores pró angiogênicos e enzimas de matriz extracelular, que facilitam angiogênese, invasão e metástase, além de sinais que podem levar a transição epitélio mesênquima (Hanahan e Weinberg 2011).



Fonte: Hanahan e Weinberg (2011).

Figura 2 – *Hallmarks* do câncer. Em 2011 a lista das características assoaciadas a carcinogênese foram atualizadas incluindo, além dos aspectos relacionados a apoptose, fatores de crescimento, angiogênese e invasão tecidual, os aspectos relacionados ao papel da instabilidade genômica, metabolismo celular, resposta imunológica e inflamação como fatores importantes do microambiente tumoral.

Por outro lado, a presença do infiltrado inflamatório no sítio tumoral indica que o hospedeiro não está alheio à presença do tumor, que a vigilância imunológica detecta algo estranho e há ou houve tentativas de controlar a pro-

gressão do tumor (Whiteside 2008). Como as células tumorais são, em parte, estranhas ao organismo, há ativação do sistema imunológico com recrutamento dessas células imunes para o sítio tumoral em resposta às disfunções associadas às células tumorais e a mediadores inflamatórios produzidos no microambiente tumoral, como aqueles produzidos devido à hipóxia, lesão tecidual ou infecções. Um microambiente tumoral então é formado, composto por células tumorais, infiltrado inflamatório e células do estroma (fibroblastos e células endoteliais, principalmente) que interagem entre si e ditam a relação tumor-hospedeiro.

A interação entre sistema imunológico e tumor envolve um processo de edição imunológica que consiste em três fases: eliminação, equilíbrio e escape (Wang et al. 2016). Na primeira fase, células imunes eliminam células tumorais pelo reconhecimento de antígenos estranhos; porém, devido a heterogeneidade tumoral algumas células neoplásicas podem sobreviver, caracterizando a segunda fase, de equilíbrio, na qual o percentual de células destruídas pelo sistema imune é contrabalanceado pelo número de células que adquirem novos mecanismos de escape imunológico. Na última fase, os mecanismos de escape superam a capacidade do sistema imune de reagir contra e eliminar as células tumorais, e as células resistentes iniciam um processo de proliferação e possível metástase, representando a fase de escape do tumor (Wang et al. 2016). Em todas essas etapas, o infiltrado imune está presente tentando eliminar as células transformadas, mas também contribuindo para inflamação crônica discutida anteriormente.

Em geral, os tumores são diagnosticados já na fase de escape com um microambiente tumoral que apresenta características de tolerância imunológica e exaustão das células imunes efetoras como:

- ignorância dos linfócitos T às células tumorais, por exemplo pela diminuição da expressão de moléculas HLA;
- ativação de mecanismos de tolerância ou inibição da resposta imunológica, como citocinas imunossupressoras e polarização de fenótipos celulares su-

pressores (como macrófagos associados ao tumor – *tumor-associated macrophages*, TAM; células mielóides supressoras - *myeloid-derived suppressor* cells, MDSC; células T reguladoras, Treg); e

- resistência das células tumorais por meio da expressão de proteínas de *checkpoint* imune, que inibem a resposta imunológica, como PD-L1 ou -2, CTLA-4, TIM-3, LAG3 e galectina-9 (Wang et al. 2016).

Esse aspecto de escape imune também foi incluído na lista das marcas do câncer em 2011, seguindo a lógica de que o sistema imunológico está sempre em vigilância. Nesse contexto, a progressão de um tumor sólido relaciona-se a sua capacidade de evasão (escape imunológico) ao evitar os vários mecanismos de detecção do sistema imune ou de limitar sua ação efetora (Hanahan e Weinberg 2011).

As interações dinâmicas no microambiente tumoral, que dependem das vias oncogênicas que promoveram o tumor, apresentam diferenças na composição, densidade e organização do infiltrado resultando em diferentes mecanismos de escape tumoral (Becht et al. 2016; Wang et al. 2016; Gajewski et al. 2013; Whiteside 2008). Segundo Becht et al. (2016) os tumores podem ser divididos em três tipos: inflamatório (relacionado à resposta imune inata; pior prognóstico), imunogênico (relacionado ação de linfócitos; melhor prognóstico) e negligenciado/ignorado imunologicamente (*escaping tumor*) (prognóstico intermediário). Estes tipos informam sobre as características do microambiente tumoral e ativação da resposta imunológica antitumoral presente (Becht et al. 2016).

Gajewski et al. (2013), por sua vez, definiram dois imunofenótipos tumorais: um relacionado a infiltrado de células T, e outro sem este infiltrado, caracterizado por um perfil de células mielóides. No primeiro caso, o infiltrado de linfócitos T caracteriza uma resposta espontânea do hospedeiro contra o tumor, com amplo painel de citocinas e interferons do tipo I; este fenótipo está associado ao melhor prognóstico. No segundo caso, o estroma e células mielóides,

como os macrófagos, apresentam expressão aumentada de fatores angiogênicos e imunossupressores, estando associado a pior prognóstico e menor resposta à imunoterapia (Gajewski et al. 2013).

É importante ressaltar que a simples presença de infiltrado inflamatório não pode ser associada diretamente ao melhor prognóstico e sobrevida do paciente (Whiteside 2008), uma vez que esta relação depende da composição e orientação funcional das células componentes deste infiltrado. Por exemplo, um infiltrado caracterizado em sua maioria por linfócitos, já foi associado a melhor taxa de resposta ou melhor prognóstico em vários tipos de tumores como colorretal, mama, renal, melanoma, ovário, próstata, pulmão, fígado, esôfago, bexiga e mesmo estômago (Zuo et al. 2020; Zhang et al. 2019a; Ye et al. 2019; Becht et al. 2016; Giraldo et al. 2015; Gajewski et al. 2013). No entanto esta relação não se mantém quando o infiltrado de células T é composto em sua maioria por células T CD4+Foxp3+, também chamadas de células T reguladoras (Treg), que são células com atividade reguladora e imunossupressora, ou seja, auxiliam na progressão do tumor inibindo atividade de outras células efetoras (Becht et al. 2016; Giraldo et al. 2015; Whiteside 2008).

Outro caso semelhante, é quando há linfócitos T citotóxicos (T CD8+) presentes, porém estes são anérgicos, ou seja, não funcionais devido inibição da sua ativação contra o tumor no microambiente tumoral (Wang et al. 2016; Gajewski et al. 2013). Por outro lado, quando o infiltrado é rico em linfócitos e há formação de "estrutura terciária de linfócitos" (*tertiary lymphoid structure*, TLS), que é a organização de linfócitos T e B em zonas e agregados peri tumoral (semelhantes aos órgãos linfoides secundários), esta composição e organização representam uma estrutura ativa de resposta antitumoral, e está associada a bons prognósticos (Becht et al. 2016; Giraldo et al. 2015).

Portanto, a composição de células do infiltrado inflamatório é a princípio uma resposta do hospedeiro contra o tumor, mas também tem um papel chave no crescimento do tumor, processo de angiogênese e progressão tumoral (Stock-

mann et al. 2014). De importância, a diversidade dos componentes inflamatórios encontra-se associada à resposta ao tratamento e ao desfecho clínico do paciente (Becht et al. 2016; Wang et al. 2016; Gajewski et al. 2013; Whiteside 2008).

1.1.2 Avaliação do perfil imunológico

O contexto imunológico do microambiente tumoral tem impacto clínico (**??**Zhang et al. 2019a; Ye et al. 2019; Becht et al. 2016; Tang et al. 2016; Fridman et al. 2012). Um dos motivos que estimularam os estudos de avaliação do contexto imunológico associado ao prognóstico em câncer foi a observação de pacientes com câncer colorretal e mesmo estadiamento clínico (TNM) que apresentaram diferentes desfechos. Havia a necessidade de classificar os pacientes de forma melhor em relação ao prognóstico e predizer resposta à terapia (Galon et al. 2012; Galon et al. 2006). Foi observado então que a caracterização do infiltrado de células imunológicas (localização, densidade e tipo celular) mostrou-se como melhor preditor de sobrevida (Galon et al. 2012; Galon et al. 2006).

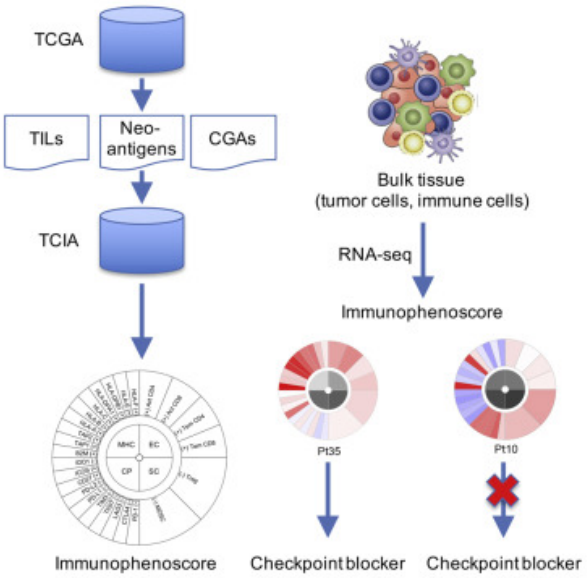
Em 2012, foi proposta uma nova classificação considerando os aspectos imunológicos do tumor (TNM-Immune ou TNM-I) por meio da utilização da ferramenta *Immunoscore* (Galon et al. 2012). Essa ferramenta propõe uma métrica do contexto imunológico baseada na avaliação de aspectos determinantes da resposta imune, como a análise da composição de linfócitos, localização e densidade dessas células no microambiente tumoral. Considerando a importância do contexto imunológico em câncer, como fator preditor de repossta a terapia, o *Immunoscore* poderia auxiliar na escolha de estratégias terapêuticas e predizer prognóstico, além de auxiliar no desenvolvimento de novas técnicas de tratamento (Becht et al. 2016; Giraldo et al. 2015; Gajewski et al. 2013; Galon et al. 2012). Desde então, um consórcio foi formado com objetivo de fazer essa classificação ser aplicada na prática clínica (Angell et al. 2020; Bruni et al. 2020; Galon e Lanzi 2020; Liu et al. 2018). A discussão ainda é atual e tem como base estudos que apontaram para relevante valor prognóstico e preditivo de resposta a terapia

utilizando a classificação pelo *Immunoscore*, atualmente baseado em técnicas de imunohistoquímica e patologia digital.

Em 2017, baseado na primeira versão do *Immunscore*, outra estratégia mais completa de estudar o contexto imunológico e sua relação prognóstica foi proposta com um *score* obtido a partir do *Immunophenoscore* (Charoentong et al. 2017). A diferença entre as duas versões, propostas pelo mesmo grupo, é que no Immunophenoscore foi incluída avaliação dos checkpoint imunes, genes *HLA* e outros efetores associados a resposta imune citotóxica (Figura 3). Este *score* funciona como um preditor de resposta a tratamento com inibidores de *checkpoint* imunes (como anti-PD-1 e anti-CTLA-4) ao considerar as informações dos aspectos de imunogenicidade em tumores sólidos como infiltrado de células imunes efetoras e reguladoras, maquinaria de apresentação de antígenos e expressão de *checkpoints* imunes (Charoentong et al. 2017). Os resultados encontrados com este *score* reforçaram a importância do contexto imunológico na conduta clínica.

Em seguida, considerando a estratégia de integrar métodos de imunogenômica para classificar o microambiente tumoral, Thorsson et al. (2018) definiram 6 subtipos tumorais imunes (C1 a C6). Estes subtipos tumorais foram definidos considerando 5 conjuntos de genes relacionados a respostas de: cicatrização (*would healing*), interferon gama, inflamatória, depleção de linfócitos e silenciamento imunológico (Figura 4). Por meio de métodos de análise de enriquecimento, a maior ou menor expressão desses conjuntos de genes direcionou a classificação nos seis subtipos imunes aplicados em vários tipos tumorais (Thorsson et al. 2018).

Avaliando os diferentes tipos tumorais, cada um dos seis subtipos imunes apresentou particularidades comuns em relação à fração de leucócitos, taxa de proliferação, heterogeneidade intratumoral e composição de infiltrado inflamatório (Figura 5). Esta foi uma tentativa de integrar informações clínicas, moleculares e imunes considerando diferentes tipos de tumores a fim de direcionar o prognóstico de uma forma unificada. De importância, este estudo endossa a



Fonte: Charoentong et al. (2017).

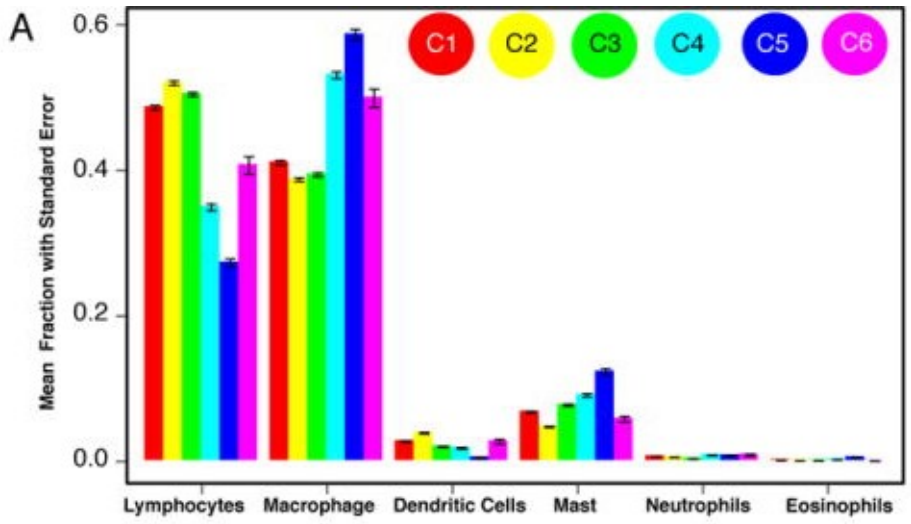
Figura 3 – *Immunophenoscore*. A análise pelo *Immunophenoscore* utiliza como base dados de RNA-seq e resulta em um *score* baseado em determinantes da imunogenicidade do tumor: composição e enriquecimento do infiltrado de linfócitos (*tumor infiltrating lymphocytes*, TILs), antígenos tumorais (neoantígenos e antígenos aberrantes CGA - *cancer germline antigen*), heterogeneidade tumoral, imunofenótipos, definidos pela associação entre o infiltrado de linfócitos e genes HLA e inibidores ou esimuladores da resposta imunológica.

	Macrophage: lymphocyte	Th1:Th2	Proliferation	Intratumoral heterogeneity	Other
Wound healing	Balanced	Low	High	High	
IFN-γ dominant	Lowest	Lowest	High	Highest	Highest M1 and highest CD8 T cells
Inflammatory	Balanced	High	Low	Lowest	Highest Th17
Lymphocyte depleted	High	Minimal Th	Moderate	Moderate	
Immunologically quiet	Highest	Minimal Th	Low	Low	Highest M2
TGF-β dominant	High	Balanced	Moderate	Moderate	Highest TGF-β signature

Fonte: Thorsson et al. (2018).

Figura 4 – Subtipos imunes (C1-6). Subtipo C1 (vermelho): *would healing*; C2 (amarelo): interferon gama dominante; C3 (verde): inflamatório ; C4 (azul claro): depleção de linfócitos; C5 (azul escuro): silenciamento imunológico; C6 (rosa): TGF-beta dominante.

aplicação de uma perspectiva imunológica como meio de avaliação prognóstica (Thorsson et al. 2018).

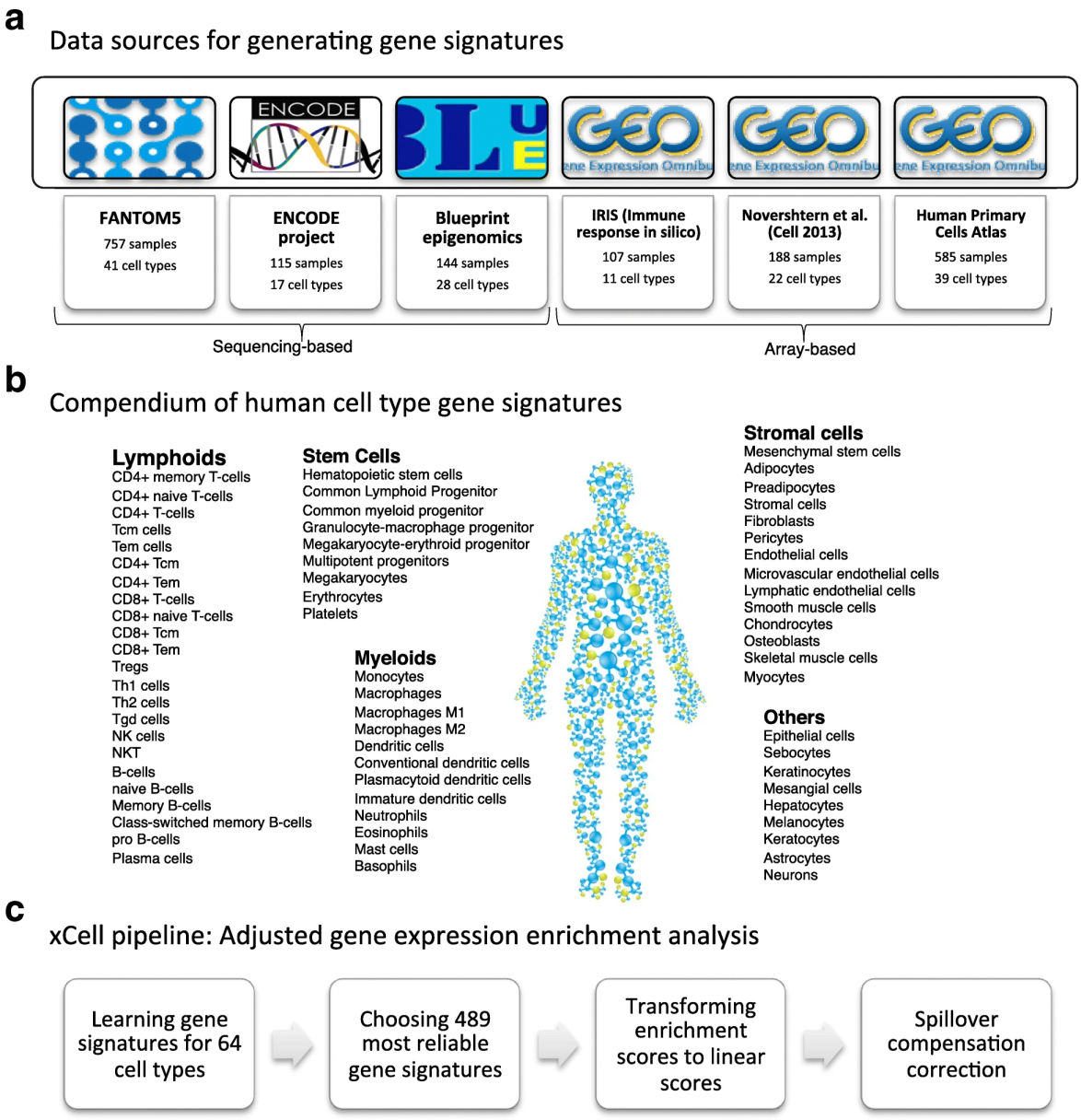


Fonte: Thorsson et al. (2018).

Figura 5 – Composição do infiltrado de células imunes nos subtipos imunes (C1-6).

A abordagem de análise de enriquecimento também foi aplicada por Aran et al. (2017) com o método *xCell*. Utilizando dados de expressão gênica de células puras obtidos por *microarray* e por RNA-Seq, foi definida uma assinatura de genes marcadores utilizada para análise de enriquecimento, ajustada para comparação linear de tipos de células. O método foi validado por meio de análises *in silico* e comparação com imunofenotipagem por citometria. *xCell* resulta em uma avaliação da composição de mais de 30 subtipos de células imunológicas, inclusive progenitoras, e células do estroma, totalizando 64 subtipos celulares avaliados (Figura 6). A desvantagem deste método é que os *scores* gerados para os diferentes tipos celulares não são comparáveis entre si, fazendo com que se perca a noção de composição do infiltrado inter amostra.

Como mencionado anteriormente, estudar a interação entre a resposta imunológica no ambiente tumoral envolve conhecer os aspectos moleculares e celulares considerando as células tumorais, estroma e composição do infiltrado



Fonte: Aran et al. (2017).

Figura 6 – Ferramenta *xCell*. (a) Bancos de dados utilizados como referência da ferramenta xCell, (b) tipos celulares considerados e (c) método de construção da ferramenta.

tumoral, incluindo sua heterogeneidade, composição e os aspectos funcionais. Considerando que cada tipo e subtipo tumoral possui uma diversidade na composição de infiltrado de células imunes já instalada no microambiente tumoral, é esperado que estratégias de tratamento mais personalizadas possam ser tomadas.

Em conjunto com as ferramentas já citadas, uma abordagem mais completa de avaliação desse contexto compreende na aplicação de métodos analíticos de deconvolução capaz de estimar a proporção de diferentes tipos celulares com base em um painel de genes e os respectivos níveis de expressão gênica (Shen-Orr e Gaujoux 2013). Esta abordagem assume que um conjunto de genes apresenta atividade transcricional em células específicas. Diante dessa premissa, a abordagem utiliza uma assinatura com expressão gênica esperada obtidas de células puras a fim de extrair as proporções das respectivas células em amostras misturadas de acordo com a equação:

$$A_{11}X_1 + A_{12}X_2 + ...A_{1j}X_j = B_1$$

$$A_{21}X_1 + A_{22}X_2 + ...A_{2j}X_j = B_2$$

$$...$$

$$A_{i1}X_1 + A_{i2}X_2 + ...A_{ij}X_j = B_j$$

$$AX = B$$

O somatório da expressão de um gene de cada célula "**A**" vezes sua proporção "**X**" na amostra resulta no valor de expressão "**B**" encontrado na amostra (mistura). O resultado da análise por deconvolução é então uma matriz contendo as proporções de cada célula por amostra cujo somatório é 1 (ou 100%).

Os primeiros algoritmos de deconvolução foram desenvolvidos para estimar a proporção de subtipos de células circulantes (amostras de sangue total), e não microambiente tumoral. Além disso, os dados de expressão gênica eram baseados

na tecnologia de *microarray* (Gaujoux e Seoighe 2013; Gong et al. 2011; Abbas et al. 2009; Abbas et al. 2005). Desde meados de 2015 (Li et al. 2017; Newman et al. 2015), vários algoritmos de deconvolução tem sido criados a fim de adaptar sua utilização com base em dados de RNA-Seq e estimar as proporções de amostras teciduais. Além da maior profundidade da tecnologia de sequenciamento de nova geração (NGS), as amostras teciduais possuem células normais, células tumorais e do estroma, sendo que as células imunes compreendem apenas uma porcentagem das células nessa mistura.

Ainda não existe uma estratégia consolidada para avaliação do contexto imune do câncer. Os estudos na área imuno-oncologia cresceram consideravelmente nos últimos anos a fim de firmar um conhecimento sólido que possa ser efetivamente aplicado na clínica, como a efetiva caracterização de uma assinatura imune com sensibilidade e relevância prognóstica. Assim, a caracterização global do contexto imunológico do microambiente tumoral e seus valores prognósticos e/ou preditivos ainda são um desafio.

1.2 CONTEXTO MUTACIONAL EM CÂNCER

1.2.1 Assinaturas mutacionais

Os cânceres surgem devido a um acúmulo de mutações em um subconjunto de genes que conferem vantagem de crescimento às células, favorecendo a transformação das células normais em neoplásicas por meio da inibição de vias de morte celular, além de todas as demais marcas do câncer descritas por Hanahan e Weinberg (2011) (Figura 2). Essas mutações podem ocorrer durante a replicação do DNA ou através da exposição a mutagênicos exógenos ou endógenos (Yates e Campbell 2012; Greenman et al. 2007).

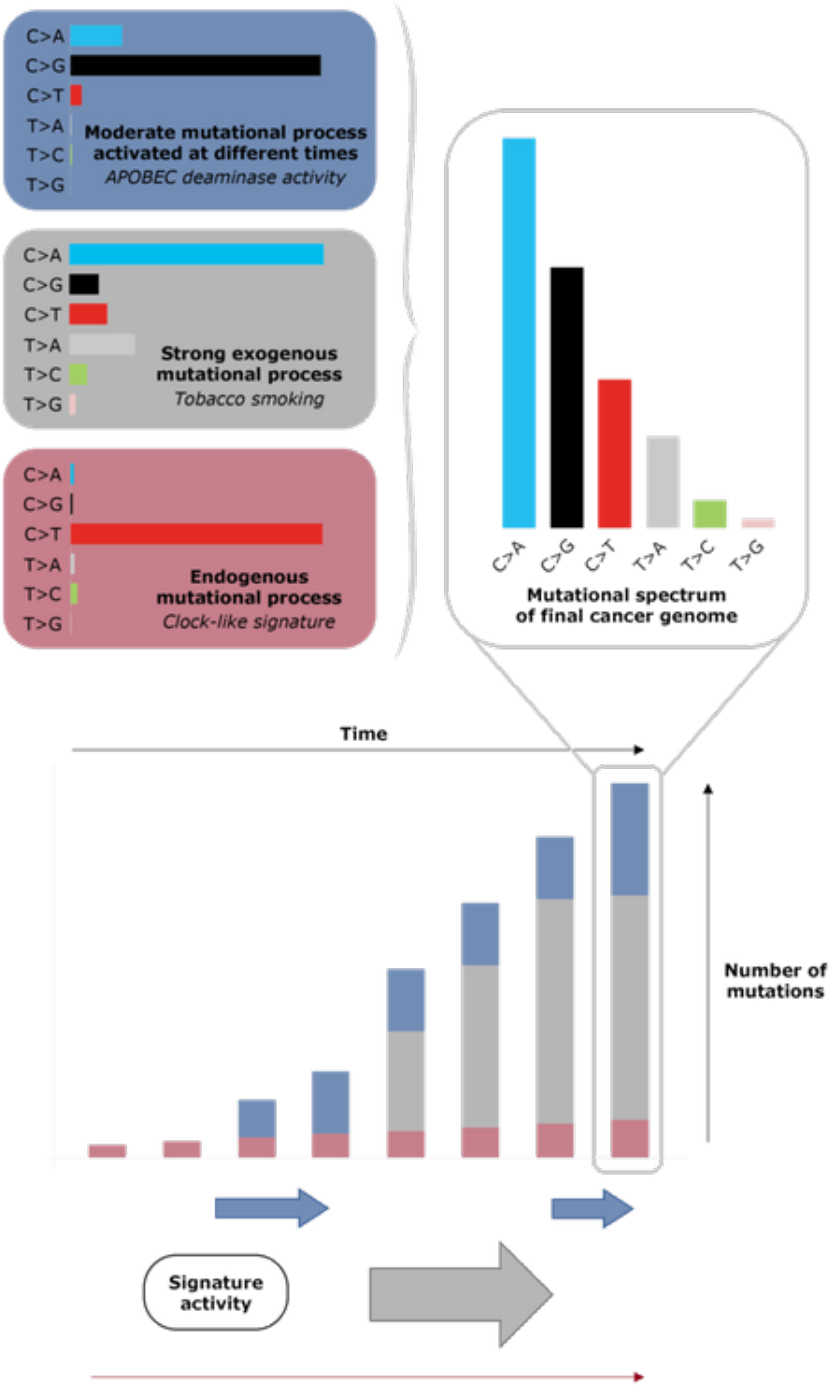
Inicialmente, os estudos na área de oncologia focaram na definição dessas mutações com objetivo de direcionar terapias alvos, realizar o diagnóstico diferencial de subtipos tumorais ou relacionar com prognóstico. No entanto, o genoma

tumoral compreende um acúmulo de muitas outras mutações (*passengers*) que, mesmo não influenciando diretamente o processo de carcinogênese, contribuem para a heterogeneidade tumoral e informam sobre a biologia do câncer em um contexto evolutivo (Helleday et al. 2014).

Os padrões de mutações que ocorrem no genoma do câncer refletem sobre quais processos mutacionais as células foram expostas. Os processos mutacionais podem ser endógenos ou exógenos e envolvem componentes de dano ou modificação de DNA, reparo de DNA e replicação de DNA (que pode ser normal ou anormal). Tais processos geram um espectro mutacional com alterações que podem ser mutações de base única, inserções, deleções, alteração do número de cópias (CNV), rearranjos e substituições. O perfil mutacional presente no genoma tumoral é um reflexo da ação de diferentes processos mutacionais, de acordo com o tempo e intensidade de exposição a estes processos (Figura 7).

As mutações de base única (SNV, *single nucleotide variation*) são as mais comuns (Stratton et al. 2009) no genoma humano. As possibilidades de trocas de bases são seis: C:G>A:T, C:G>G:C, C:G>T:A, T:A>A:T, T:A>C:G e T:A>G:C. Considerando uma base adjacente, antes e depois, de cada uma dessas possíveis trocas, soma-se um total de 96 trincas de possibilidades mutacionais. Considerando todo o conjunto de mutações somáticas de base única no genoma de um tumor, é possível avaliar a frequência de cada uma dessas trincas, cujas combinações refletem assinaturas mutacionais que são resultantes dos processos mutacionais individuais que estão ativos no tumor, ou seja, informam sobre os mecanismos mutacionais ativos naquele tumor, que participaram da sua formação.

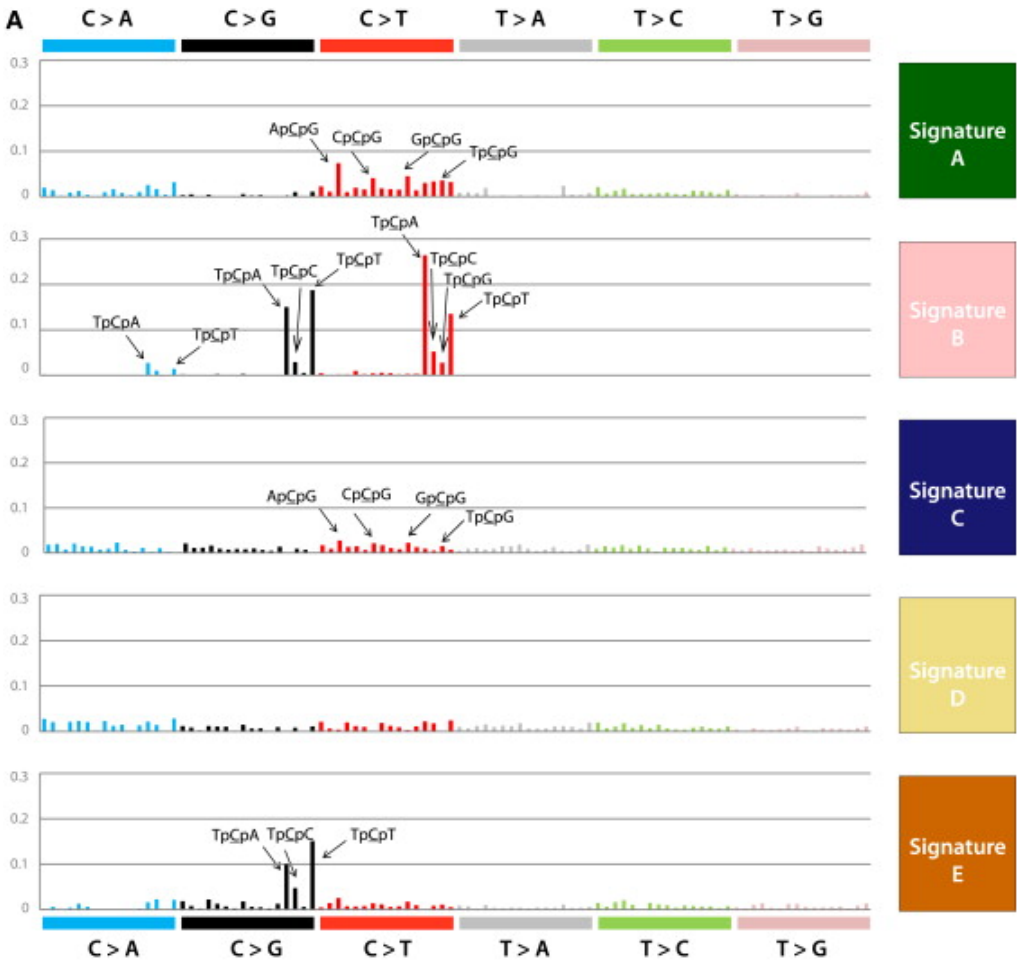
Os estudos de assinaturas mutacionais iniciaram em meados 2012 quando Nik-Zainal et al. (2012), com base nessas 96 trincas de possíveis substituições de base, utilizaram uma generalização do método NMF (*Non-negative Matrix Factorization*) para avaliar as diferentes combinações e frequências destas trocas. Considerando 21 genomas tumorais de pacientes com câncer de mama, foram



Fonte: https://cancer.sanger.ac.uk/signatures/

Figura 7 – Assinaturas mutacionais como junção de vários processos mutacionais. Cada processo mutacional de diferentes etiologias estão atuando no câncer e originando mutações (trocas) características no DNA das células tumorais. Quando vistas em conjunto, a coleção de mutações caracteriza cada assinatura mutacional mais ou menos ativa no tumor.

encontradas 5 assinaturas mutacionais denominadas como A, B, C D e E (Figura 8). O que define cada assinatura é justamente as diferentes frequências de cada troca e as combinações entre elas. É importante ressaltar que todas as assinaturas podem estar presentes no genoma tumoral, variando em frequência, ou exposição.



Fonte: Nik-Zainal et al. (2012)

Figura 8 – Primeiras assinaturas mutacionais identificadas em câncer de mama. A assinatura **A** tem maior frequência de trocas no contexto C>T e C>A; assinatura **B** caracterizada por trocas C>A, C>G e C>T; e destaque também para assinatura **E** com trocas no contexto C>G.

Além da caracterização dessas 5 assinaturas mutacionais em câncer de mama, foi possível associar fatores etiológicos, ou seja, quais os aspectos geradores daquele perfil mutacional. Por exemplo, a assinatura A associou-se ao processo mutacional de desaminação espontânea de 5-metilcitosina, cujas substituições ca-

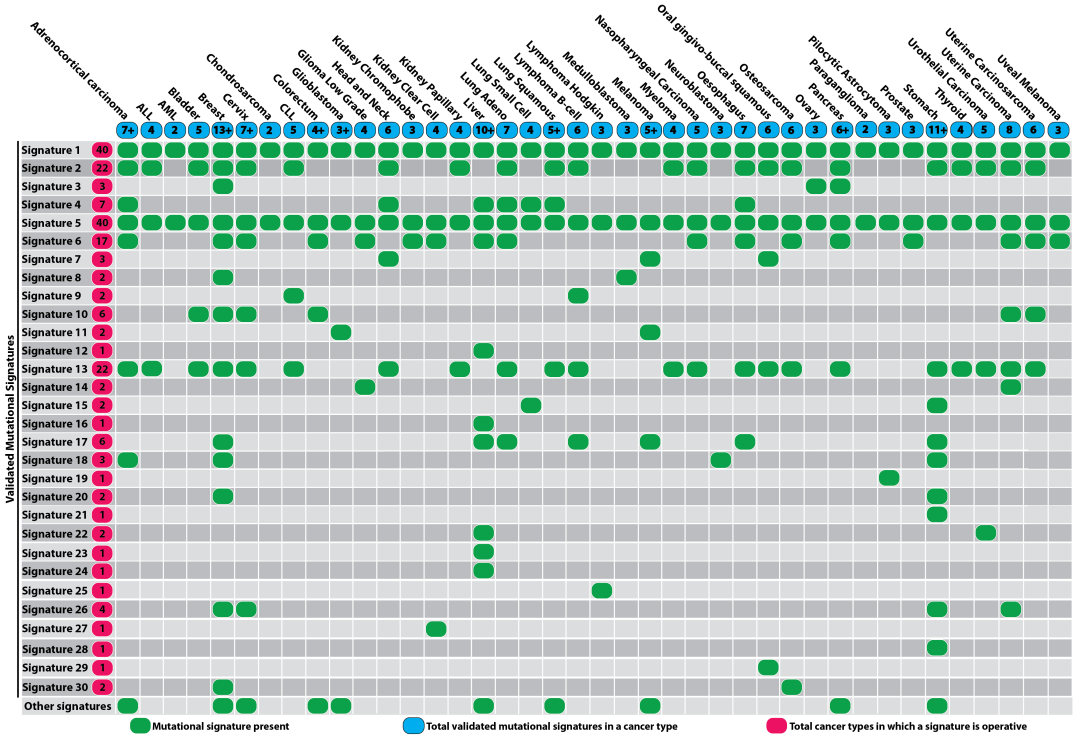
racterísticas deste processo são as trocas no contexto C>T principalmente (Figura 8).

O repertório de assinaturas mutacionais tem sido expandido desde pelo grupo de Alexandrov e colaboradores por meio de uma análise ampla de cerca de SNVs provenientes de tumores primários disponíveis no TCGA e em outros bancos. Além da detecção de diferentes assinaturas, a definição de suas etiologia e caracterização da exposição em cada um dos subtipos tumorais (Figura 9) formaram o banco de dados de assinaturas mutacionais no COSMIC [1]. Em março deste ano (2021), uma nova versão de assinaturas mutacionais foi publicada, atualizando o banco de dados do COSMIC para versão 3.2 que compreende quase 100 assinaturas mutacionais de substituições de base única (SBS), entre outras de substituições duplas (DBS) e assinaturas de pequenas inserções/deleções (ID) (Alexandrov et al. 2020).

Entre as etiologias conhecidas estão os processos associados à idade (desaminação espontânea de 5-metilcitosina), falha em mecanismos de reparo de DNA, ação de enzimas da família APOBEC (*Apolipoprotein B mRNA Editing Enzyme Catalytic Polypeptide-like*), tabagismo, radiação UV (ultravioleta), exposição a quimioterápicos, entre outros (Alexandrov et al. 2020). Estes últimos caracterizam processos exógenos, pela exposição a fatores mutagênicos como o tabaco, quimioterápicos e radiação.

O processo mutacional associado a idade, que caracteriza a primeira assinatura mutacional identificada, é observada em todos os subtipos tumorais (assinatura A no estudo de Nik-Zainal et al. (2012) (Figura 8); assinatura 1 (Figura 9) ou SBS1, como conhecida atualmente) e pode ser considerado como um dos principais processos endógenos não patológicos. Outros processos endógenos são as falhas nos mecanismos de reparo do DNA e ação de enzimas da família APOBEC. Os processos mutacionais endógenos possuem importante papel na instabilidade genômica, facilitando a ocorrência de mutações ou a não correção destas, que por

[1] https://cancer.sanger.ac.uk/cosmic/signatures



Fonte: https://cancer.sanger.ac.uk/signatures/signatures_v2/

Figura 9 – Distribuição das 30 assinaturas mutacionais por subtipo tumoral. Matriz de assinaturas da versão 2 do COSMIC. A assinaturas eram denominadas *Signatures 1-30* e consideravam apenas mutações pontuais (Alexandrov et al. 2015).

sua vez, como já comentado, auxiliam no processo de carcinogênese, aumentando as chances de alterações em genes *drivers*.

A atividade de enzimas APOBEC consiste na edição de DNA e pode ser induzida em resposta a infecções virais, por exemplo. A desregulação desta função pode ser um mecanismo de oncogênese associado a vírus, como sugerido para EBV em câncer gástrico (Bobrovnitchaia et al. 2020) e HPV em câncer cervical e câncer de cabeça e pescoço (Revathidevi et al. 2021; Cannataro et al. 2019). No contexto das assinaturas mutacionais, as assinaturas SBS2 e SBS13 estão associadas à atividade de APOBEC (Alexandrov et al. 2020).

Estudos mais recentes associaram as assinaturas mutacionais ao fator prognóstico de câncer (Wang et al. 2018; Ma et al. 2018). Apesar de alguns estudos considerarem apenas a alta carga mutacional como um fator associado com melhor prognóstico ou resposta a certos tratamento, como a imunoterapia, outros

apontam para tipos específicos de mutações como mais eficazes na predição de resposta à imunoterápicos (Hugo et al. 2016; Smid et al. 2016). Considerando as assinaturas relacionadas a APOBEC, por exemplo, Wang et al. (2018) observaram estas assinaturas como marcadora e preditiva de resposta à imunoterapia em tumores de pulmão de células não-pequenas (Wang et al. 2018).

1.2.2 Mecanismos de reparo

Outro tipo de processo mutacional endógeno é aquele associado à falha dos mecanismos de reparo do DNA. Encontram-se bem documentados pelo menos seis tipos de vias de reparo que são ativadas em cenários específicos para minimizar os danos ao DNA (Tabela 1) (Torgovnick e Schumacher 2015).

Tabela 1 – Mecanismos de reparo do DNA

Mecanismos de reparo	**Cenário**
Base Excision Repair (BER)	Lesões oxidativas (espécies reativas de oxigênio)
Nucleotide Excision Repair (NER)	Distorção da dupla hélice de DNA (radiação UV)
Mismatch Repair (MMR)	Erros de replicação
Single Strand Break Repair (SSBR)	Quebras de fita simples (radiação ionizante, espécies reativas de oxigênio)
Homologous Recombination Repair (HRR)	Quebras da fita dupla (radiação ionizante, espécies reativas de oxigênio)

A importância dos mecanismos de reparo do DNA é destacada pela existência de muitas condições clínicas e síndromes causadas por defeitos nos genes dessas vias (Ma et al. 2018; Torgovnick e Schumacher 2015). Por exemplo, defeitos no mecanismo NER podem causar o xeroderma pigmentoso (doença autossômica recessiva) e defeitos nos genes *ATM* e *ATR*, relacionados ás lesões de fita simples e dupla, estão associados a ataxia-telangiectasia, uma doença de imunodeficiência primária de herança recessiva. Por outro lado, quando não a disfunção desses genes não afeta ambas cópias, o que acontece é uma predispo-

sição a diferentes tipos de câncer. Falhas na via NER predispõe ao câncer de pele; mutações no gene *ATM* aumentam o risco de câncer de mama, colorretal e de estômago (Thompson et al. 2005; Paglia et al. 2009); mutações nos genes *BRCA1* e *BRCA2*, da via HRR, predispõe ao câncer de mama e ovário (Davies et al. 2017).

Mesmo as mutações nos genes de mecanismos de reparo do DNA estarem associadas a predisposição ao câncer e portanto constituir o evento inicial para transformação celular, após estabelecido, esses mecanismos também podem indicar uma fraqueza do tumor uma vez que, o gene mutado é frequentemente necessário para a célula cancerosa manter sua própria sobrevivência (Furgason e Bahassi 2013). Em outras palavras, as alterações destas vias podem ser utilizadas como marcadores de um determinado fenótipo ou mesmo ser utilizadas como alvos de terapia.

Semelhante a APOBEC, muitas assinaturas mutacionais já foram associadas aos defeitos nos mecanismos de reparo ao DNA e podem ser utilizadas como fatores prognósticos. Por exemplo, a assinatura mutacional SBS3 (antiga CS-3) já foi associada a mutações nos genes *BRCA1/BRCA2* (genes do mecanismo HRR) e também observada como preditora da resposta ao tratamento com inibidores de PARP (Davies et al. 2017; Van Hoeck et al. 2019). Pacientes com falhas nos mecanismos NER, que possuem mutações no gene *ERCC2* (perfil com assinatura mutacional SBS5 ou SBS8 (Van Hoeck et al. 2019)), também podem ser mais sensíveis a certos tratamentos (Kim et al. 2016; Jager et al. 2019) e pacientes com falha na via BER (assinaturas SBS18 ou SBS30) podem ser beneficiados com imunoterapia (Volkov et al. 2019).

Outro importante mecanismo de reparo do DNA com pelo menos quatro assinaturas mutacionais associadas é o MMR, relacionado às assinaturas 6,15, 20 e 26. O mecanismo MMR age nas correções de mutações que ocorrem durante a replicação celular, e por isso, considerando que uma das marcas do câncer é a proliferação celular, a falha neste mecanismo tem grande importância na predis-

posição ao câncer. A falha no mecanismo de MMR está associada à síndrome de Lynch e é um fator de risco para câncer gastrointestinal (estômago e colorretal), endometrial e diversos outros subtipos (Sehgal et al. 2014). Geralmente mutações inativadoras nos genes da via MMR também estão associadas ao fenótipo tumoral de instabilidade de microssatélites (MSI) (Boland e Goel 2010).

Microssatélites são sequências repetitivas e curtas de DNA, distribuídas difusamente pelo genoma humano (Leung et al. 2000). O fenótipo MSI está associado a falha no mecanismo de MMR, pois sendo uma região repetitiva, o acúmulo de erros durante a replicação, como inserções ou deleções de algumas bases dessas sequências acontece em alta frequência, uma vez que não podem ser corrigidos (Arzimanoglou et al. 1998; Fang et al. 2012). Um dos métodos de detecção de MSI é por meio da imunohistoquímica de quatro proteínas da via MMR (MSH2, MSH6, MLH1 e PMS2) (Boland et al. 1998).

Outra forma de avaliar este fenótipo e a falha no mecanismo de MMR é por meio do perfil hipermutado que pode ser mensurado pela quantidade de mutações no genoma tumoral, por meio do *Tumor Mutation Burden* (TMB). No entanto, a definição de um ponto de corte não é bem estabelecida neste caso (Samstein et al. 2019). De qualquer forma, identificar este cenário de falha no mecanismo de MMR é promissor considerando que pacientes com este perfil podem ser beneficiados com imunoterapia. Alguns estudos já demonstraram que tumores (glioblastomas, gastrointestinais, endometrial e de próstata) com falha em MMR são mais responsivos a tratamento com anti-PD1 (Abida et al. 2019; Reichert et al. 2019; Le et al. 2017). Além da resposta à terapia, a relevância clínica deste perfil molecular é bem estabelecida em câncer colorretal associado ao melhor prognóstico (Buckowitz et al. 2005; Benatti et al. 2005).

Uma das características que podem explicar o melhor prognóstico desses subtipos tumorais é que a grande quantidade de mutações está diretamente relacionada a maior quantidade de neoantígenos. Neoantígenos são definidos como novos antígenos produzidos por mutações não sinônimas, ou seja, mutações que

alteram a proteína final sintetizada. Estes novos antígenos podem ser mais facilmente reconhecidos como estranhos pelo nosso sistema imunológico. De fato, tumores associados a falha no mecanismo de MMR parecem induzir respostas inflamatórias mais expressivas (Maby et al. 2021; Zhang et al. 2021; Wang et al. 2016).

1.3 CÂNCER GÁSTRICO

1.3.1 Epidemiologia

O câncer gástrico (CG) é o quinto tipo mais frequente de câncer e o quarto com maior taxa de mortalidade no mundo. É esperado mais de 1 milhão de novos casos e uma estimativa de quase 770 mil óbitos em 2020 [2]. Embora tenha sido observada uma queda na incidência global de CG desde a década de 1980, devido às melhorias no armazenamento de alimentos e controle da infecção por *Helicobacter pylori* (*H. pylori*), as taxas de mortalidade se mantiveram relativamente constantes (em torno de 8-10% de óbitos por câncer) (Bray et al. 2018; Torre et al. 2015; Ferlay et al. 2010; Parkin et al. 2005).

Além da diminuição na incidência de CG (do mais incidente antes de 1975 ao quinto mais incidente em 2020) (Bray et al. 2018), é possível observar que houve também uma mudança nas características epidemiológicas e incidência geográfica de CG, sendo que atualmente países em desenvolvimento concentram a maioria dos casos. Além disso, o gênero e lugar de moradia influenciam o número de casos de CG, sendo 16 vezes mais alta entre homens chineses quando comparado ao acometimento em mulheres da América do Norte ou da maioria dos países africanos. Além dos países asiáticos, a incidência de CG é mais alta em países da América do Sul e Leste Europeu (Torre et al. 2015) e, de forma geral, está associada com a idade avançada, sendo mais prevalente em pacientes com idade por volta dos 60-70 anos.

[2] https://gco.iarc.fr/today/home

A mudança geográfica no perfil epidemiológico da doença, dos países mais desenvolvidos para os em desenvolvimento, também vem sendo acompanhada por alterações na distribuição da localização do tumor primário na anatomia do estômago. Há um aumento expressivo nos tumores que surgem na proximidade da transição esôfago-gástrica, também chamada cárdia ou região proximal, em detrimento dos tumores da porção distal (fundo) do estômago (Anderson et al. 2010). Essas mudanças provavelmente ocorreram devido a dinâmica dos elementos carcinogênicos envolvidos na tumorigênese do CG, sendo que os tumores de região distal do estômago estão associados à infecção crônica por *H. pylori* (que teve sua prevalência diminuída) enquanto que os tumores proximais estariam relacionados à obesidade e dieta rica em sal (Bray et al. 2018).

No Brasil, CG aparece em terceiro lugar na incidência entre os homens e em quinto entre as mulheres [3], com variações de acordo com a região do país. A maior incidência de CG no Brasil ocorre nas regiões Norte e Nordeste, onde o CG é o segundo e terceiro tipo de câncer mais incidente em homens, respectivamente (Ministério da Saúde e Instituto Nacional de Câncer José Alencar Gomes da Silva (INCA) 2019; Ministério da Saúde e Instituto Nacional de Câncer José Alencar Gomes da Silva (INCA) 2017; Ministério da Saúde e Instituto Nacional de Câncer José Alencar Gomes da Silva (INCA) 2015). Segundo o INCA, a estimativa de novos casos para cada ano do triênio 2020-2022 foi de 21.230, sendo 13.360 casos novos de câncer de estômago entre homens e 7.870 nas mulheres, e o número de mortes em 2018 foi de 14.761, sendo 9.387 em homens e 5.374 mulheres (Ministério da Saúde e Instituto Nacional de Câncer José Alencar Gomes da Silva (INCA) 2019) [3].

1.3.2 Aspectos clínicos e moleculares

O câncer gástrico é um tumor sólido caracterizado por complexas interações genéticas e ambientais. Classicamente, a infecção por *H. pylori*, obesidade,

[3] https://www.inca.gov.br/tipos-de-cancer/cancer-de-estomago

refluxo gastroesofágico e tabagismo são os principais agentes conhecidos na sua oncogênese. Há fatores genéticos de risco que são herdáveis, porém, a maioria dos fatores genéticos associados ao CG são adquiridos ao longo da vida e não ocorrem de forma isolada, mas são cumulativos ao longo do processo de carcinogênese (Lordick e Janjigian 2016; McLean e El-Omar 2014), o que também caracteriza o CG como uma doença altamente heterogênea. Outros fatores de risco associados ao CG são consumo de bebidas alcoólicas, dieta rica em sal e gorduras, baixo consumo de frutas e vegetais, anemia perniciosa e doenças crônicas gástricas (entre elas, algumas síndromes hereditárias) (Wadhwa et al. 2013).

O *H. pylori* é considerado como carcinógeno tipo 1 segundo a Organização Mundial de Saúde (OMS), podendo aumentar o risco de desenvolver CG de até 6X. Porém, há baixa penetrância considerando que menos de 1% das pessoas infectadas desenvolvem CG. Isso significa que a inter-relação entre os fatores virulentos do patógeno e a predisposição genética do indivíduo são determinantes para desenvolvimento de CG (Lordick e Janjigian 2016). Ainda assim, a infecção por *H. pylori* é considerado importante fator de risco visto que contribui para instalação de quadro de inflamação crônica no tecido gástrico além de promover modulações epigenéticas e produzir oncogenes, como CagA, que podem induzir transformação das células gástricas por meio de diversas vias (Wadhwa et al. 2013). Além de *H. pylori*, o CG também pode ser associado com infecção por Epstein-Barr vírus (EBV) (Lordick e Janjigian 2016).

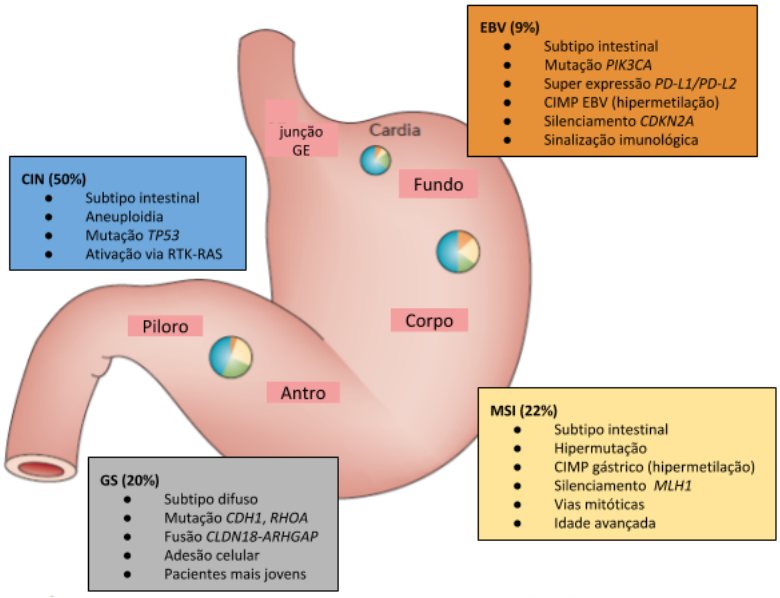
A forma histológica predominante do CG é o adenocarcinoma gástrico (AdG), responsável por aproximadamente 95% dos casos. Existem diferentes tipos de classificação de AdGs: considerando seus (a) aspectos histológicos que, por sua vez, é a classificação mais utilizada na prática clínica e, mais recentemente, (b) aspectos moleculares, que caracterizam a alta heterogeneidade genética do CG (Cristescu et al. 2015; Li et al. 2016). A classificação histológica clássica de Lauren (1965) (Lordick e Janjigian 2016; McLean e El-Omar 2014; Wadhwa et al. 2013), subdivide os AdGs em dois subtipos principais: difuso e intestinal,

além dos subtipos mistos e indeterminados (não-classificáveis).

O subtipo intestinal caracteriza-se por células tumorais que se arranjam em estruturas tubulares ou similares às de glândulas, além de haver limitado conteúdo estromal. A carcinogênese do subtipo intestinal está relacionada à progressão sistemática a partir de estágios pré-neoplásicos, com a evolução progressiva de alterações pré-malignas como gastrite atrófica, evoluindo para metaplasia intestinal e displasia, envolvendo inflamação crônica e geralmente associada à infecção por *H. pylori*. O subtipo difuso, por sua vez, caracteriza-se por células pouco diferenciadas ou indiferenciadas intercaladas com células do estroma. A carcinogênese do subtipo difuso estaria relacionada a processos de perda de adesão celular, células em anel de sinete e sem infecção por *H. pylori* (Lordick e Janjigian 2016; McLean e El-Omar 2014; Wadhwa et al. 2013; PDQ Adult Treatment Editorial Board 2002).

A classificação molecular proposta pelo grupo do *The Cancer Genome Atlas* (TCGA) considera a grande complexidade mutacional, cariotípica e de expressão gênica/proteica, além de instabilidades cromossômicas e de microssatélites (Bass et al. 2014). A caracterização de 295 amostras de AdGs (TCGA) definiu que o CG não é uma entidade única e pode ser subdividida em 4 grupos (Figura 10) denominados EBV (Epstein-Barr vírus), MSI (*MicroSatellite Instability*), CIN (*Chromosomal Instability*) e GS (*Genomically Stable*) Figura 10).

O subtipo denominado EBV (9% dos casos) está associado à infecção por Epstein-Barr vírus, caracteriza-se por hipermetilação genômica, maior tendência de ocorrência no fundo ou corpo gástrico (62%, p=0,02) e 80% dos casos carregam mutação em *PIK3CA*. O subtipo denominado como MSI (22% dos casos) foi mais frequente em pacientes com idade mais avançada (72 anos, p=5x10-5), apresenta hipermetilação (padrão distinto do subtipo EBV) e altas taxas de mutação somática nos genes *PIK3CA* (42%), *ERBB3* (26%) e ambos (12%). O subtipo denominado CIN (50% dos casos) caracteriza-se por aneuploidia e amplificação de genes de receptores tirosina quinase (RTK), principalmente *ERBB2* (HER2)



Fonte: Adaptado de: Lordick e Janjigian (2016)

Figura 10 – Subtipos moleculares de câncer gástrico de acordo com *The Cancer Genome Atlas* (TCGA). O câncer gástrico pode ser classificado em 4 tipos considerando seus aspectos moleculares de acordo com estudo do TCGA (Bass et al. 2014). Os aspectos moleculares consideram padrão de metilação, mutações somáticas, expressão gênica, e infecção por EBV. GE: gastro esofágico; CIMP: *CpG island methylation phenotype*; CIN: chromossome instability; GS: *genomically stable*; MSI: *microsatellite instability*.

(24%) e *VEGFR2*. Esses três casos estão mais associados ao subtipo histológico intestinal.

Finalmente, o subtipo GS (aproximadamente 20% dos casos), que foi mais frequente em pacientes mais jovens (59 anos, p=4x10^{-7}) e predominante no subtipo histológico difuso (40/55 = 73%, p=8x10^{-7}); não apresenta aneuploidias, altas taxas de mutação ou hipermetilação. Cerca de 30% do subtipo GS apresentam novas alterações nos componentes de sinalização da via Rho (gene *RHOA* ou fusão com genes de proteínas ativadoras GTPase) (Bass et al. 2014; Li et al. 2016; Lordick e Janjigian 2016) [4].

A maioria dos casos de AdGs ocorre esporadicamente, sendo que até 3% dos casos estão relacionados às síndromes hereditárias e variantes genéticas de origem germinativa (Lordick e Janjigian 2016; McLean e El-Omar 2014). Entre

[4] https://cancergenome.nih.gov/cancersselected/stomachcancer

as principais mutações germinativas associadas a casos de CG familiar estão alterações nos genes *CDH1*, *PLCE1* e *PSCA*, este último associado ao subtipo difuso (Wadhwa et al. 2013). Adicionalmente, algumas síndromes associadas ao câncer hereditário são:

- HDGC (*Hereditary Diffuse Gastric Cancer*), de caráter autossômico dominante caracterizado por mutação com perda de função em *CDH1* (que codifica E-caderina) ou *CTNNA1* (E-catenina);
- GAPPS (*Gastric Adenocarcinoma and Proximal Polyposis of the Stomach*);
- HNPCC (*Hereditary Nonpolyposis Colon Cancer*) ou Síndrome de Lynch, que aumenta o risco de AdG subtipo intestinal de 2 a 19X, geralmente associado a mutações em genes relacionados a reparação de DNA da via MMR, como *MSH2* e *MLH1*;
- Síndrome de Li-Fraumeni, que possui mutação germinativa em *TP53*;
- Síndrome Peutz-Jeghers, devido às alterações encontradas no gene *STK11* que regula a via de mTOR, importante na diferenciação e ativação de linfócitos T; e
- Polipose Adenomatosa Familiar, caracterizada por mutação no gene *APC*, que aumenta o risco de transformação maligna (McLean e El-Omar 2014).

O CG é considerado uma doença altamente heterogênea com diferentes entidades histológicas e moleculares reconhecidas. O processo de tumorigênese está baseado no acúmulo de diferentes anormalidades genéticas que desregula as vias de sinalização celular e altera o perfil de interação entre as células (como do tumor com próprio hospedeiro). Sendo assim, a capacidade de resposta antitumoral também é variável entre os pacientes, o que influencia diretamente o prognóstico e a resposta à terapia.

Com isso, nesse contexto, além da caracterização histológica e molecular do tumor, a caracterização do infiltrado inflamatório e do perfil de resposta imunológica presente no microambiente tumoral, todos são fatores que devem ser integrados e considerados a fim de auxiliar no entendimento da relação tumor-hospedeiro, podendo contribuir na resolução dos desafios ainda existentes relacionados ao CG.

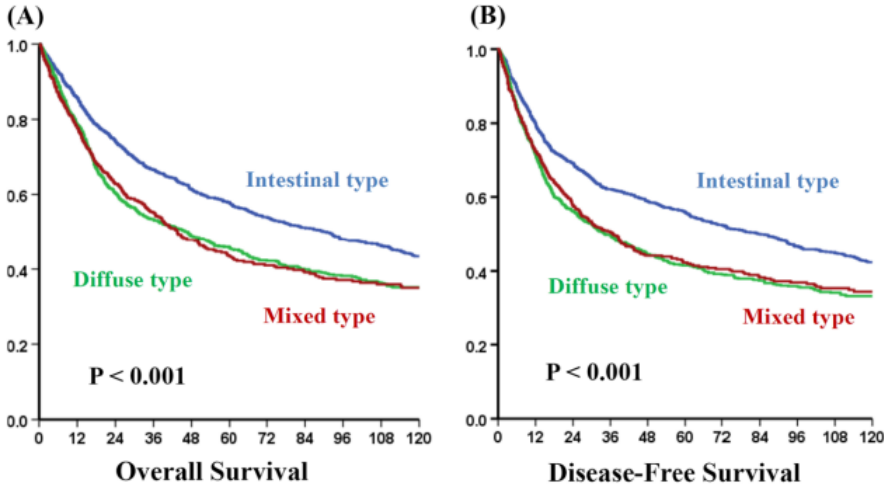
1.3.3 Prognóstico e tratamento

O mal prognóstico do CG está intimamente relacionado ao diagnóstico tardio de uma doença sem sintomas de alerta nas fases precoces. Na época do diagnóstico, a maioria dos pacientes apresenta queixas incaracterísticas, como náusea, alteração de apetite, perda de peso e dor abdominal, que podem passar despercebidas e não serem consideradas suspeitas dessa neoplasia. Assim, a doença é descoberta já em fase avançada, quando as chances de cura são baixas ou mesmo inexistentes [5].

Um dos maiores desafios em oncologia é estratifcar os pacientes de forma eficiente no que tange a resposta à terapia e prognóstico. Em câncer gástrico, os subtipos histológicos de Lauren, classificação comumente utilizada na prática clínica, estão relacionados ao prognóstico, no sentido de que o subtipo intestinal geralmente apresenta sobrevida global e sobrevida livre de doença mais longa do que o subtipo difuso. Tendo como base uma coorte coorte asiática, o prognóstico associado ao subtipo intestinal também é melhor do que aquele associado ao subtipo misto (Chen et al. 2016) (Figura 11). Outra variável clínica associada ao prognóstico em câncer gástrico é o sítio anatômico do tumor. Em geral, tumores gástricos mais distais têm sobrevida mais longa (Figura 12) (Zhao et al. 2016).

De importância, a principal característica clínica que se apresenta como importante fator preditor em câncer gástrico é o estadiamento. Incluisive, as relações prognósticas observadas para os subtipos histológicos e sítio tumoral são

[5] https://www.inca.gov.br/tipos-de-cancer/cancer-de-estomago



Fonte: Chen et al. (2016)

Figura 11 – Sobrevida global e livre de doença entre os subtipos histológicos de Lauren em coorte chinesa. O subtipo intestinal apresenta melhor (A) sobrevida global e (B) livre de doença do que os subtipos difuso e misto; teste de log-rank p<0,001.

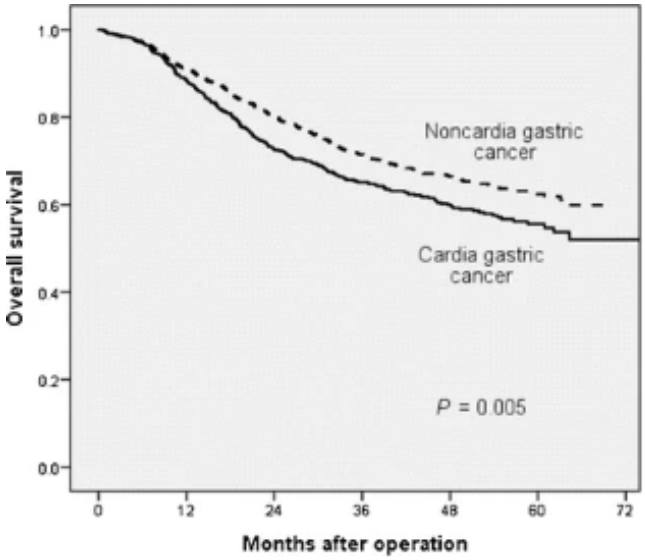
minimizadas quando é considerado o estadiamento da doença. Pode-se observar diferença no prognóstico entre os subtipos histológicos apenas no estágio inicial (Figura 13). Já para o sítio anatômico, nenhuma diferença é obervada quando considerado o estadiamento (Figura 11 e 12).

Considerando a diversidade molecular em câncer gástrico, mesmo com essas variáveis clínicas relevantes, ainda restam dúvidas sobre como estratiicar melhhor os pacientes em relação ao prognóstico, além de relacionar a melhor estratégia terapêutica.

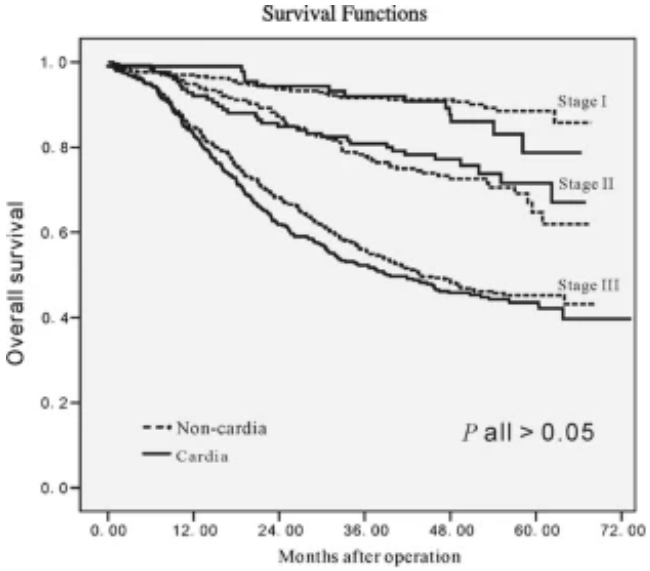
O principal tratamento utilizado para doença não metastática é a cirurgia, retirando parte ou todo o estômago, além de linfonodos próximos, sendo esta considerada estratégia associada como a melhor chance de cura. A radioterapia e a quimioterapia (terapias convencionais) são consideradas como tratamentos secundários que podem complementar a cirurgia [6].

Outro tratamento importante para CG é o medicamento *trastuzumabe* que pode ser associado à quimioterapia no tratamento da doença metastática, cujo

[6] https://www.inca.gov.br/tipos-de-cancer/cancer-de-estomago



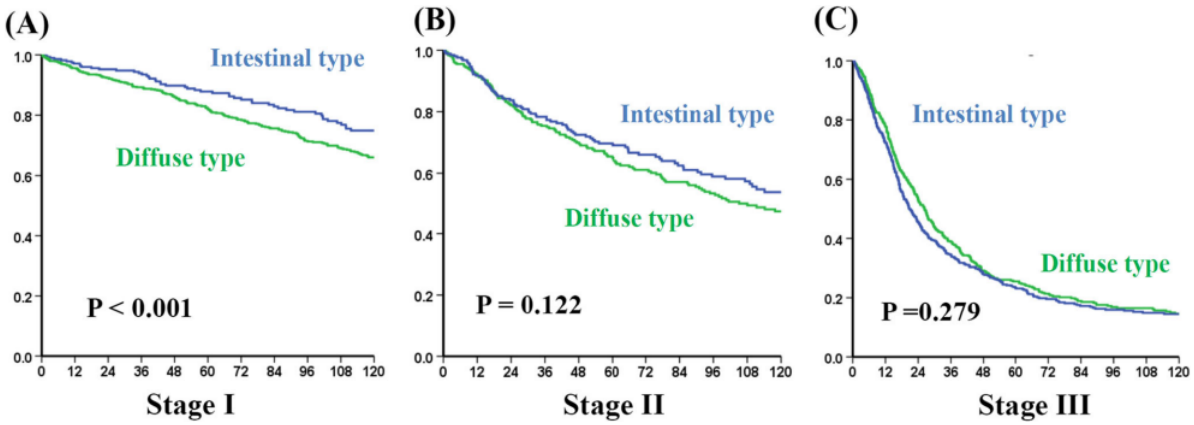
(a) Cárdia e não cárdia



(b) Cárdia e não cárdia por estadiamento

Fonte: Zhao et al. (2016)

Figura 12 – Sobrevida global associada a tumores gástricos localizados na cárdia e não cárdia. (a) Tumores localizados na cárdia estão associados a pior sobrevida global quando comparados àqueles localizados na cárdia. (b) Do ponto de vista de estadiamento, estágios mais precoces apresentam sobrevida global melhor independente do sítio anatômico do tumor; teste log-rank p>0,05.



Fonte: Chen et al. (2016)

Figura 13 – Sobrevida global entre os subtipos histológicos de Lauren em coorte chinesa por estadiamento. Não há diferença significativa entre os subtipos intestinal e difuso para os estágio II e III; teste de log-rank p>0,05.

alvo é a proteína HER2, a qual está altamente expressa nas células tumorais em cerca de 22% dos casos (Lordick e Janjigian 2016; McLean e El-Omar 2014). Mais recentemente, outro medicamento, o *ramucirumabe*, que tem como alvo a proteína VEGFR2, associada a processos de angiogênese, foi incorporada às opções para o tratamento da doença metastática. Além de não serem indicados para todos os pacientes, o desenvolvimento de resistência ao tratamento, devido, por exemplo, a heterogeneidade tumoral, limita a duração da resposta (Lordick e Janjigian 2016).

Além das abordagens discutidas acima, a imunoterapia representa uma nova opção de tratamento para os pacientes com câncer gástrico avançado ou metastático. Mais recentemente, a utilização como tratamento de primeira linha, combinado com quimioterapia, foi aprovado no FDA (FDA a; Chao ; Staff). Em 2020, foi aprovado o tratamento de primeira linha de pacientes com câncer colorretal irressecável ou metastático com instabilidade de microssatélites (MSI-H) ou com deficiência mecanismo de reparo (dMMR) sem a combinação com quimioterapia (FDA b).

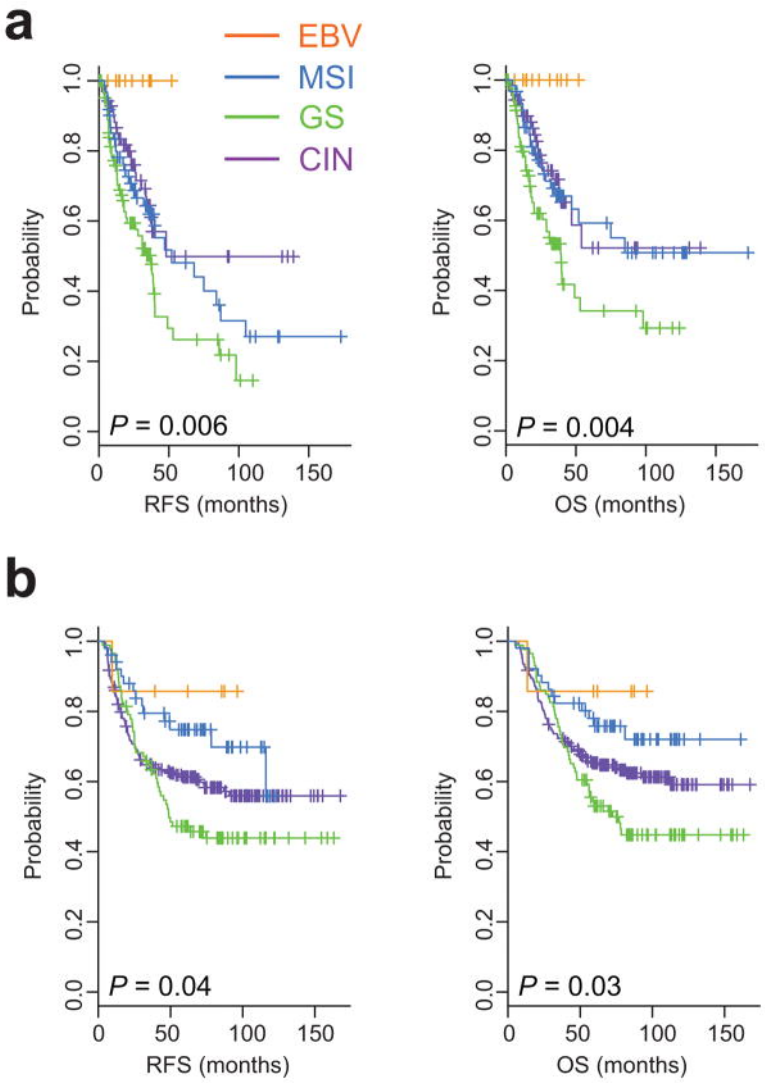
Diante desta nova abordagem de tratamento, entender a composição do

infiltrado inflamatório no microambiente tumoral tem se mostrado importante. Foi observado que nas entidades EBV e MSI, associados ao subtipo histológico intestinal, há forte assinatura imunológica com infiltrado inflamatório do tipo linfóide abundante, devido provavelmente ao caráter mais imunogênico destes subtipos, seja pela inflamação induzida pela infecção, no caso de EBV, seja pela alta taxa mutacional e produção de neoantígenos (característico de MSI). Perfil diferente foi encontrado para as entidades GS (relacionado ao subtipo histológico difuso) e CIN (intestinal), que têm característica menos imunogênica.

Esses subtipos moleculares também estão relacionados ao prognóstico, com os subtipos EBV e MSI apresentando sobrevida global mais longa e GS o pior prognóstico (Figura 14). Diferente de outras características clínicas, como a classificação de Lauren e sítios anatômicos, já foi proposto que a classificação em subtipos moleculares seria um fator de preditivo mais significativo do que o estadiamento, principalmente o estadiamento de tamanho do tumor (T) e acometimento linfonodal (N).

O infiltrado linfóide, associado aos subtipos EBV e MSI, sugere melhor prognóstico e resposta à imunoterapia, porém estudos com anti-CTLA4 falharam e com anti-PDL1 a resposta foi variável (Wang et al. 2016; Panda et al. 2018) sugerindo que, além da caracterização do tipo de infiltrado, é necessário uma avaliação ampla da composição celular e orientação funcional no microambiente tumoral além da associação com as demais variáveis clínicas e moleculares para melhor entendimento e aplicação de estratégias imunoterápicas no tratamento de CG.

Em resumo, existem características clínicas e moleculares importantes em câncer gástrico que estão associadas a diferentes prognósticos. No entanto, ainda é um desafio estratificar os pacientes de forma preditiva quanto a resposta a terapia. Considerando o avanço da imunoterapia nos últimos anos, a aplicação do contexto imunológico do microambiente tumoral pode ser o fator chave na determinação de pacientes respondedores e não respondedores.



Fonte: Sohn et al. (2017)

Figura 14 – Sobrevida global e livre de recorrênica associada aos quatro subtipos moleculares de câncer gástrico (EBV, MSI, CIN e GS) em duas coortes independentes. (a) coorte *MD Anderson Cancer Center* (MDACC); (b) coorte *Samsung Medical Center* (SMC; Seoul, Korea). Em ambas coortes foram considerados apenas pacientes em estágio II ou III. É possível observar que o subtipo EBV apresenta melhor prognóstico enquanto que o subtipo GS está associado a pior prognóstico em ambas coortes. As relações prognósticas dos subtipos MSI e CIN foram variáveis em entre as coortes.

2 JUSTIFICATIVA

Considerar o contexto imunológico do câncer é imprescindível para entender quais os mecanismos de escape estão atuando no microambiente tumoral e qual a melhor estratégia terapêutica pode trazer melhor benefício ao paciente. Para entender esse microambiente, o desenvolvimento de um painel capaz de determinar os tipos celulares existentes na amostra de forma sensível, diferenciando as células do sistema imunológico das células estromais e tumorais, a partir dos dados de RNA-Seq, permanece um problema a ser resolvido. Embora este trabalho tenha como alvo principal a caracterização do infiltrado em câncer gástrico, estabelecer uma estratégia de análise do infiltrado em amostras de tecido poderá favorecer e ser aplicada em outros tumores e situações clínicas, como doenças autoimunes e infecções.

Em CG, embora o estadiamento pareça ser a característica que melhor estratifica os pacientes em relação ao prognóstico, avaliar o microambiente tumoral de uma forma mais integrada e ampla, considerando seus aspectos imunológicos e evolutivos, pode trazer uma nova visão sobre as perspectivas no prognóstico dessa doença.

3 OBJETIVOS

3.1 OBJETIVO GERAL

Descrever a composição de células do infiltrado e sua atividade no microambiente tumoral de AdGs com base nos dados de perfil de expressão gênica utilizando, para tanto, métodos analíticos de deconvolução.

3.2 OBJETIVOS ESPECÍFICOS

- Definir um painel de genes marcadores para distinguir tipos celulares presentes no infiltrado tumoral, considerando perfil de expressão gênica associado a orientação funcional;
- Descrever composição de células do infiltrado inflamatório do tumor primário nos diferentes subtipos histológicos e moleculares de adenocarcinoma gástrico;
- Correlacionar a composição do infiltrado tumoral variáveis clínicas e moleculares e sua relação com prognóstico;
- Propor *pipeline* que permita uma avaliação global da imunologia do câncer.

4 METODOLOGIA

4.1 DADOS CLÍNICOS E MOLECULARES

Os dados clínicos e moleculares de 439 pacientes de adenocarcinoma gástrico foram obtidos do banco de dados *The Cancer Genome Atlas* (TCGA). Os níveis de expressão gênica, normalizados via FPKM[1], estavam disponíveis para 380 pacientes. Para a análise das assinaturas mutacionais, utilizamos as variantes somáticas de uma única base detectadas previamente em estudos anteriores além da coorte do TCGA (Bass et al. 2014; Kakiuchi et al. 2014; Wang et al. 2014; Ciriello et al. 2013; Wang et al. 2011). Mais detalhes estão descritos no APÊNDICE A.

4.2 ESTIMAÇÃO DO MICROAMBIENTE TUMORAL E CONTEXTO IMUNOLÓGICO

4.2.1 Genes marcadores e construção da matriz de assinatura

Para aplicação de métodos de deconvolução é necessária uma lista de genes marcadores dos alvos de interesse. Nesse sentido, primeiramente foi realizado um levantamento sistemático de listas de genes marcadores de células imunológicas. Devido aos primeiros estudos avaliarem a expressão gênica por meio da metodologia de *microarray*, muitos marcadores eram referenciados por sondas (probes) e, dessa forma, foi necessário realizar uma reanotação do número de acesso de cada gene para que as listas pudessem ser comparadas entre si. Consideramos os aliases dos genes aos respectivos *probes* (*Affymetrix GeneChip Human Genome U133 Plus 2.0 Array*)[2]. Os genes com nomes duplicados, descontinuados, não encontrados ou sem nenhuma anotação foram removidos. Ao final, as listas atualizadas com a anotação correta foram utilizadas para construção de matri-

[1] Fragments Per Kilobase Million
[2] http://refdic.rcai.riken.jp/dataset.cgi?form=download

zes de assinatura de referência a fim de aplicar em conjunto com métodos de deconvolução.

Para a construção da matriz de assinaturas, além dos genes marcadores obtidos com o levantamente prévio, utilizamos dados de expressão gênica das células de interesse que foram purificadas. Estes dados foram obtidos a partir de uma coleção de amostras previamente analisadas pela plataforma de *microarray* Affymetrix *HGU133plus2* denominada **HPCA** (*Human Primary Cell Atlas*). O **HPCA** é um banco abrangente que inclui mais de 100 estudos (Mabbott et al. 2013), totalizando 385 amostras de células imunes. A partir das informações contidas nesse banco, agrupamos as amostras em 8 grupos de células imunes considerando a mediana dos níveis de expressão gênica para cada gene.

Ao final, foram extraídas as listas de genes marcadores dessa matriz de referência para 8 células. Para visualizar os resultados, utilizamos um mapa de calor (*heatmap*) com os valores de expressão normalizados com o $z - score$. Foi utilizado o pacote "pheatmap", disponível no ambiente R (www.r-project.org), para construção de *heatmap* de cada lista.

4.2.2 Avaliação de desempenho dos métodos de deconvolução

Para avaliação dos métodos de deconvolução, inicialmente foram utilizados dados de expressão gênica (também normalizados via FPKM) de 13 subtipos de linfócitos distribuídos em 63 amostras (Bonnal et al. 2015) de subtipos conhecidos.

Análises preliminares foram realizadas a fim avaliar diferentes métodos de deconvolução utilizando a coorte dos 13 subtipos. As proporções de células encontradas foram então comparadas com o subtipo esperado de cada amostra. Com esta primeira avaliação, foram definidos os métodos de deconvolução para aplicação nas amostras de AdG em um segundo momento. Assim, utilizamos as amostras de adenocarcinoma gástrico (AdG) disponíveis no **TCGA** (TCGA-STAD).

A primeira estratégia foi a comparação entre as proporções estimadas das células imunes detectadas pelos métodos de deconvolução com os resultados obtidos das mesmas amostras de AdG caracterizados previamente(Saltz et al. 2018). Nesse estudo, foi utilizado inteligência artifical, especificamente aprendizado profundo, para estimar as proporções de linfócitos, neutrófilos, mastócitos, eosinófilos, macrófagos e células dendríticas diretamente a partir de imagens histopatológicas. O desempenho dessas abordagens foi avaliado pelo erro quadrático médio (**MSE**, *Mean Squared Error*), que basicamente calcula a distância da estimativa em relação à proporção esperada, para cada tipo celular. Esta análise foi feita utilizando a função *mse* do pacote *hydroGOF*, disponível no ambiente R. Quanto menor o valor de **MSE**, mais similar é a estimativa das proporções avaliadas.

Na segunda estratégia de avaliação dos métodos de deconvolução em amostras de AdG, utilizamos a correlação linear de Pearson entre as quantidades de células estimadas e a expressão de genes marcadores principais, descritos na Tabela 2. A expressão dos genes foi normalizada pela quantidade de infiltrado tumoral estimada pelo método ESTIMATE (Yoshihara et al. 2013), considerando o valor de 1-pureza do tumor. Somente as correlações com $p<0,05$ foram consideradas como significativas.

Tabela 2 – Genes marcadores utilizados para avaliação da estimação de células imunes.

Célula	**Genes**
Linfócito B	*MS4A1* (CD20), *CD19*
Linfócito T CD4	*CD28*, *CD4*, *CD80*, *CD86*
Linfócito T regulador (Treg)	*FOXP3*
Linfócito T CD8	*CD8A*, *PRF1*
Células NK	*GZMB*, *KIR2DL4*
Macrófagos	*CD14*, *CD68*, *CD163*
Células dendríticas	*CD1A*
Neutrófilos, Mastócitos	*CXCR1*, *CXCR2*

A terceira abordagem considerou os resultados obtidos por meio de uma análise de enriquecimento (*gene set enrichment analysis*) calculados pelo software

xCell (Aran et al. 2017). Este método, inclui 64 tipos celulares (considerando células imunológicas, progenitoras e estromais). Esse enriquecimento resulta em *scores* associados às proporções das células na amostra. Novamente, utilizamos correlação linear de Pearson para comparar os resultados obtidos pelos métodos de deconvolução aos *scores* gerados pelo método xCell (Aran et al. 2017).

4.3 ANÁLISES ESTATÍSTICAS

Para as comparações da composição de células entre grupos de pacientes com AdG (subtipos histológicos e moleculares), foram utilizados os testes não paramétricos de U de Mann-Whitney (para comparação de dois grupos) e Kruskal Wallis (para comparação de mais de dois grupos). As curvas de sobrevida global foram obtidas pelo estimador de Kaplan-Meier e o teste de log-rank foi utilizado na comparação entre grupos. A relação entre o tempo de sobrevida global, as principais variáveis clínicas (como idade, estadiamento, classificação histológica de Lauren, sítio anatômico e etnia) e a composição de células imunes estimadas foi realizada com a análise multivariada. Hazard ratio (HR) e o intervalo de confiança a 95% (IC (95%)) foram calculados para todas as variáveis.

Foram considerados significativos os resultados quando $p < 0{,}05$. Todas as análises estatísticas foram feitas no ambiente R v3.6 (https://www.r-project.org/).

As análises relacionadas às assinaturas mutacionais estão descritas na seção de métodos do APÊNDICE A.

5 RESULTADOS E DISCUSSÃO

Os resultados deste trabalho serão apresentados em duas partes principais considerando as fases de desenvolvimento do projeto. Primeiro, foi feita revisão da literatura com levantamento e avaliação das listas de genes marcadores de células imunes e dos algoritmos de análises de deconvolução, bem como a avaliação da performance de cada algoritmo. Com isso, foi proposto um *pipeline* de análise do contexto imunológico em câncer utilizando dados de RNA-Seq. Em seguida, com base nas conclusões desta avaliação, foi investigado o contexto imunológico em conjunto com outros fatores clínicos e moleculares que apresentam relevância prognóstica em câncer gástrico.

5.1 *BENCHMARKING* DA AVALIAÇÃO DA COMPOSIÇÃO DO INFILTRADO INFLAMATÓRIO

5.1.1 Revisão de genes marcadores de células imunes

Para avaliação da composição do infiltrado inflamatório por meio de métodos de deconvolução, é necessária a construção de um painel de referência de expressão gênica dos subtipos celulares de interesse. Este painel de referência é denominado matriz de assinatura a qual idealmente deve ser baseada em genes marcadores específicos, ou seja, espera-se que um gene, ou um grupo de genes, apresente(m) atividade transcricional maior em um tipo celular quando comparado aos demais.

A definição de uma lista de genes marcadores de células imunes é um grande desafio, uma vez que os níveis de expressão gênica são intrinsecamente muito variáveis e, adicionalmente, sensíveis às condições funcionais de cada tipo celular. Adicionalmente, subtipos de células imunológicas compartilham entre si marcadores funcionais e fenotípicos. De importância, no microambiente tumoral,

além de células imunes infiltrantes, há uma diversidade de outros tipos celulares, incluindo as células estromais, tumorais e células normais do tecido adjacente.

Desta forma, a complexidade de composição e fenótipos das células em amostras tumorais é muito maior, o que se reflete na expressão gênica, que é um somatório de todas as células presentes nesse ambiente. Dessa forma, é necessário não apenas considerar genes que diferenciam os tipos celulares imunes entre si, mas também que estes genes não deverão estar expressos em células do estroma e/ou tumorais, células que também estão presentes no microambiente tumoral. Assim, a determinação exata de marcadores específicos representa um dos maiores desafios dessa abordagem.

Inicialmente, foi feito um levantamento de listas de genes marcadores dos vários subtipos de células imunológicas a partir de uma extensa revisão bibliográfica, pesquisa em bancos de dados e de listas associadas aos métodos pré existentes de deconvolução. Este compilado encontra-se detalhado na Tabela 3.

Os primeiros estudos de estimação da composição de células imunes tiveram como base dados de experimentos realizados por *microarray* (Charoentong et al. 2017; Bindea et al. 2013; Abbas et al. 2009; Abbas et al. 2005) (Tabela 3). Cada estudo utilizou diferentes subtipos de células imunes, que variam de acordo com grau de diferenciação e aspectos funcionais. Assim, cada lista de genes descritas na Tabela 3 considera genes diferencialmente expressos em um conjunto de células distinto. A fim de avaliar a consistência dos genes marcadores entre as listas, primeiramente comparamos os subtipos celulares de cada lista. Com isso, foi possível agrupar as listas em 3 *clusters*, considerando os subtipos de células semelhantes entre elas (Figura 15). Algumas listas consideraram subtipos de células imunes mais especializadas, outras, tipos mais gerais sem especificações em subtipos.

Diante desta divergência dos tipos e subtipos de células em cada uma das listas, comparamos os genes considerando os grupos com número e tipo de células mais semelhantes (Figura 16). Esta análise de sobreposição de genes mostrou que,

Tabela 3 – Revisão de listas de genes marcadores de células imunes. Células: quantidade de subtipos celulares; Genes: quantidade de genes considerados em cada lista.

Lista	**Referência**	**Dados de origem**	**Células**	**Genes**
Abbas2005 (IRIS)	Abbas et al. (2005)	*microarray*	16	1627[a]
Abbas2009	Abbas et al. (2009)	*microarray*	18	267[a]
HaemAtlas	Watkins et al. (2009)	*microarray*	6	1214[a]
Allantaz2012	Allantaz et al. (2012)	*microarray*	8	701[a]
ImmPort		*microarray*		
D-MAP	Novershtern et al. (2011)	*microarray*	3	2847[a]
Bindea2013	Bindea et al. (2013)	*microarray*	28	499[a]
Angelova2015	Angelova et al. (2015)	*microarray*	31	792[a]
NanoString	Cesano (2015)	*microarray*	24	109
Newman2015 (LM22[d])	Newman et al. (2015)	*microarray*	22	547
Becht2016	Becht et al. (2016)	*microarray*	8	70
Aran2017 (xCell)	Aran et al. (2017)	*microarray* e RNA-seq	34[b]	2702[a]
Charoentong2017	Charoentong et al. (2017)	*microarray*	28	779[a]
Danaher2017	Danaher et al. (2017)	*microarray*	13	60
Racle2017	Racle et al. (2017)	RNA-seq	8	78[a]
BRef[cd]	Racle et al. (2017)	RNA-seq	6	65
TRef[cd]	Racle et al. (2017)	RNA-seq	7	98
BioLegend	BioLegend(R)	Citometria de fluxo	18	214
CellMarker	Zhang et al. (2019b)	*microarray* e RNA-seq	69	6227[a]
LM6[cd]	Newman et al. (2019)	*microarray*	6	684
Sugio2018	Sugio et al. (2018)	*microarray*	14	119
Finotello2019 (TIL10)[d]	Finotello et al. (2019)		10	170

[a] Número de genes únicos; um mesmo gene pode estar associado a mais de um subtipo celular. Nos casos de listas mais antigas baseadas em microarray, as assinaturas estavam associadas às sondas. Uma vez que diferentes sondas podem se ligar a um mesmo gene, as sondas foram traduzidas para os nomes oficiais dos genes (NCBI) e então foi verificado o número de genes únicos válidos.
[b] *xCell* avalia 34 células imunes e 30 do estroma ou células precursoras, totalizando 64 células.
[c] Listas de genes não incluídas na comparação inicial.
[d] Listas de genes obtidas de assinaturas de referência construídas para o método de deconvolução do próprio estudo.

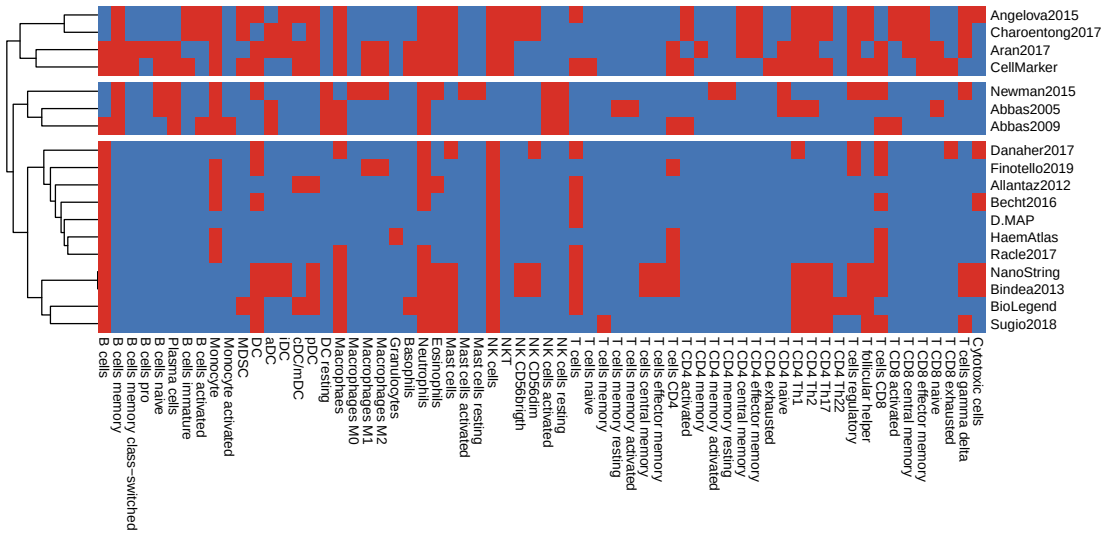
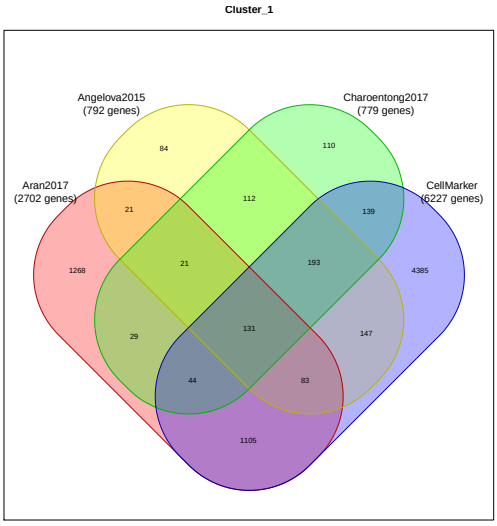


Figura 15 – Similaridade entre diferentes listas de genes marcadores de células imunes. As listas de genes foram agrupadas considerando os tipos de células incluídas em cada um. Vermelho indicam que as listas nas linhas continham tal célula descrita nas colunas. Azul indicam que tal lista não considerava o tipo celular em questão. Três grupos principais foram encontrados de forma não supervisionada.

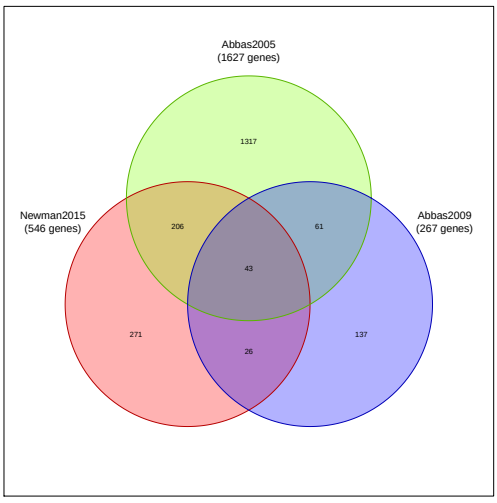
mesmo sendo feita a comparação entre as listas mais semelhantes de acordo com os tipos celulares considerados em cada uma, o número de genes compartilhados entre 3 ou mais listas é relativamente pequeno (Figura 16) e a maioria dos genes são exclusivos para determinada lista. Por exemplo, 5120 genes de um total de 8143 (62,9%) aparecem em uma única lista de genes marcadores. Nesta etapa, foi desconsiderada a lista "D-MAP" no grupamento 3 pois embora tenha sido agrupado, compreende apenas três tipos celulares imunes e da mesma linhagem linfóide (linfócitos T e B e células Natural Killer - NK).

Outro ponto de divergência foi o número variável de genes que compõe cada lista, sendo umas compostas por mais de 1000 genes, como Abbas2005 e Aran2017, e outras listas com menos de 100 genes como Racle2017 e Danaher2017 (Tabela 3).

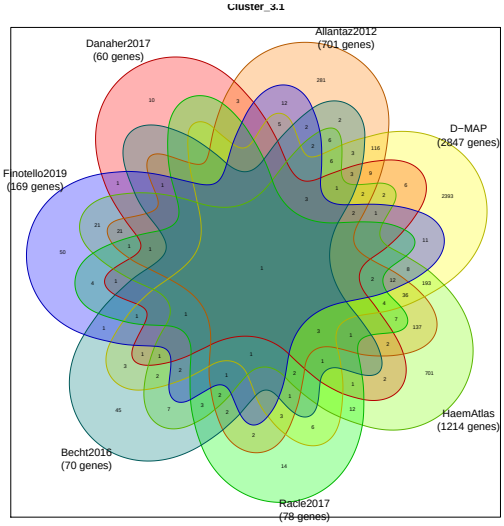
A fim de avaliar a especificidade dos genes em cada uma das listas de genes marcadores, foi construído um *heatmap* considerando oito células imunes básicas utilizando a referência HPCA (Figura 18). A maioria das listas agruparam as



(a) Grupo 1



(b) Grupo 2



(c) Grupo 3

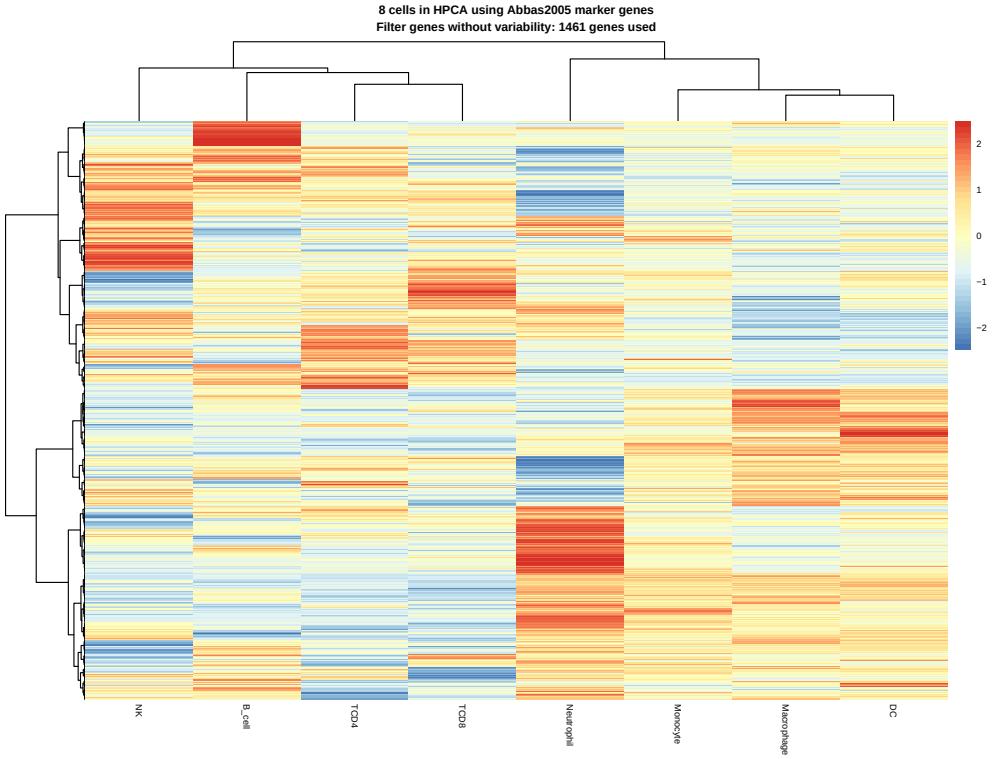
Figura 16 – Diagrama de Venn ilustrando a sobreposição dos genes. As listas de genes encontram-se previamente definidas na Figura 15, incluindo (a) o grupo 1 com mais subtipos celulares, (b) grupo 2 com quantidade intermediária de subtipos celulares e (c) grupo 3 com subtipos celulares mais básicos.

células das duas linhagens (linfóide e mielóide) como esperado: em dois grandes grupos clusterizados de forma não supervisionada. A exceção foi a lista de "NanoString", cujo perfil de expressão de genes associados aos linfócitos B ficou mais semelhante ao grupo de células mielóides (Figura 18f). Como comentado anteriormente, a lista ideal de genes marcadores contém genes específicos associados a cada tipo celular. Dessa forma, no *heatmap* é esperado um padrão em blocos ou escada, ou seja, genes fortemente expressos em um tipo celular, não estariam nos demais.

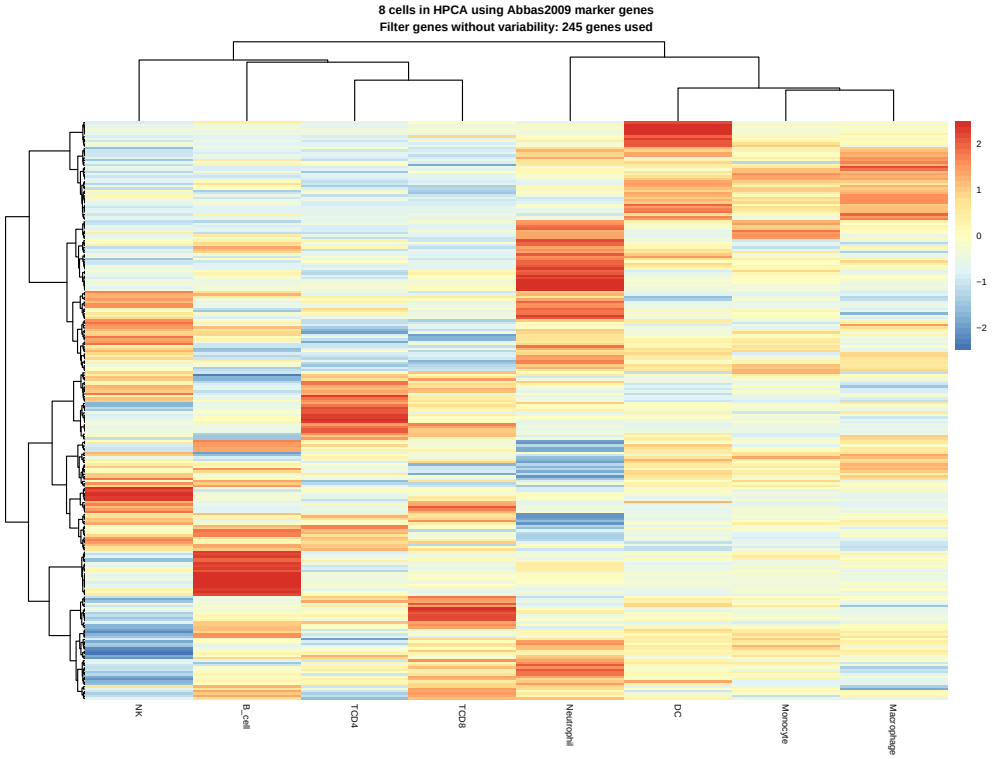
Pelo padrão de expressão de cada grupo de genes marcadores, considerando os oito subtipos básicos de células imunes, observamos que o conjunto de genes parece ser mais importante do que o número de genes em cada lista, visto que o padrão em blocos/escada esperado pode ser observado em listas com quantidades maior ou menor de genes, como para Charoentong2017 e Racle2017, respectivamente (Figuras 18k e 18m).

A partir desta revisão de genes marcadores imunes, pode-se concluir que, embora alguns genes sejam comuns entre as listas, principalmente os marcadores clássicos utilizados também na técnica de citometria de fluxo (padrão outro para imunofenotipagem), a maior parte das listas é composta por genes que não se repetem em outras. O fato de diferentes subtipos de células imunes serem consideradas em cada lista (número e nível de diferenciação) pode aumentar esta não sopreposição entre os genes.

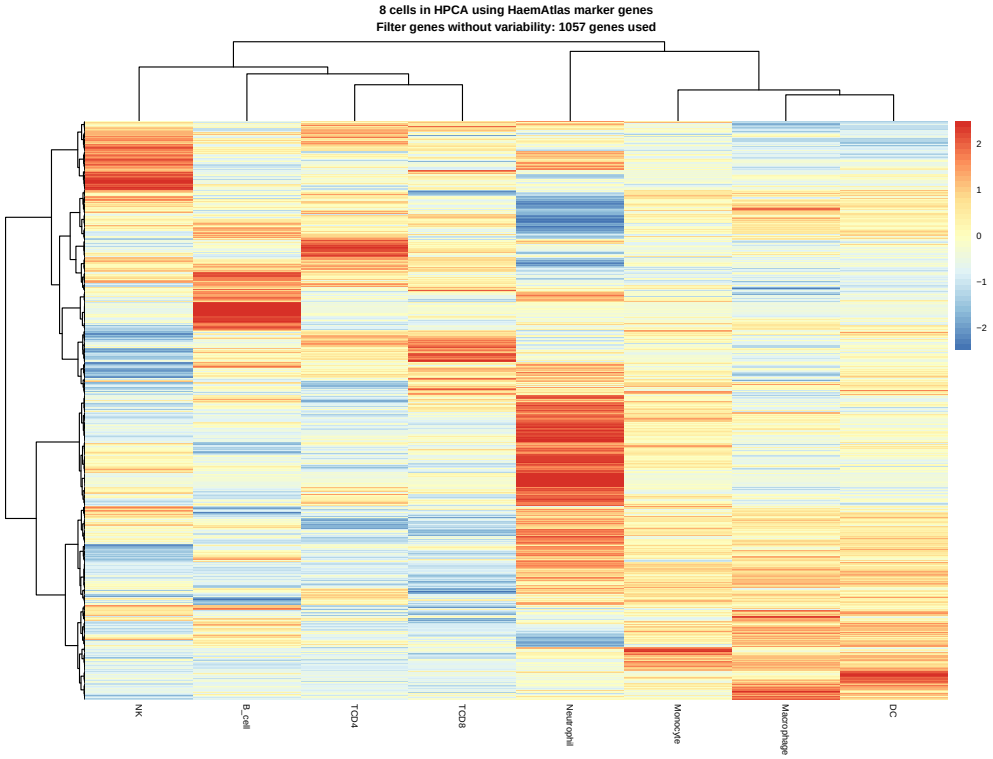
O passo seguinte foi avaliar o desempenho de cada lista de genes, em conjunto com os algoritmos de deconvolução disponíveis, para a estimação da composição de células imunes no microambiente tumoral. Nesta etapa, duas listas adicionais, baseada nesta revisão, foram construídas a partir de genes presentes em 5 ou 4 listas. Estas listas foram denominadas "review-4"e "review-5", compostas por 695 e 399 genes, respectivamente.



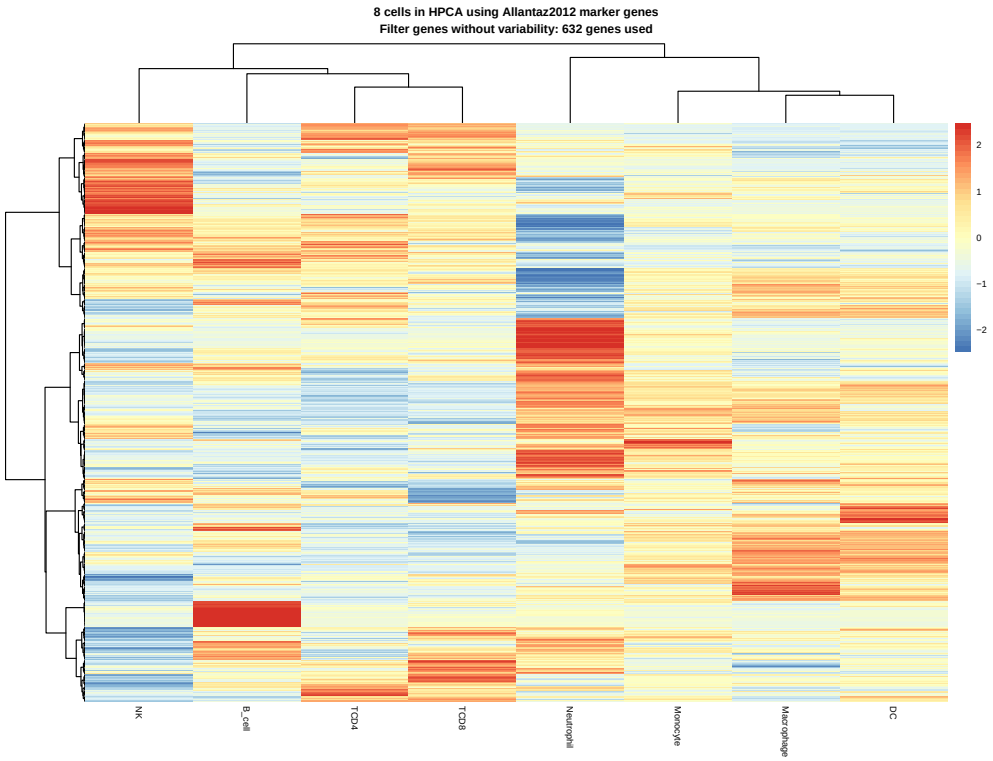
(a) Abbas2005



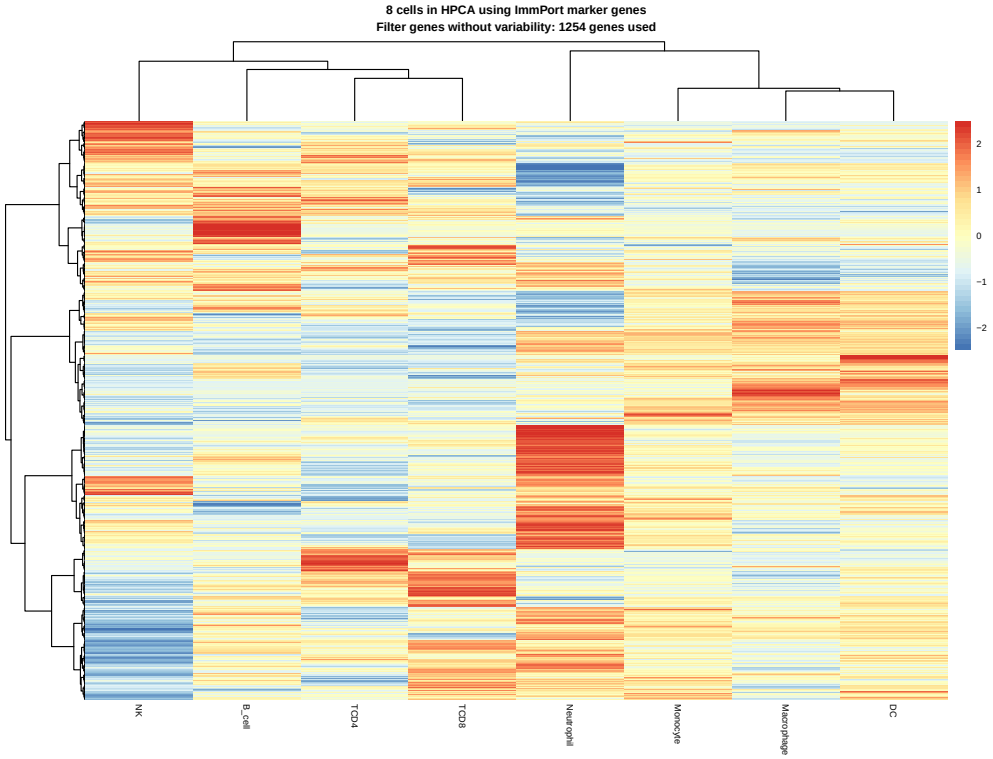
(b) Abbas2009



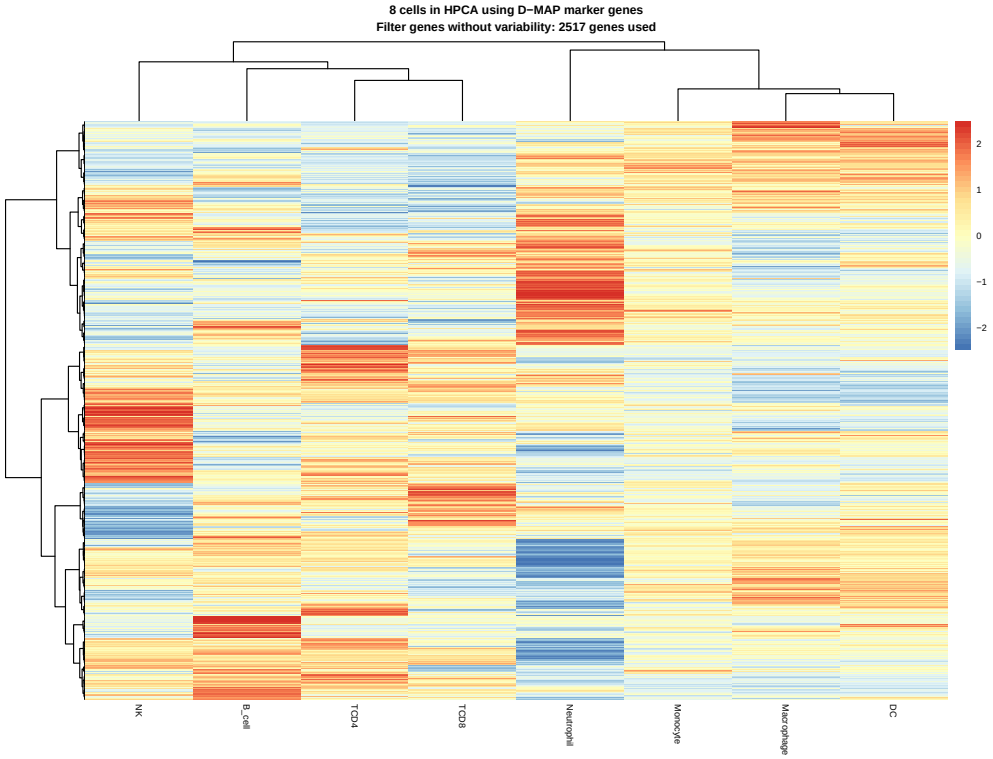
(c) HaemAtlas



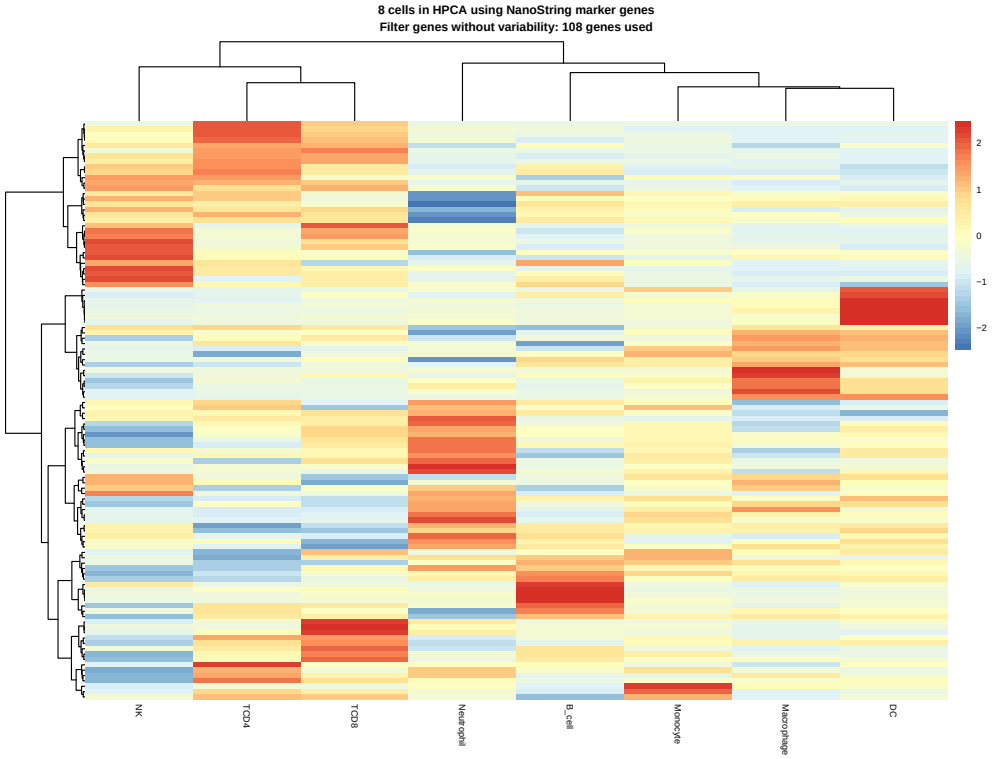
(d) Allantaz2021



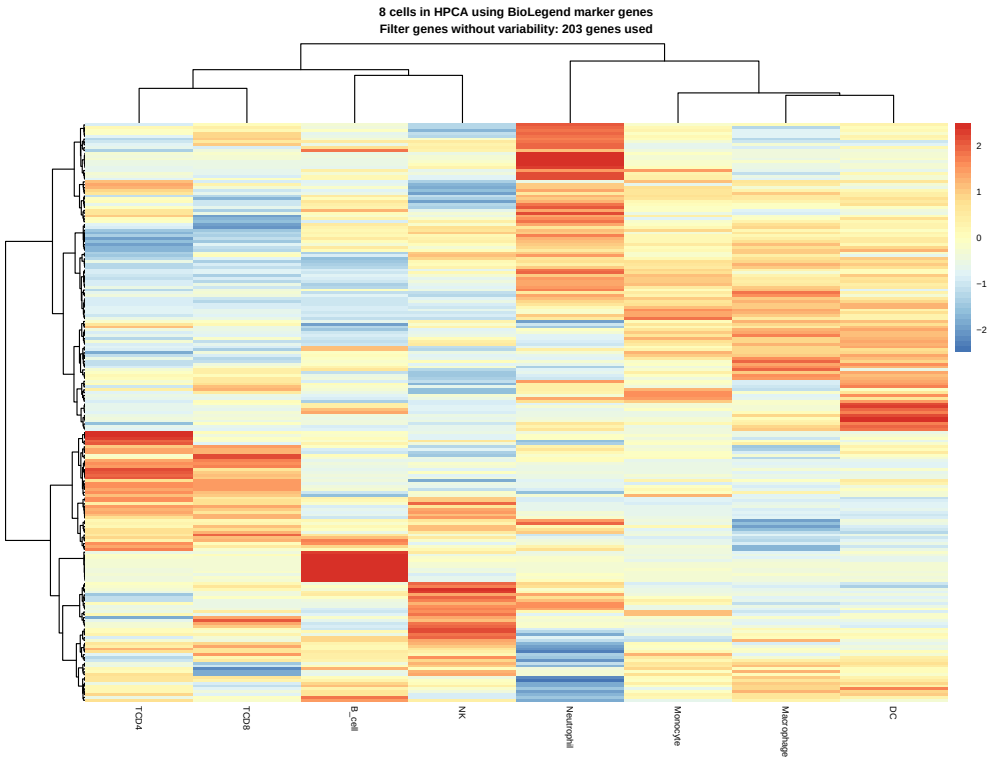
(e) ImmPort



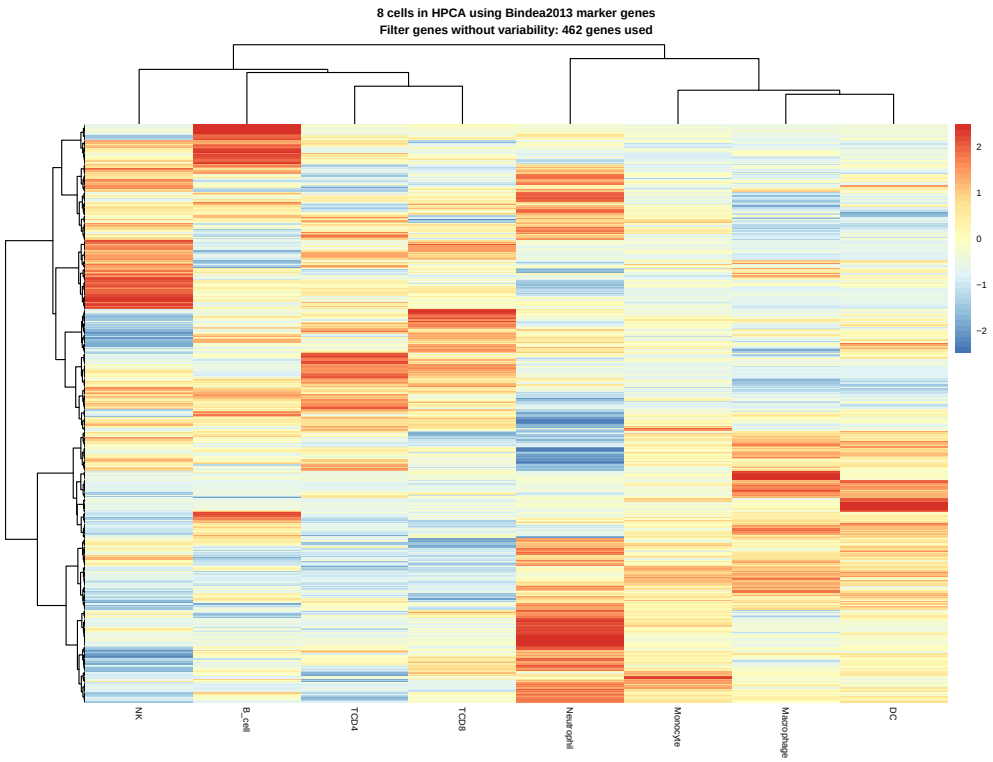
(f) D-MAP



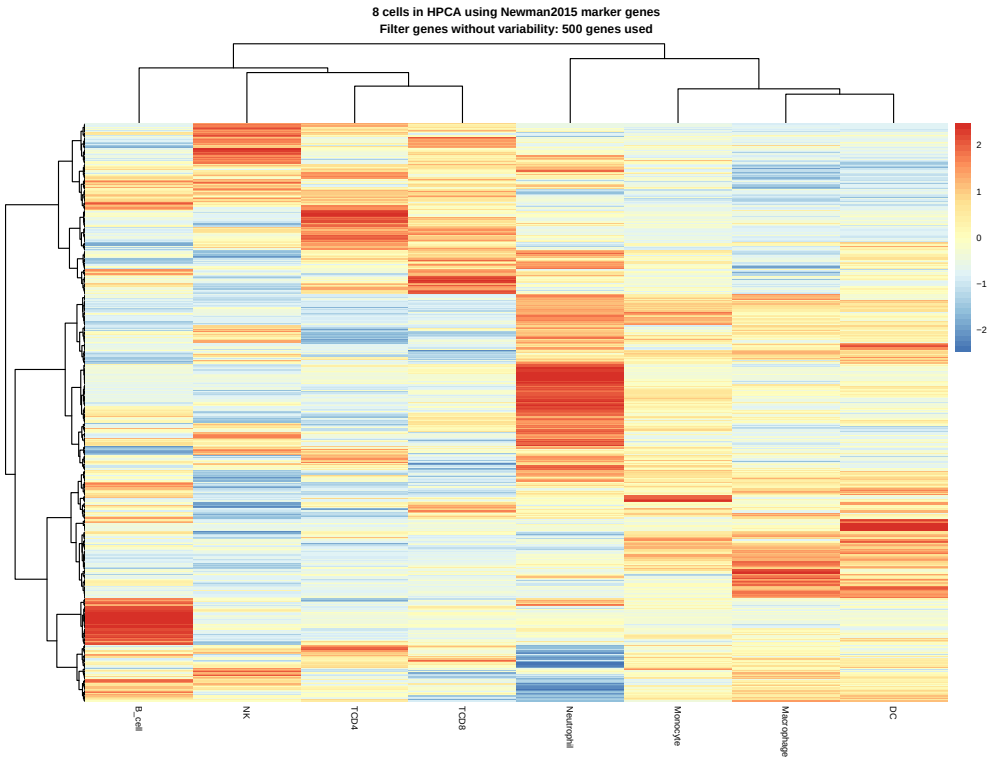
(g) NanoString



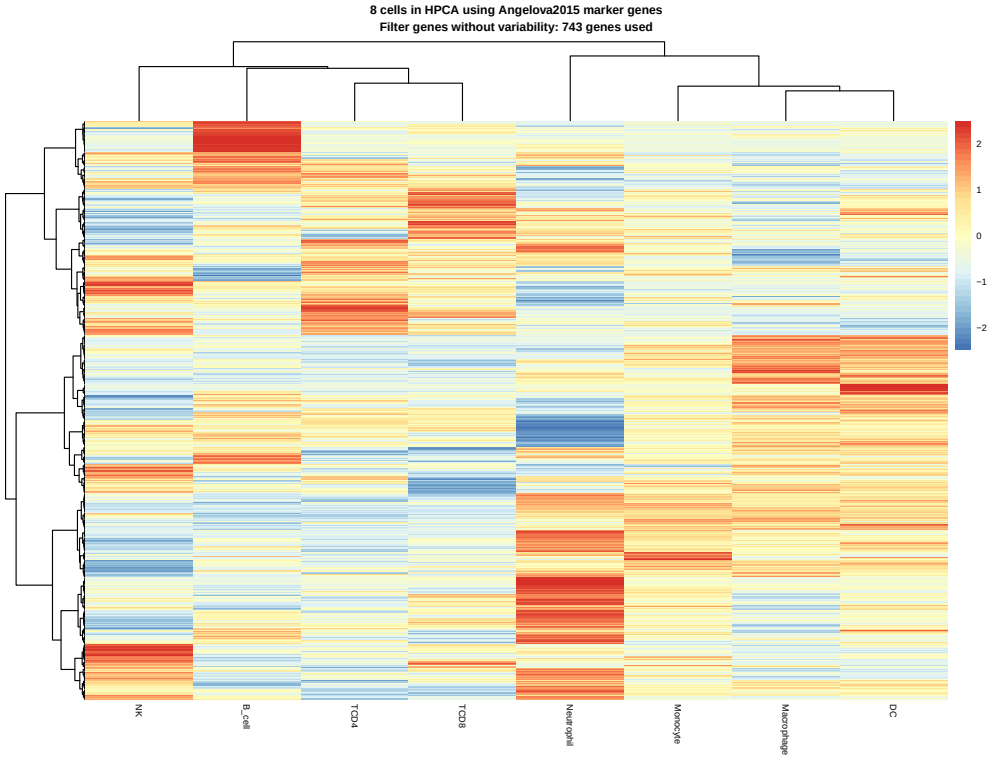
(h) BioLegend



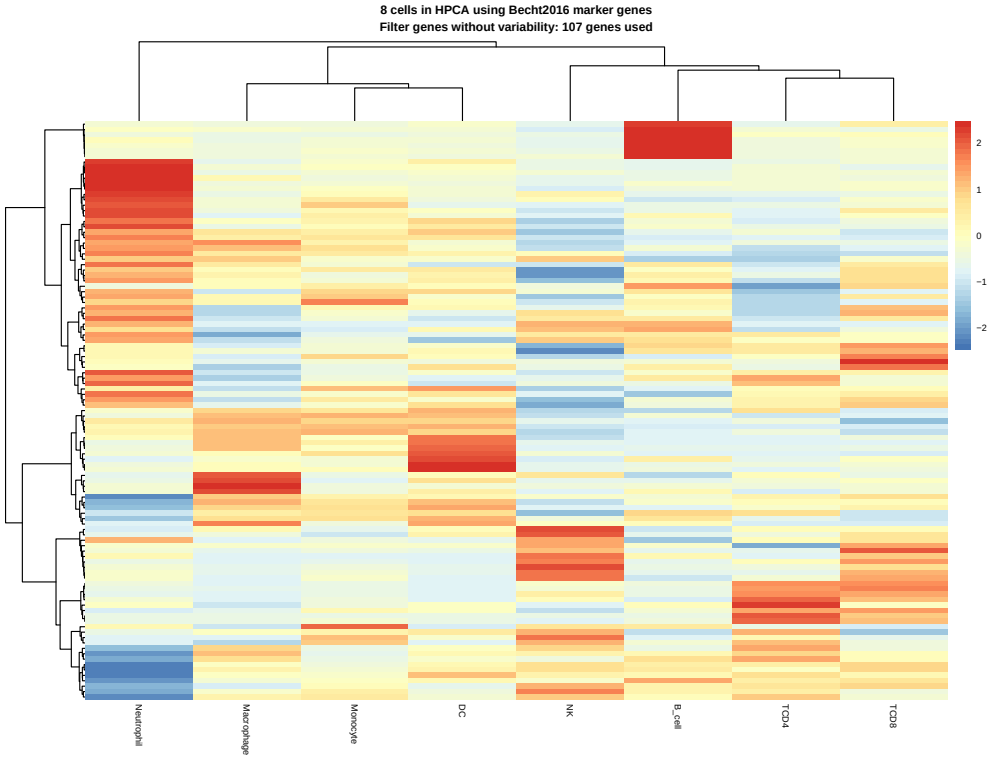
(i) Bindea2013



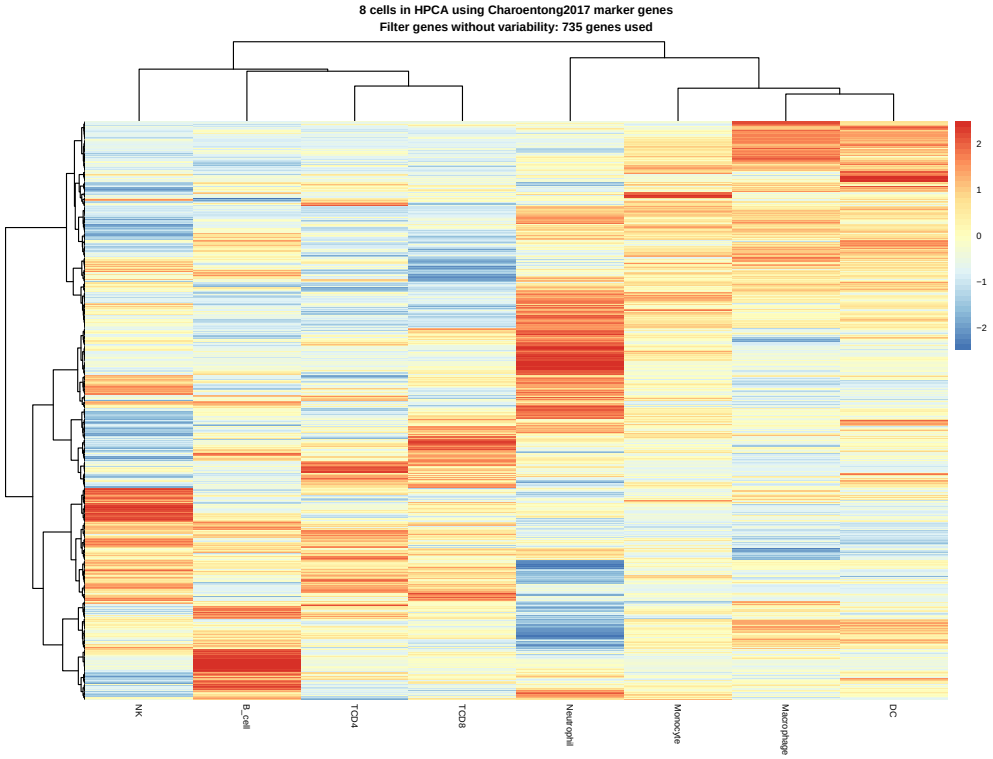
(j) Newman2015



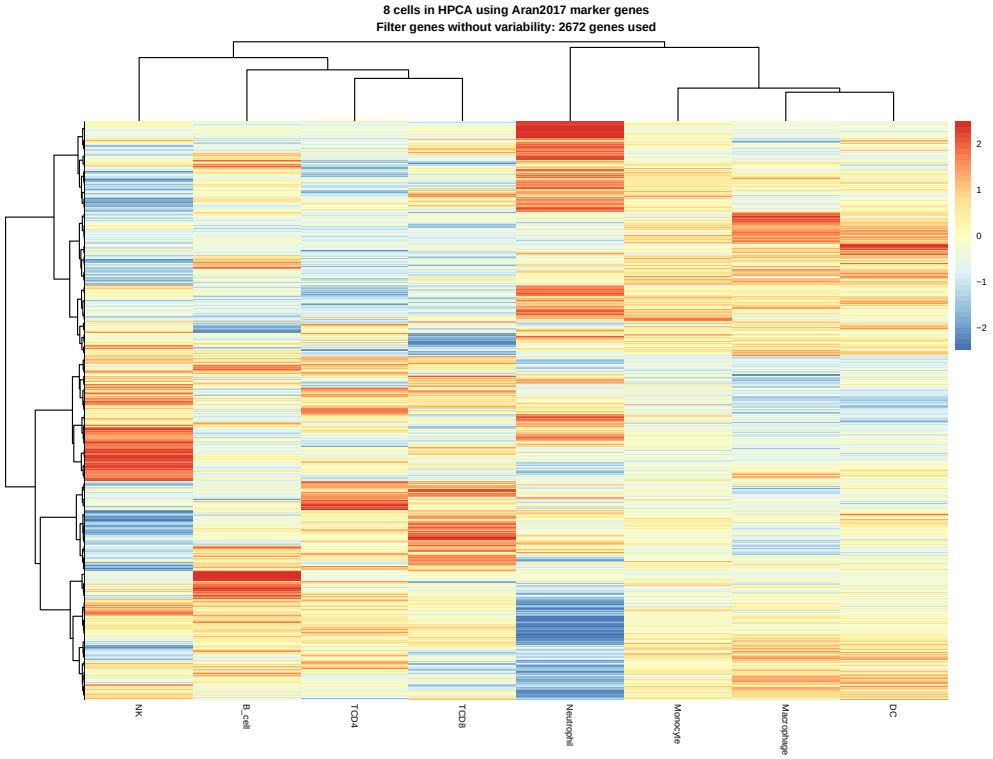
(k) Angelova2015



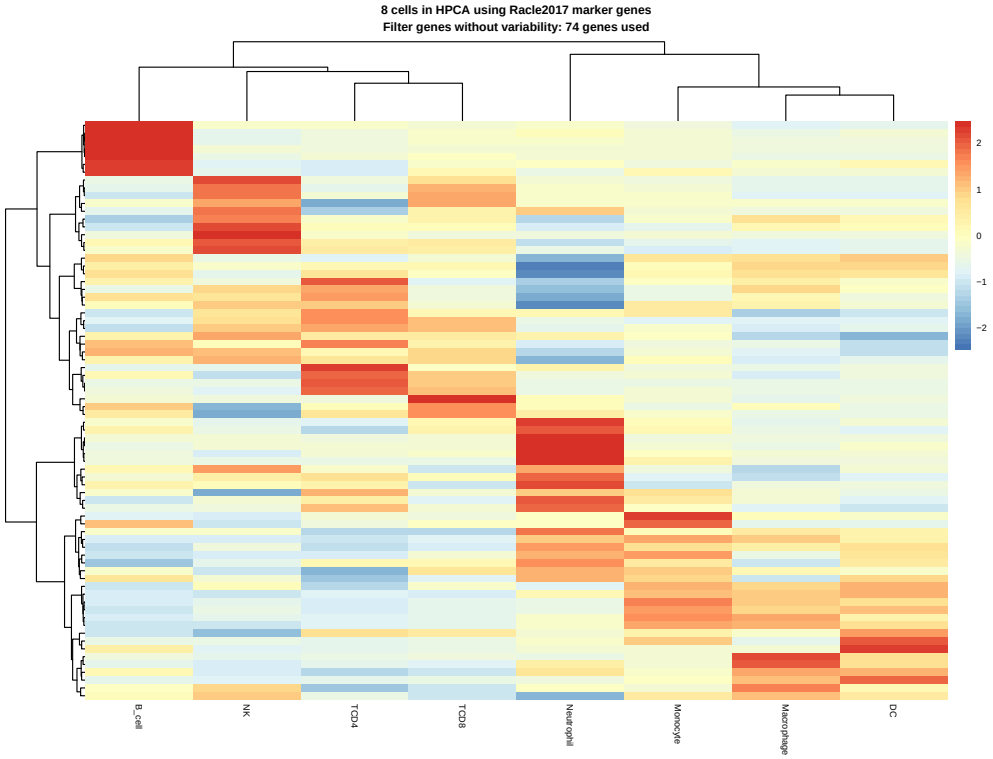
(l) Becht2016



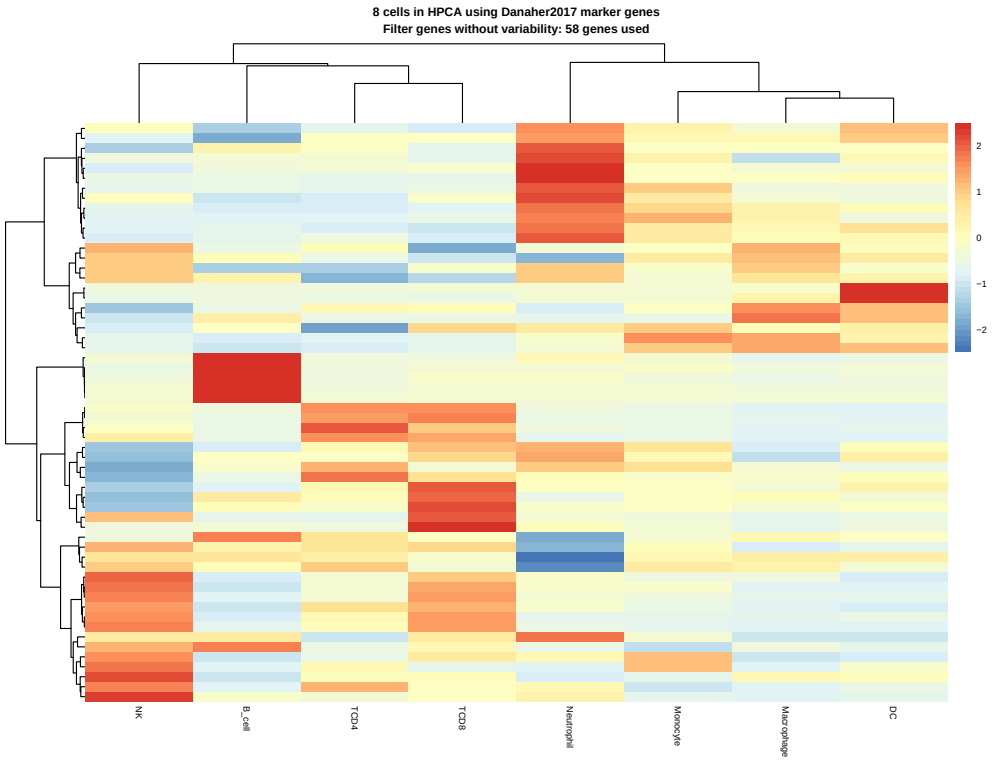
(m) Charoentong2017



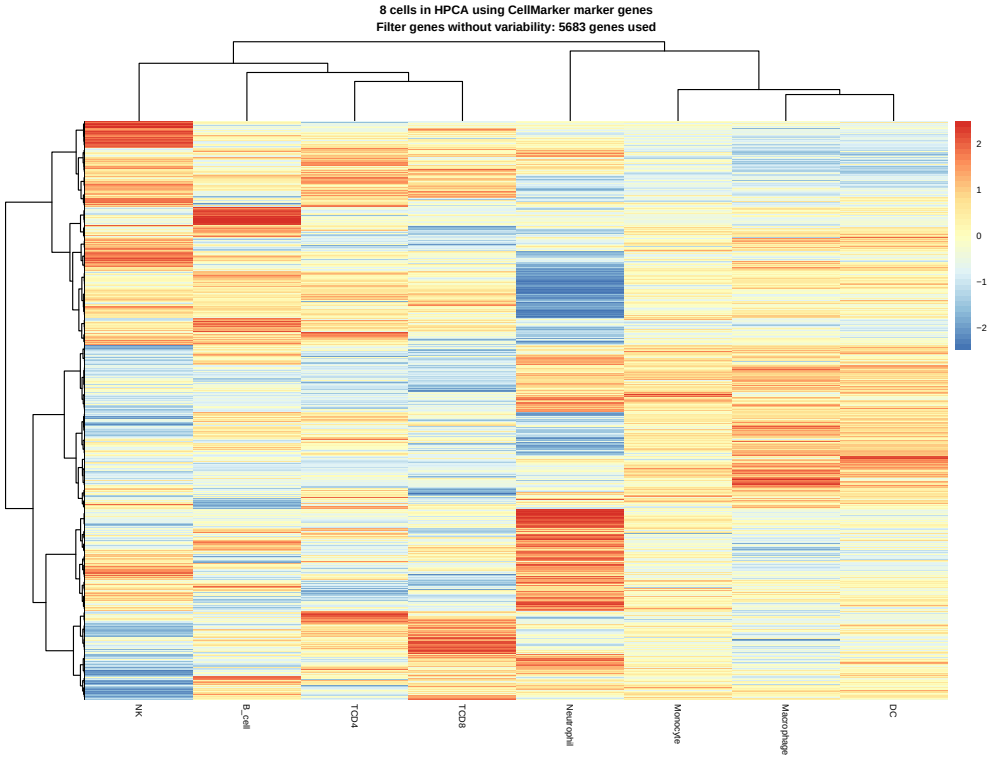
(n) Aran2017



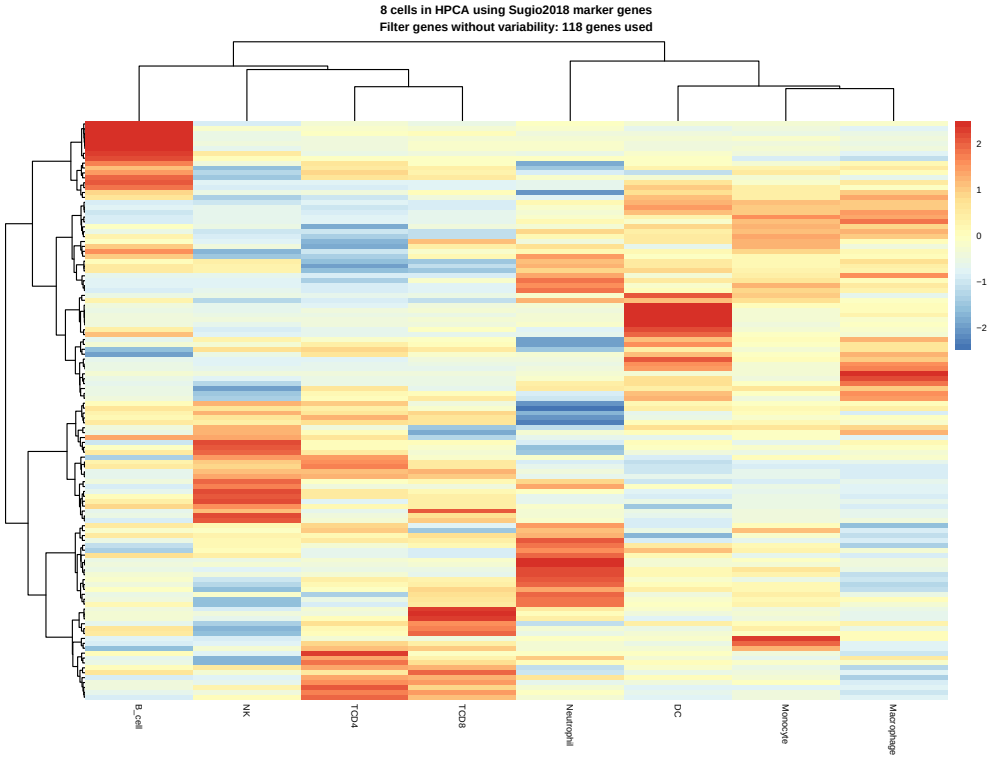
(o) Racle2017



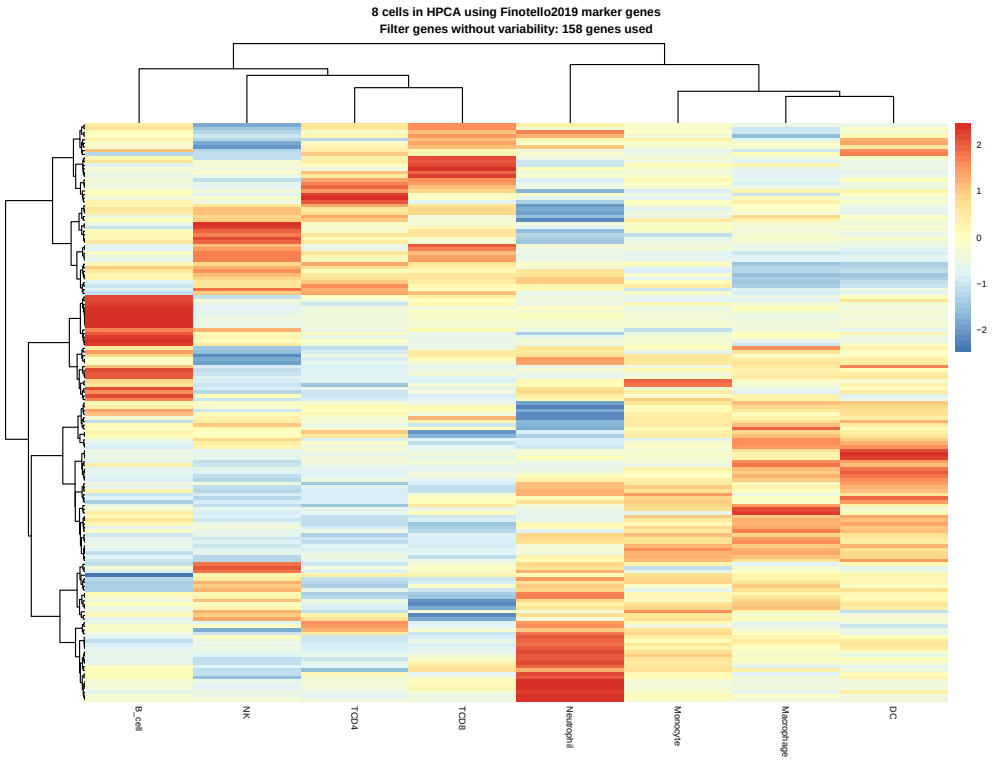
(p) Danaher2017



(q) CellMarker



(r) Sugio2018



(s) Finotello2019

Figura 18 – *Heatmaps* de expressão de cada lista de genes considerando 8 tipos celulares imunes básicos e os dados de expressão de referência do HPCA (Mabbott et al. 2013). Os valores de expressão dos genes de cada lista foram calculados pela mediana dos valores das células de um mesmo subtipo e normalizados com *z-score* e plotados em *log*.

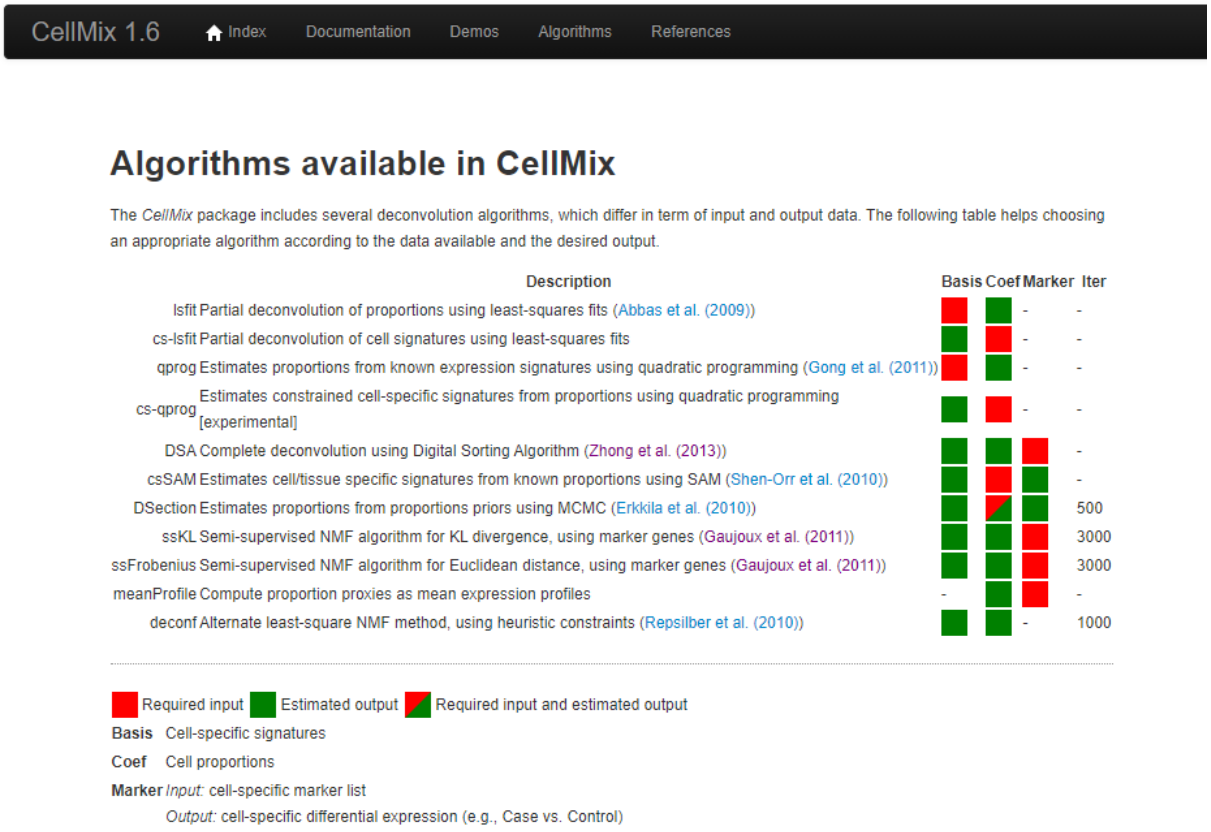
5.1.2 Revisão dos métodos de deconvolução

Juntamente com o levantamento de listas de genes marcadores de células imunes, foi feita uma revisão do métodos de deconvolução disponíveis (Tabela 4). Não raro, alguns métodos foram desenvolvidos em conjunto com listas de genes marcadores, matriz com assinatura de referência ou mesmo algum método para construção de assinatura de referência.

Os primeiros métodos de deconvolução foram agrupados no pacote CellMix (Gaujoux e Seoighe 2013) disponível em linguagem R. O pacote foi uma proposta para reunir os métodos de deconvolução, além de estratégias de busca de marcadores por meio da expressão gênica (Figura 19).

Um dos pontos críticos da abordagem por deconvolução é definir o conjunto de marcadores específicos das células de interesse que estão presentes em

uma mistura (ou amostra). A partir destes marcadores, há dois tipos de análise de deconvolução: a completa e a parcial. A deconvolução completa utiliza apenas uma lista de genes marcadores para estimar a proporção de células. Já a deconvolução parcial utiliza uma assinatura, ou seja, valores de referência de um conjunto de genes marcadores para cada célula de interesse. Os métodos de deconvolução levantados pela revisão de literatura encontram-se disponíveis na Tabela 4.



Fonte: http://web.cbio.uct.ac.za/~renaud/CRAN/web/CellMix/_PAGE-Algorithms.html

Figura 19 – Ferramenta CellMix (Gaujoux e Seoighe 2013).

Os principais métodos de deconvolução presentes no pacote CellMix são *lsfit* (Abbas et al. 2009) e *qprog* (Gong et al. 2011) e consideram um painel de referência (deconvolução parcial). Por outro lado, os métodos *ssKL* (Gaujoux e Seoighe 2012) e *DSA* (Zhong et al. 2013) utilizam apenas lista de marcadores (deconvolução completa) (Figura 19). Para entender melhor como os dois tipos de

Tabela 4 – Revisão de métodos de avaliação da composição de infiltrado inflamatório utilizando dados de expressão gênica.

Métodos de deconvolução:				
Algoritmo	**Referência**	**Tipo**	***Input***	**Estimativa**
lsfit[a]	Abbas et al. (2009)	parcial	matriz	proporção
qprog[a]	Gong et al. (2011)	parcial	matriz	proporção
ssKL[a]	Gaujoux e Seoighe (2012)	completa	lista	proporção
DSA[a]	Zhong et al. (2013)	completa	lusta	proporção
CIBERSORT(x)	Newman et al. (2015), Newman et al. (2019)	parcial	matriz	proporção e abundância
MCPcounter	Becht et al. (2016)		lista	abundância
EPIC	Racle et al. (2017)	parcial	matriz	proporção
TIMER	Li et al. (2017)	parcial	matriz	proporção
DTANGLE	Hunt et al. (2019)			proporção
"DSA"	Zaitsev et al. (2019)		matriz	proporção
ABIS	Monaco et al. (2019)	completa		abundância
MIXTURE	Fernández et al. (2019)	parcial	matriz	proporção
Outras ferramentas:				
Algoritmo	**Referência**	**Tipo**	***Input***	**Estimativa**
xCell	Aran et al. (2017)	GSEA	lista	*score* de abundância
immunedeconv	Sturm et al. (2020)	*pipeline*		
qaunTIseq	Finotello et al. (2019)	*pipeline*	matrix (TIL10)	abundância/mm^2
ConsesusTME[c]	Jiménez-Sánchez et al. (2019)	*pipeline*		

[a] Métodos do pacote CellMix (Gaujoux e Seoighe 2013) que compilou diferentes ferramentas em um mesmo ambiente. No entanto, o projeto não foi mais atualizado e foi então descontinuado, pois tinha como base a linguagem R compatível apenas com versões anteriores a 3.

[b] Inclui, no mesmo pacote, os métodos quanTIseq, TIMER, CIBERSORT, MCPcounter, *xCell* e EPIC.

[c] Ferramenta integrativa de vários métodos atuais a fim de geral um resultado consenso e auxiliar na escolha do melhor método para avaliação do microambiente tumoral.

abordagens (completa e parcial) com os dados de RNA-seq, cada um dos quatro algoritmos principais do pacote CellMix foi utilizado para avaliar 63 amostras de dados a partir de células puras com 13 subtipos de linfócitos T e B (Bonnal et al. 2015).

A assinatura de referência foi construída utilizando a mediana do valor de expressão normalizado (FPKM) de 9 dos 13 subtipos celulares, a partir das 63 amostras. Esta assinatura foi então filtrada pela lista de genes marcadores (Charoentong et al. 2017), que também foi utilizada como parâmetro para os métodos *ssKL* e *DSA*. Em seguida, foi escolhida esta lista de marcadores pois os subtipos celulares incluídos continham a maioria dos 13 subtipos celulares a serem testados, os 9 selecionados para construção da matriz de assinatura (Figura 20).

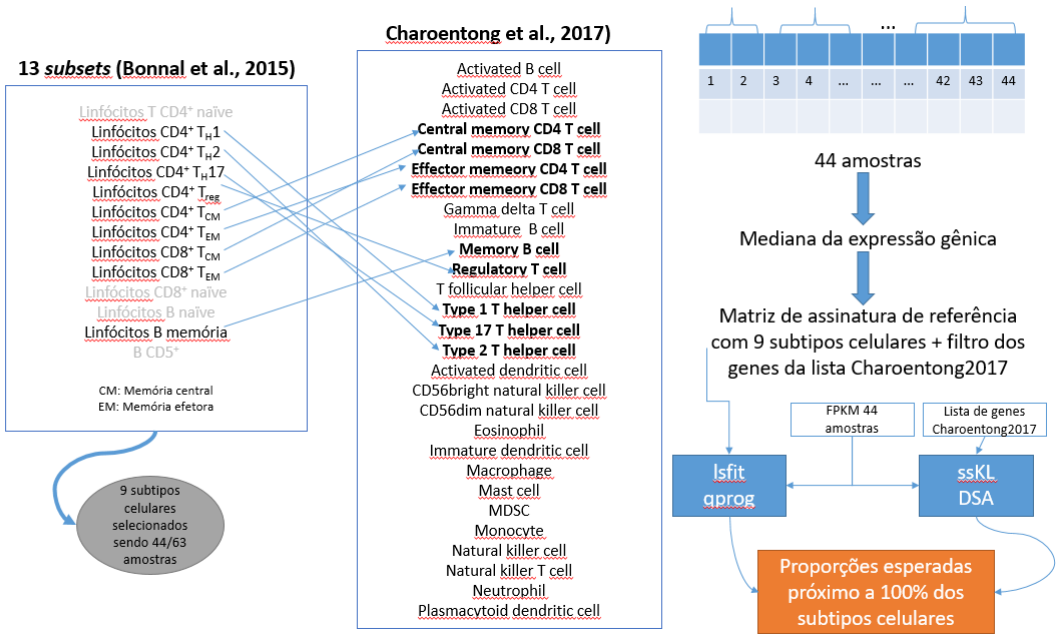


Figura 20 – Esquema de análises utilizando os *13 subsets* para comparar os métodos de deconvolução parcial e completa.

Os métodos *lsfit* e *qprog* estimaram 90% e 86%, respectivamente, do resultado esperado considerando 44 amostras dos subtipos celulares compatíveis com a assinatura. Já os métodos *ssKL* e *DSA* associaram todas as amostras com subtipos de células NK. Desta forma, como os resultados com os métodos de

deconvolução completa foram inadequados, apenas os métodos de deconvolução parcial foram então considerados no *benchmarking*.

É importante ressaltar que os primeiros métodos de estimação da composição de células imunes tiveram como base dados de *microarray*. Atualmente, os dados de expressão gênica são gerados, em sua maioria, a partir da técnica de sequenciamento de nova geração, RNA-Seq. Consequentemente, uma vez que os níveis de expressão gênica foram obtidos a patir da tecnologia de *microarray*, é necessário realizar uma normalização dos níveis de expressão utilizados nas assinaturas de referência.

Além dos tipos de deconvolução existentes, as motivações destes primeiros métodos e sua aplicação na imuno-oncologia foi outro ponto importante de ser considerado. A princípio, os métodos de deconvolução foram desenvolvidos com objetivo de determinar a composição de células circulantes no sangue, ou seja, a partir de amostras de plasma sanguíneo (Abbas et al. 2009). Nesse sentido, espera-se que a maior parte das células sejam os leucócitos, que são as próprias células imunológicas. Assim, faz sentido que o somatório dessas células seja próximo a 100% da sua amostra, ou seja, que a expressão gênica seja o reflexo apenas dessas células de interesse e não de outras. Essa ideia muda ao considerarmos as amostras de tumores, uma vez que a expressão gênica resulta de uma mistura de células tumorais, células de tecido normal adjacente, células imunes do infiltrado inflamatório e do estroma.

Os primeiros métodos que adaptaram os algoritmos de deconvolução considerando dados de amostras de tecido tumoral foram *CIBERSORT* (Newman et al. 2015), *TIMER* (Li et al. 2017) e *EPIC* (Racle et al. 2017) (Tabela 4).

TIMER foi uma estratégia desenvolvida baseada nos dados do *The Cancer Genome Atlas* (TCGA). O método de deconvolução é semelhante ao *lsfit*, porém o *TIMER* compreende um *pipeline* para normalização dos dados, montagem de uma assinatura específica para cada tipo tumoral e aplicação da deconvolução que resulta em um valor de abundância de células. Para construção da matriz de

assinatura, o *TIMER* utiliza o *dataset* de referência HPCA que contém o valor de expressão de mais de 300 células puras (Mabbott et al. 2013). Basicamente, a mediana é utilizada como principal método para obter a expressão de referência das células de interesse (Figura 21), a partir de várias amostras do mesmo tipo celular. No entando, o que diferencia este método é que, para montagem da matriz de assinatura final, há inclusão dos dados de expressão das misturas (amostra tumoral) e a informação da pureza do tumor de cada amostra no refinamento desta assinatura de referência.

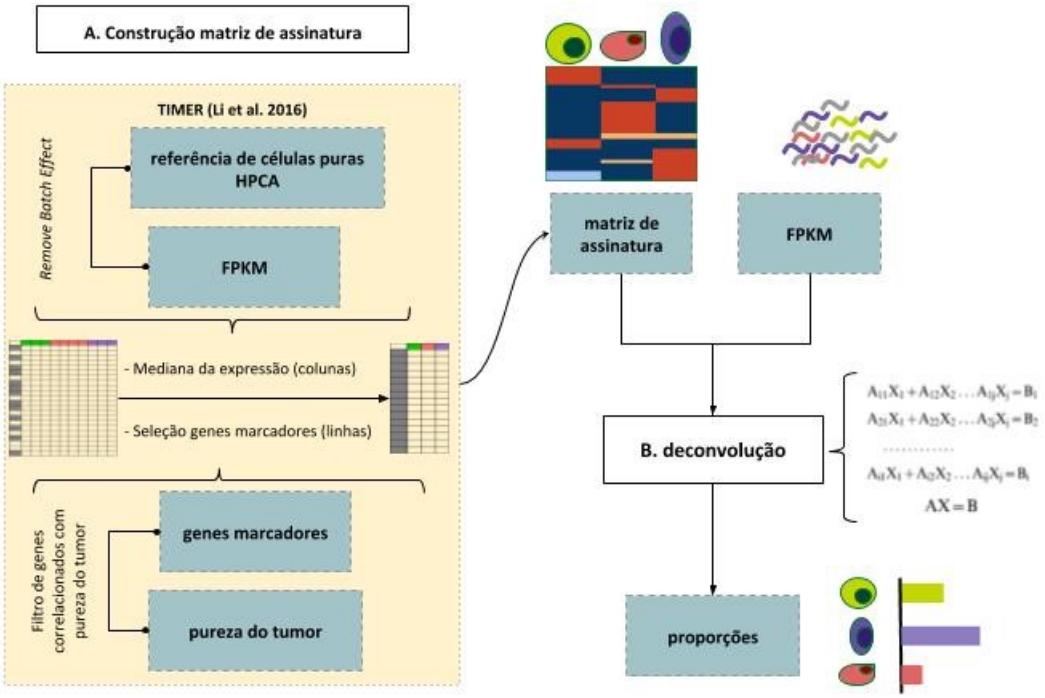


Figura 21 – Esquema da análise de infiltrado inflamatório utilizando método TIMER (Li et al. 2017) para construção das assinaturas de refêrencia. A) A construção da matriz de assinatura de referência, importante para deconvolução dos subtipos celulares de interesse, baseia-se nos dados de expressão gênica de células puras, lista de genes de interesse e normalizações considerando pureza do tumor e variabilidade dos dados de expressão gênica (*batch effect*). B) O passo de deconvolução utiliza a matriz de assinatura de referência com as expressão esperada para cada tipo celular e os dados de expressão gênica da mistura (*bulk*).

Este refinamento considera que alguns genes marcadores de células imunes podem estar expressos nas células tumorais e, portanto, não seriam interessantes marcadores para deconvolução. Assim, é feita correlação da expressão dos genes

com a pureza do tumor em cada amostra, sendo filtrados aqueles com correlação positiva (ou seja, presentes no tumor). Além disso, o *dataset* de referência (originado de dados de *microarray*) é normalizado em conjunto com o valor de expressão (FPKM) de cada gene na amostras tumorais, utilizando estratégia de "*Remove Batch Effect*". Com os dados de expressão normalizados e a referência refinada, é aplicado o método de deconvolução que resulta na abundância de células de interesse. Esta estratégia foi adaptada no presente trabalho para a construção de outras assinaturas e, adicionalmente, com a opção de escolha de lista de marcadores e número de células de interesse.

O software *EPIC*, por sua vez, tem duas funcionalidades de particular interessante. Primeiro, seu desenvolvimento foi baseado em dados de RNA-Seq. Além disso, é o único que inclui na estimativa da composição de células proporções de "outras"células. Desta forma, esse software estima, além das células contidas na matriz de assinatura, a proporção de células "desconhecidas" que não estariam presentes na matriz inicial usadas para estimar as 8 células (Racle et al. 2017).

O *CIBERSORT*, por sua vez, foi desenvolvido para estimar células circulantes baseado em dados de *microarray* (Newman et al. 2015). Seguindo as novas abordagens para quantificar os níveis de expressão gênica em larga escala, foi incluído um parâmetro de normalização para aplicação em dados de RNA-Seq e passou a ser aplicado em estudos de imuno oncologia (Craven et al. 2021; Zhong et al. 2019; Thorsson et al. 2018). Seu diferencial é a matriz de assinatura de referência (LM22) composta por 22 subtipos de células imunes e um valor de p associado a estimativa de proporções destas células. Outro aspecto interessante é a possibilidade de estimar proporções (estimativas de células que somam 100%) e abundância, que são valores normalizados mais próximos à quantidade real de cada célula. Em 2018, o método foi atualizado com novas funcionalidades e, no caso da deconvolução para estimativa da composição de células, foi acrescentado um parâmetro semelhante de remoção do *batch effect*, já que a assinatura LM22 provém de dados de *microarray* e estão sendo aplicados em dados de RNA-seq

(Newman et al. 2019). O *CIBERSORTx* está disponível apenas online [1].

A estimativa resultante dos primeiros métodos de deconvolução é o formato de proporções de células consideradas na matriz de referência que somam 100% (Tabela 4). Como comentado anteriormente, para avaliação da composição de células em amostras de tecido essa abordagem não faz muito sentido se não for considerada a informação adicional de pureza do tumor, ou da quantidade de infiltrado inflamatório presente em cada amostra. Uma estimativa de 90% de infiltrado de linfócitos possui um peso diferente se este infiltrado representa 10% ou 70% da amostra de tumor. Neste ponto, considerações importantes de normalização devem ser feitas na interpretação dessas proporções.

No método *EPIC*, a solução proposta foi de acrescentar a métrica de células "desconhecidas" no resultado. No método *TIMER*, a pureza do tumor é considerada e a estimativa do infiltrado resulta em valores absolutos. Outros métodos, como *CIBERSORT*, também estimam valores de abundância que podem ser melhor correlacionados com os valores "reais". De qualquer forma, espera-se que estes valores de abundância sejam mais informativos que as proporções. Para avaliar qual a melhor estratégia de análise em câncer gástrico (lista de marcadores, assinatura de referência e método de deconvolução), o próximo passo foi realizar uma análise sistemática dos algoritmos e métricas de desempenho.

O crescimento desta área nestes últimos anos foi considerável e, desde o início deste trabalho, novas ferramentas foram publicadas, adaptadas de métodos existentes ou mesmo propondo um consenso das atuais (Sturm et al. 2020; Jiménez-Sánchez et al. 2019; Finotello et al. 2019; Monaco et al. 2019; Hunt et al. 2019). Consequentemente, para este *benchmarking*, nem todos os métodos listados foram utilizados, enquanto outros foram levantados posteriormente e considerados para *pipeline* proposto nesta tese.

É importante mencionar que o pacote CellMix foi descontinuado (não foi mais atualizado) e, portanto, fez parte apenas do início de desenvolvimento deste

[1] https://cibersortx.stanford.edu/

trabalho e do *benchmarking*. No entanto, seu principal método (*lsfit*) foi a base para o método *TIMER*, adaptado para dados de tecido (microambiente tumoral), que é considerado no *pipeline* final.

5.1.3 Desempenho dos algoritmos de deconvolução de células imunes

Muitas listas de genes marcadores, bem como diferentes matrizes de assinatura, foram desenvolvidas previamente em conjunto com as estratégias de análise. Assim, buscamos investigar quais dessas abordagens possuem melhor desempenho a fim de aplicá-la(s) no nosso *pipeline* para estimação da composição do infiltrado inflamatório, considerando padrões de expressão gênica e composição celular no microambiente tumoral.

Para avaliar o desempenho dos algoritmos e listas de marcadores em estimar as proporções de células imunes, utilizamos o *pipeline TIMER* (Figura 22) para construção da matriz de assinatura para 5 e 7 células, presentes no painel de referência HPCA e esperadas serem encontrados no infiltrado tumoral. Dados de expressão gênica (FPKM) de 380 amostras de adenocarcinoma gástrico disponíveis pelo banco de dados público do consórcio *The Cancer Genome Atlas* (TCGA) foram utilizadas como mistura. As proporções estimadas foram comparadas com os resultados prévios de *deep learning* com a análise de lâminas coradas com HE das mesmas amostras de AdG (Saltz et al. 2018). A Figura 22 ilustra o *workflow* dessas análises de *benchmarking* dos algoritmos de deconvolução na análise de infiltrado inflamatório.

O método *lsfit* contém duas variações aqui denominadas de *nnls-lsfit* e *qnnls-lsfit*, sendo que esta última considera a opção de normalização dos dados por quantis. Ambas estratégias foram testadas. Para o método *EPIC*, para que seus resultados pudessem ser comparados com os demais, a proporção de "outras células" foi desconsiderada e as proporções remanescentes normalizadas para somarem 1 (ou 100%); aqui os resultados do *EPIC* estão indicados como *nEPIC*.

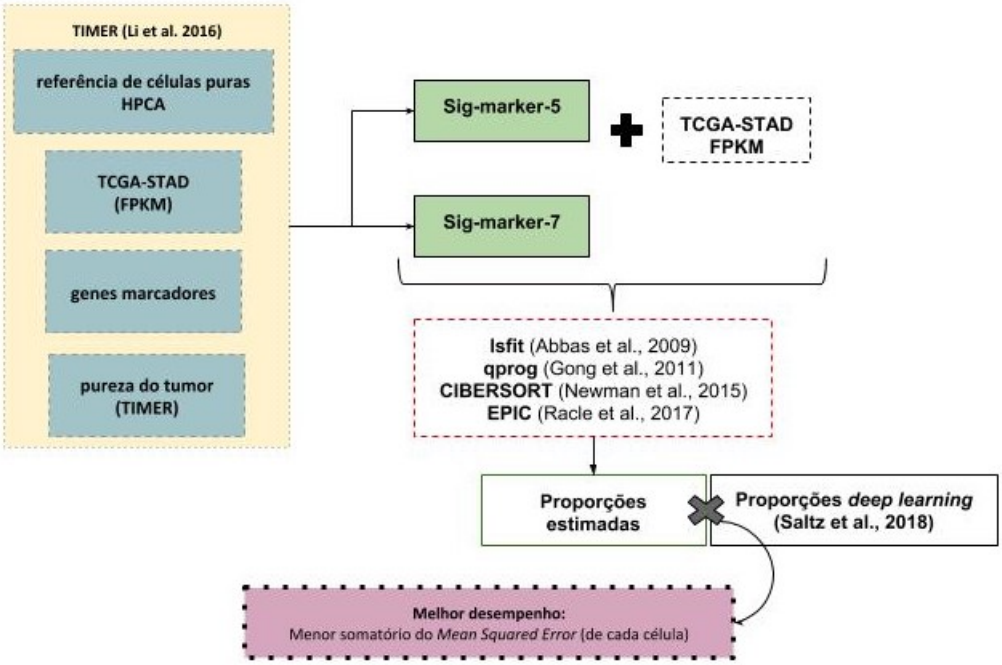


Figura 22 – Esquema das análises de *benchmarking*. O método TIMER (Li et al. 2017) foi utilizado para testar assinaturas customizadas com as listas de genes encontradas pela revisão de literatura considerando 5 e 7 células imunes principais. Quatro métodos principais foram selecionados para serem comparados utilizando dados de câncer gástrico do TCGA.

A matriz de assinatura de 5 células era composta por: linfócitos B, T CD4+, T CD8+, macrófagos e células dendríticas (DC). Para comparação com os resultados de *deep learning*, as proporções de linfócitos B, T CD4+ e T CD8+ obtidas pela deconvolução foram somadas e denominadas como *"Lymphocytes"*. Dos resultados de *deep learning* foram desconsideradas as proporções de eosinófilos e mastócitos, tipos celulares que não estão representados no *dataset* de referência HPCA. Além disso, neste teste de 5 células, os neutrófilos que não foram abordados na matriz de assinatura de referência. Todos os valores de proporções foram normalizados para somar 1. Com isso, das 5 células foi feita comparação com 3 células dos resultados de *deep learning*: linfócitos, macrógafos e DC (Figura 23).

A matriz de assinatura de 7 células era composta por: linfócitos B, T CD4+, T CD8+, células NK, neutrófilos, macrófagos e DC. Para comparação com os resultados de *deep learning*, as proporções de linfócitos B, T CD4+ e T

CD8+ e células NK obtidas pela deconvolução foram somadas e denominadas como "*Lymphocytes*". Dos resultados de *deep learning* foram desconsiderados eosinófilos e mastócitos, da mesma maneira que anteriormente. Com isso, das 7 células foi feita comparação com 4 células: linfócitos, neutrófilos, macrógafos e DC (Figura 24).

É possível verificar que a combinação do algoritmo e a lista de marcadores gera valores distintos quando comparado ao método MSE em ambos os testes (5 e 7 células). No geral, as combinações com pior desempenho foram observadas nas análises com 5 células (Figura 23) e 7 células (Figura 24). Em ambos os testes, pode-se observar que todos os algoritmos combinados com a lista Newman2015 apresentam desempenho semelhante e com MSE relativamente baixo quando comparado com as demais combinações. Esses resultados nos levam a concluir que Newman2015 pode ser considerada uma boa opção para análise de infiltrado inflamatório em AdG com qualquer algoritmo.

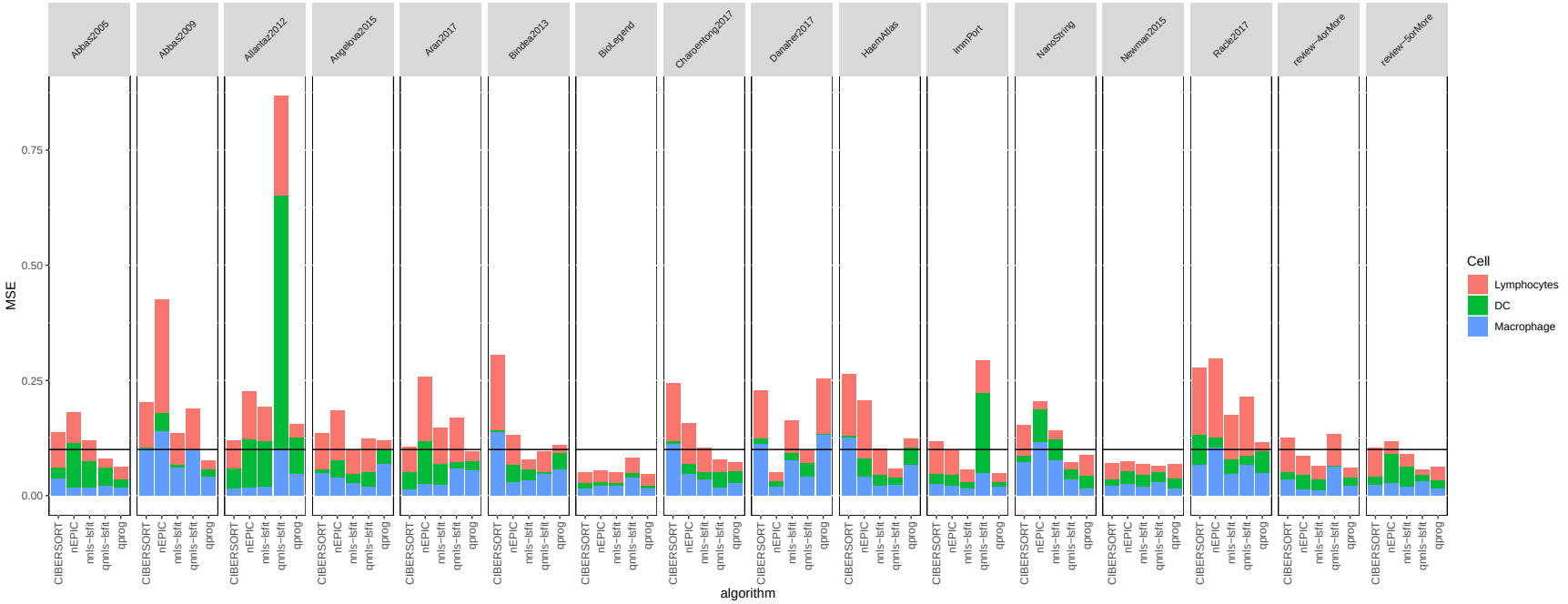


Figura 23 – Distribuição do MSE para cada lista de assinaturas de genes considerando 3 células analisadas (linfócitos (B, T CD4+ e T CD8+), macrófagos e DC), para cada um dos métodos. Em cada barra, os valores de MSE comparando as estimativas de proporção de cada célula, geradas a partir da análise de 5 células com os resultados de *deep learning* (Saltz et al. 2018), resultando o erro total para cada teste.

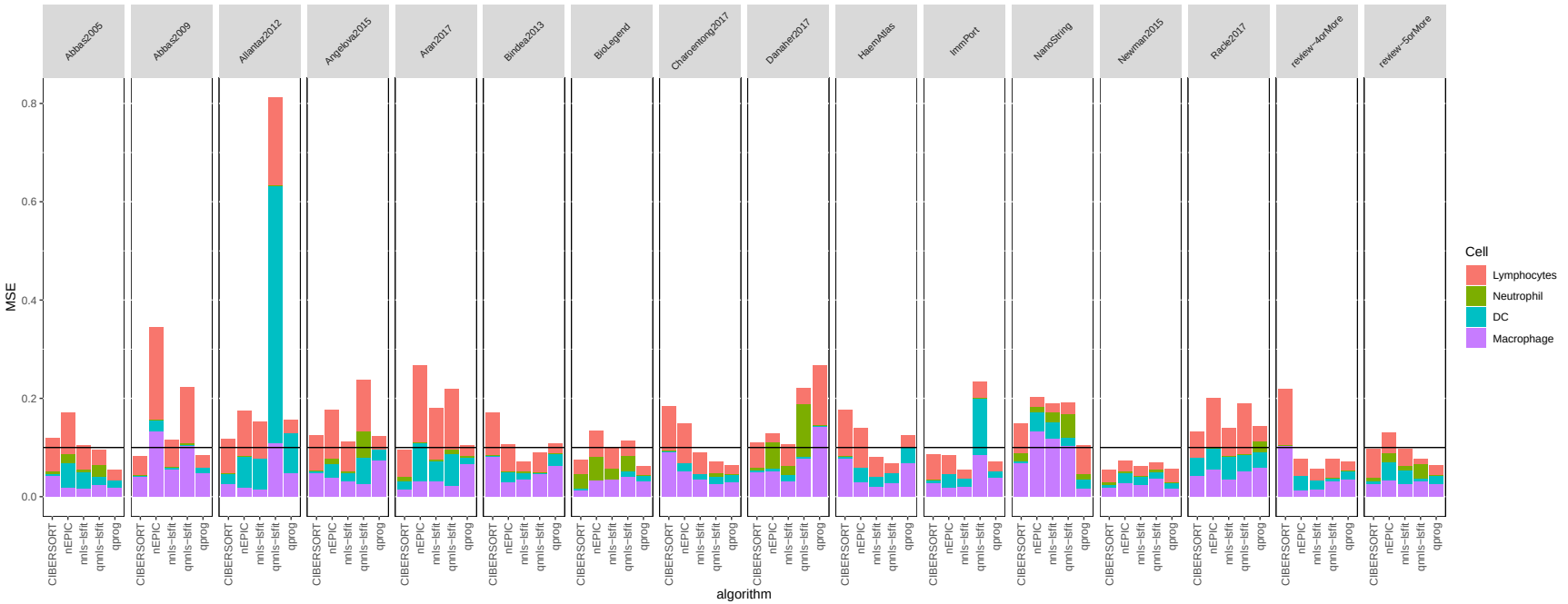


Figura 24 – Distribuição do MSE para cada lista de assinaturas de genes considerando 4 células analisadas (linfócitos (B, T CD4+, T CD8+ e células NK), neutrófilos, macrófagos e DC), para cada um dos métodos. Em cada barra, os valores de MSE comparando as estimativas de proporção de cada célula, geradas a partir da análise de 7 células com os resultados de *deep learning* (Saltz et al. 2018), resultando o erro total para cada teste.

Fica claro que o melhor desempenho de alguns algoritmos é dependente da lista de marcadores considerada. A combinação com menor somatório de MSE foi qprog-BioLegend com somatório de 0,0454, na análise de 5 células. Já para 7 células, a combinação com melhor desempenho foi CIBERSORT-Newman2015 com somatório de MSE de 0,0546. No geral, linfócitos e macrófagos são os tipos celulares que apresentaram maior erro entre as listas; sendo os neutrófilos o tipo celular com melhor estimativa (Figuras 23 e 24).

A fim de avaliar o desempenho de CIBERSORT-Newman-2015 foi feita a correlação entre as estimativas de cada célula com genes marcadores característicos, indicados na Tabela 2 (Seção 4.2). De importância, os valores proporções de células estimadas foram normalizados com os valores de infiltrado (1 - pureza do tumor), estimado pelo algoritmo ESTIMATE (Yoshihara et al. 2013). Como pode ser observado, a maioria das proporções estimadas de células e seus genes marcadores tiveram correlações significativas ($p<0,001$) positivas com valores de intermediário a forte (Figura 25).

Outra forma de avaliar o desempenho considerando mais subtipos celulares foi por meio da correlação entre as estimativas de proporção por deconvolução (normalizadas) com os *scores* gerados pelo método *xCell* (Aran et al. 2017). Foi utilizada esta segunda abordagem por ser uma estratégia de análise muito utilizada em estudos de expressão gênica e vias ativas. No entanto, uma desvantagem dessa abordagem, superada pelos métodos de deconvolução, é que os valores (*scores*) gerados para um tipo celular não são comparáveis com outra célula, podendo as amostras serem comparadas entre si para um mesmo tipo celular. Assim, considerando as 7 células e os resultados de CIBERSORT-Newman2015, foi feita a correlação das estimativas com os *scores* do *xCell*, mostrado na Figura 26.

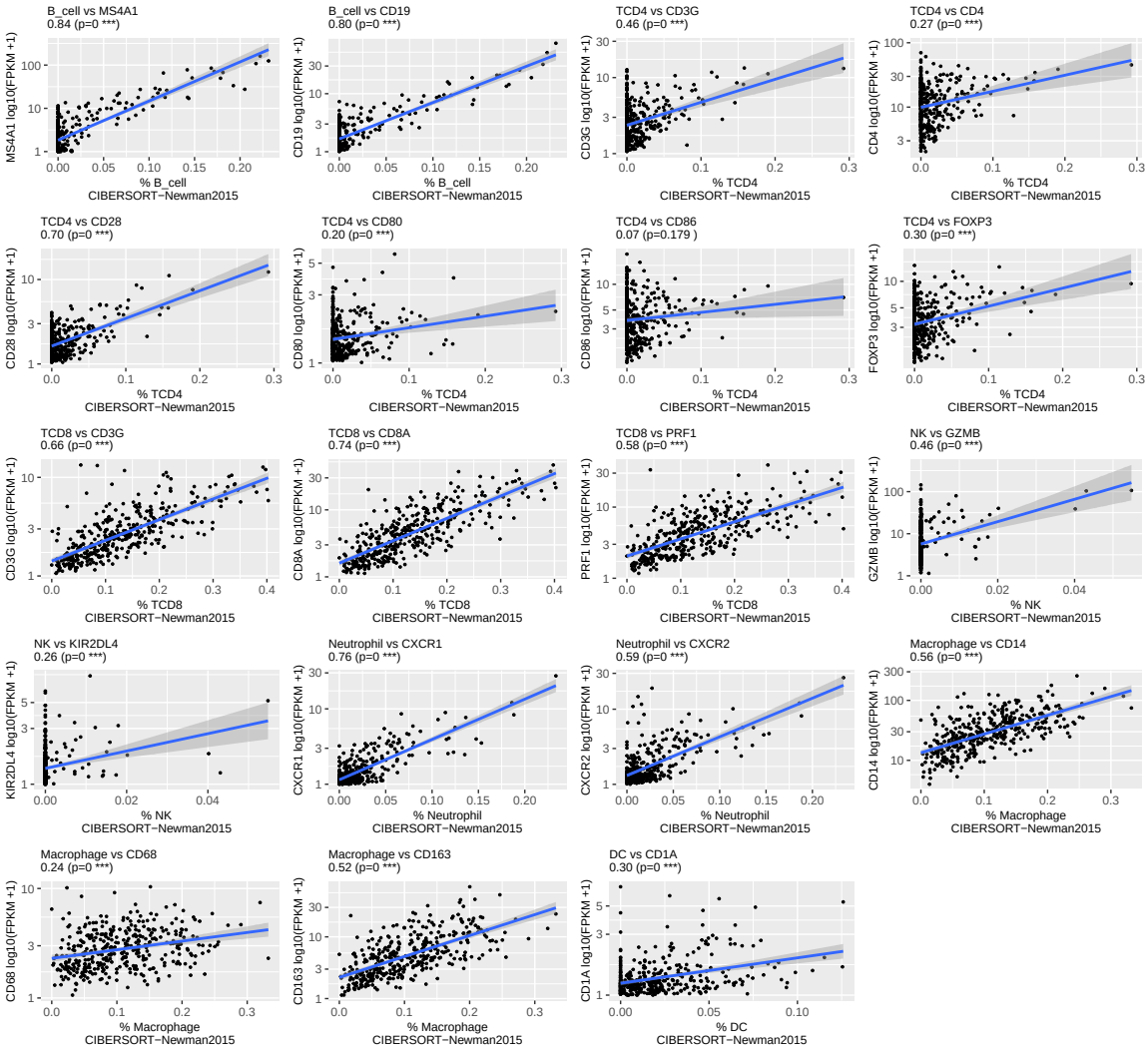


Figura 25 – Correlação de Pearson entre as proporções estimadas de 7 células pelo CIBERSORT-Newman2015 e a expressão de genes marcadores principais característicos de cada célula.

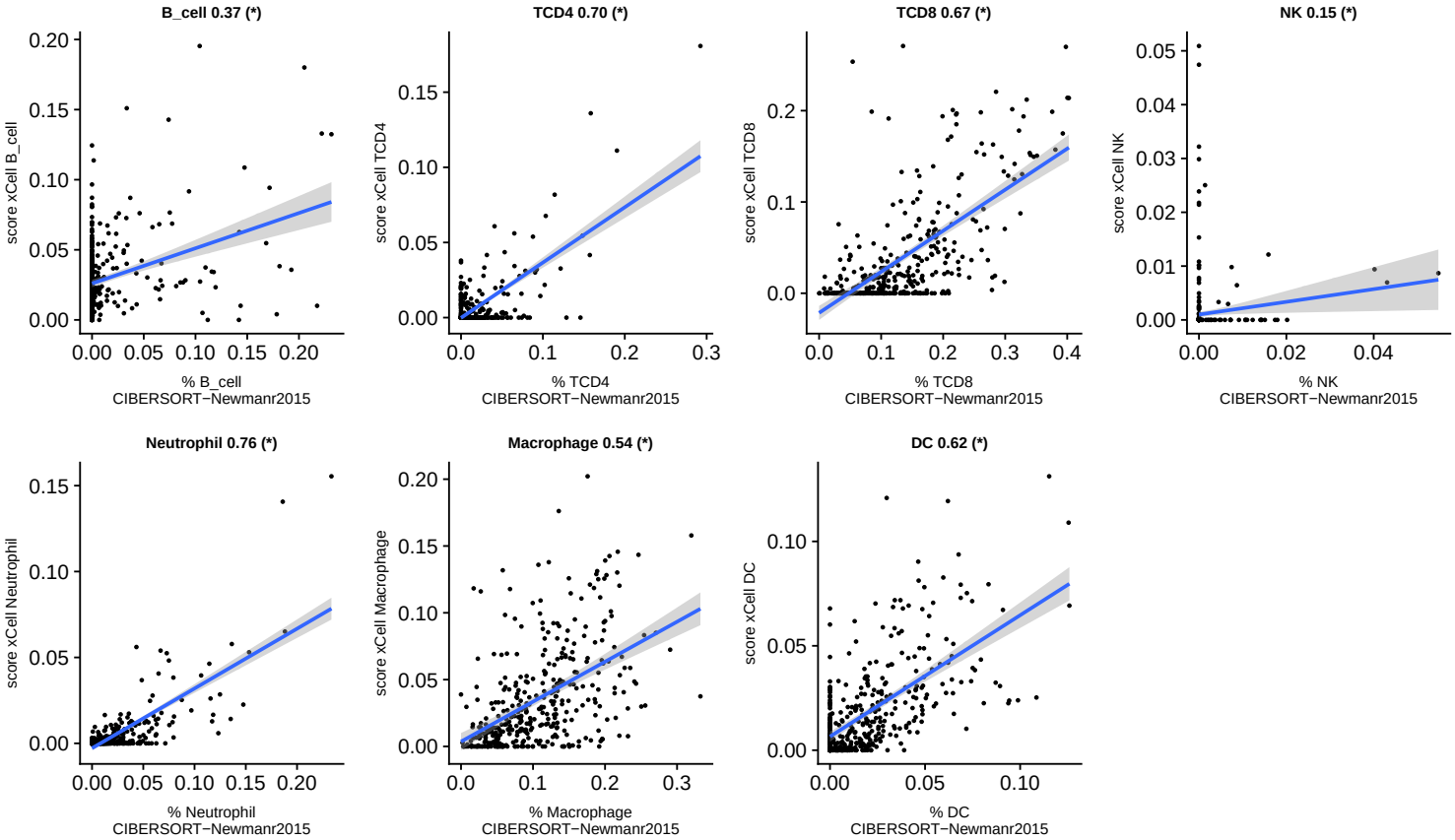


Figura 26 – Correlação de Pearson das proporções estimadas de células CIBERSORT-Newman2015 com *scores* de *xCell* (Aran et al. 2017). * p<0,05 .

É possível verificar que todas as correlações foram positivas e significativas entre os *scores* das células pelo *xCell* com as estimativas da melhor combinação encontrada para 7 células (Figura 26). Entre os linfócitos, a melhor correlação foi com os linfócitos T (rho=0,70), enquanto que a pior foi com as células NK (rho=0,15).

Nesta primeira parte do *benchmarking*, concluímos que a estimação do infiltrado inflamatório é dependente de uma lista apropriada de genes marcadores, uma vez que a performance dos algoritmos varia conforme essa lista. Para análise prévia da composição do infiltrado inflamatório em AdG foi escolhida a combinação mais abrangente e com menor erro, ou seja, CIBERSORT-Newman2015 considerando 7 células imunes básicas. Os monócitos não foram incluídos pois para o contexto de microambienta tumoral espera-se encontrar células já diferenciadas em macrófagos.

Considerando a lista de genes de Newman2015, que mostrou melhor desempenho global independente do algoritmo, continuamos sua avaliação em estimar a composição do infiltrado em AdG. Esta lista compreende 547 genes extraída da assinatura de referência denominada LM22, contida no algorítmo *CIBERSORT* e contém 22 subtipos celulares. As amostras de AdG foram então analisadas pelo o método *CIBERSORT* original a fim de observar seu desempenho e, consequentemente, investigar detalhadamente a composição do infiltrado inflamatório considerando um número amplo de subtipos imunes com objetivo de entender melhor as respostas imunes envolvidas e mecanismos de escape imunológico.

Além de mais subtipos celulares, utilizamos o parâmetro de estimar as quantidades absolutas (e não relativas), cujos valores são mais informativos sobre a quantidade real de infiltrado no geral. Também consideramos o nível de significância da estimativa ($p<0,05$), que considera a distribuição dos dados normalizados de expressão, não adequados para algumas amostras. Utilizamos o parâmetro máximo de 1000 permutações e com a opção de normalização por quan-

tis desabilitada (como recomendado para dados de RNA-Seq). Das 380 amostras tumorais do projeto de AdG no TCGA (STAD), 215 (56%) tiveram estimativas significativas ($p < 0{,}05$) e foram consideradas para avaliação de desempenho e análises posteriores de prognóstico em AdG.

Utilizando os dados prévios de *deep learning* das mesmas amostras, devido aos tipos de dados serem diferentes (quantidades absolutas e proporções), a comparação foi feita utilizando a correlação de Pearson entre a composição absoluta do infiltrado inflamatório de 22 células estimada pelo *CIBERSORT* e as proporções das 6 células por *deep learning* (Figura 27).

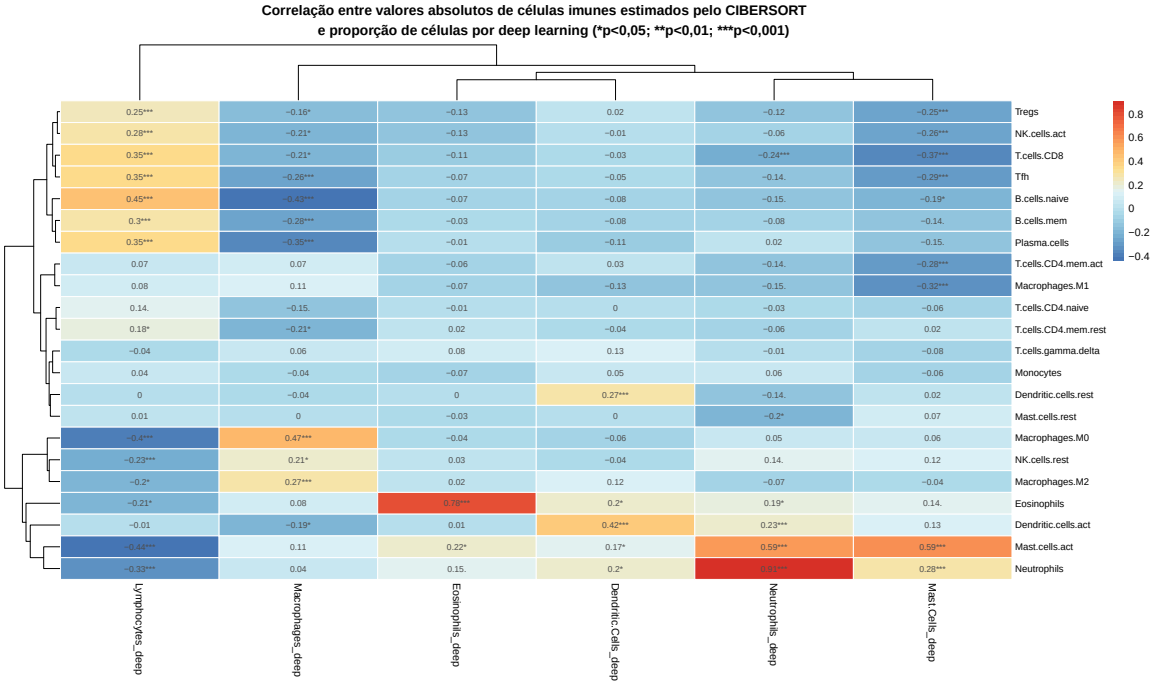


Figura 27 – *Heatmap* de correlação de Pearson entre as quantidades absolutas estimadas com método *CIBERSORT* (linhas) e proporções de células por *deep learning* (colunas) (Saltz et al. 2018).

Pode-se observar correlações significativas entre os tipos celulares correspondentes, mais fortes para células mieloides da resposta imune inata (mastócitos, neutrófilos e eosinófilos), bem como formação de *clusters* entre os dois tipos principais tipo celulares linfócitos B e T, com correlações intermediárias (Figura 27).

Para correlação com a expressão gênica de marcadores principais das célu-

las imunes, utilizamos os genes descritos na Tabela 2 (Seção 4.2). Como observado anteriormente na avaliação de marcadores imunes, não existe um marcador ideal, específico para tipo celular, pois o conjunto se mostra mais importante e por essa razão é que a análise por deconvolução é mais completa. No entanto, mesmo os marcadores imunes serem compartilhados entre os tipos celulares, essa característica torna-se interessante sob o ponto de vista funcional, pois a maior expressão de um certo marcador associado a um determinado tipo celular auxilia no melhor entendimento da resposta envolvida, principalmente considerando a composição de mais subtipos celulares.

A correlação global entre a composição absoluta de células imunes e a expressão de genes marcadores principais, além de servir como uma avaliação de desempenho, indica quais células estão envolvidas na resposta imune. Como validação destaca-se correlações significativas para estimativas de Macrófagos, como entre *CD163* e *CD86* e Macrófagos M2 (rho=0,68 e rho=069, respectivamente). Além disso, correlações significativas entre células T CD8+ e *CD8A* (rho=0,88), células Treg e *FOXP3* (rho=0,62), células memória com *MS4A1* (CD20) (rho=0,62), e Neutrófilos com *CXCR1*, *CXCR2* (rho=0,91 e rho=0,67) (Figura 28, padj<0,01).

A correlação global entre a composição absoluta de células imunes pelo *CIBERSORT* e os *scores* pelo *xCell* (Figura 29) também foram altas e significativas para linfócitos T CD8+, subtipos de linfócitos T CD4+ e B, macrófagos, neutrófilos. As correlações foram intermediárias, mas significativas para células dendríticas e eosinófilos e baixas ou não significativas para os subtipos de mastócitos e células NK. Os diferentes subtipos funcionais destas células entre as duas abordagens podem explicar estes valores de correlação menores.

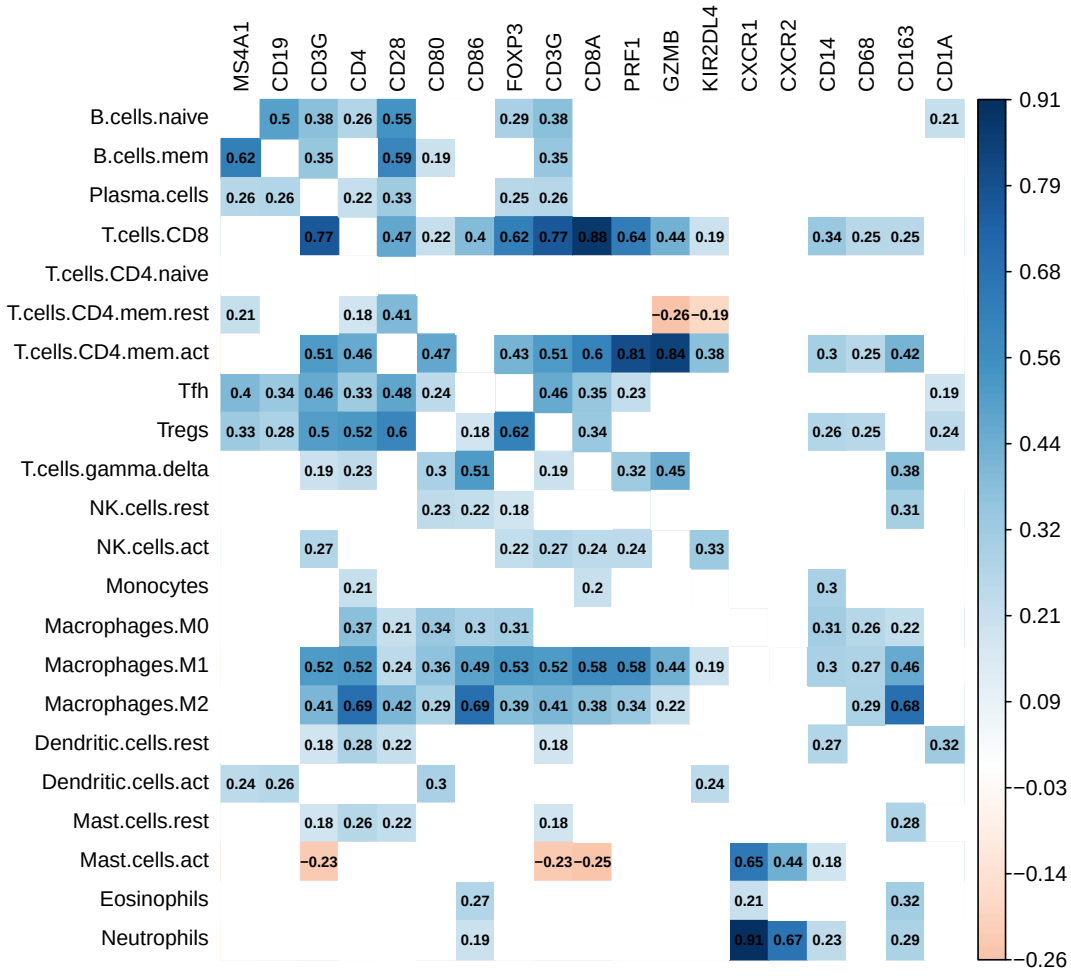


Figura 28 – Correlograma de Pearson das quantidades absolutas estimadas com método *CIBERSORT* (linhas) e expressão de genes principais marcadores (colunas). Valores apresentados considerando p *value* ajustado <0,01

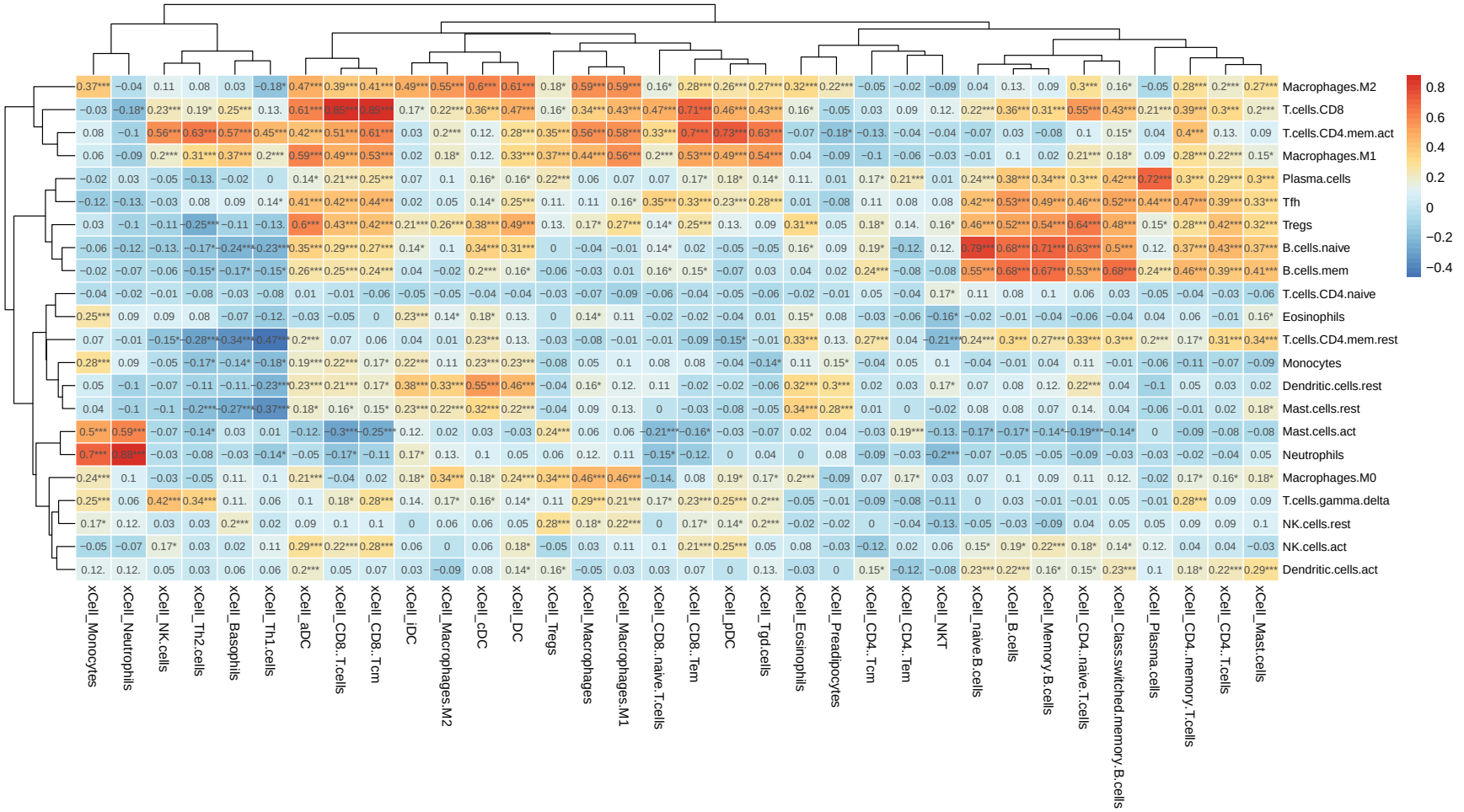


Figura 29 – *Heatmap* de correlação de Pearson entre as quantidades absolutas de células imunes estimadas com método *CIBERSORT* (linhas) e *scores* estimados pelo método *xCell* (colunas) a partir do FPKM de amostras TCGA-STAD.

Em resumo, as análises de *benchmarking* dos métodos de deconvolução aplicados às amostras de AdG do TCGA indicaram que o método CIBERSORT apresentou melhor desempenho global. Este método então foi escolhido e aplicado para avaliação do infiltrado inflamatório em câncer gástrico. Considerando a grande quantidade de listas de genes marcadores e métodos de deconvolução avaliadas no momento e outros que tem sido publicados na literatura, concluimos que para cada subtipo tumoral deve haver uma melhor combinação de lista de marcadores/assinatura de referência com algoritmo de deconvolução. Assim, somado à estimatição da composição do infiltrado inflamatório por deconvolução, deve haver estratégias de validação que indiquem que aquela é a melhor combinação a ser aplicada. Além disso, é importante considerar a normalização dos dados de expressão gênica e remoção de *batch effect* e *outliers*.

5.1.4 Proposta de *pipeline* para avaliação do contexto imunológico

A maioria dos algoritmos de deconvolução estão disponiveis em pacotes do R, e portanto, o *pipeline* proposto reúne as opções de métodos disponíveis nesta linguagem. Além disso, reúne listas de genes marcadores e assinaturas de referência levantados pela revisão de literatura bem como avalição dos resultados encontrados com as combinações de interesse. Com base nos resultados obtidos pelo *benchmarking*, a proposta de *workflow* de análise do contexto imunológico do câncer utilizando dados de expressão gênica envolve os seguintes pontos, esquematizados na Figura 30:

- Lista de marcadores e assinaturas de referência:

  A lista de genes marcadores compreende as opções descritas no *benchmarking*, na Tabela 3. Definida a lista de genes de interesse, a assinatura de referência é construída com base no método do TIMER (Li et al. 2017) considerando 7, 8 ou até x subtipos celulares, de acordo com a referência HPCA (Mabbott et al. 2013). Há também opções de assinaturas prontas:

LM22, LM6, TIL10, BRef e TRef, construídas em conjunto com alguns métodos de deconvolução, como descrito na Tabela 3.

- Lista de algortimos de deconvolução:

  A lista final de algoritmos de deconvolução desconsidera os primeiros métodos do pacote CellMix (Gaujoux e Seoighe 2013) que foi descontinuado. Há métodos que resultam em proporções e valores de abundância (Tabela 4). O método *CIBERSORT* (Newman et al. 2015) é o único que resulta em ambos tipos de resultado. O método mais atualizado CIBERSORTx está disponível apenas de forma online, mas o tratamento dos dados de expressão gênica por meio de abordagens de normalização, remoção de *batch effect*, ou remoção de amostras *ouliers* podem aumentar a qualidade dos resultados.

- Avaliação do método por tipo tumoral com dados de *deep learning*:

  Para 13 tipos de tumores estão disponíveis dados da composição de infiltrado baseado em *deep learning* (Saltz et al. 2018). Uma das formas disponíveis de avaliação da estimação da composição do infiltrado por deconvolução é a correlação entre os resultados para amostras do TCGA.

- Avaliação dos resultados com *xCell*:

  Outra forma de avaliar e fortalecer os achados da composição do infiltrado inflamatório é por meio da correlação com os resultados do método *xCell* (Aran et al. 2017).

- Avaliação dos resultados com correlação com genes marcadores principais e normalização pela pureza do tumor:

  Por fim, outra maneira de corroborar os achados da composição do infiltrado (proporções) é a correlação entre a composição de cada célula absoluta ou proporções normalizadas e a expressão de genes marcadores principais.

A normalização proposta é a multiplicação dos valores de proporção pelo equivalente a quatidade de células não tumorais na amostra, valor obtido por 1 - pureza do tumor. A pureza do tumor pode ser obtida pelo método ESTIMATE (Yoshihara et al. 2013), que também se baseia em dados de expressão gênica.

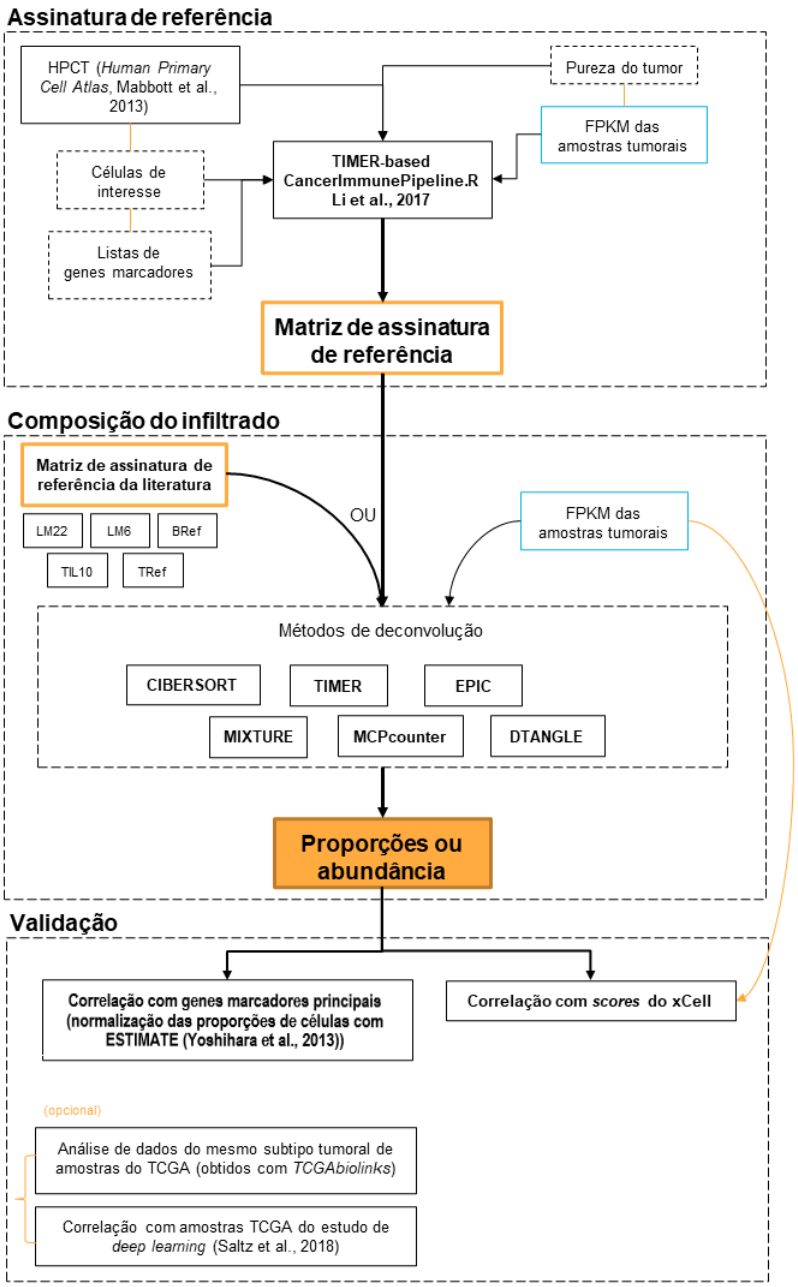


Figura 30 – Proposta e considerações para análise da composição do infiltrado inflamatório utilizando métodos de deconvolução.

5.2 CONTEXTO IMUNOLÓGICO EM CÂNCER GÁSTRICO E SEUS ASPECTOS PROGNÓSTICOS

5.2.1 Infiltrado inflamatório nos subtipos de câncer gástrico

Com base nos resultados do *benchmarking* descritos na seção anterior, a composição do infiltrado inflamatório foi avaliada de acordo com o os subtipos histológicos e moleculares em câncer gástrico. Para isso, utilizamos a melhor combinação de algoritmo e lista de marcadores (CIBERSORT e Newman2015) para estimar a composição de sete tipos básicos de células imunes.

Estudos prévios já associaram os subtipos histológicos de Lauren a diferentes prognósticos de sobrevida em AdG, porém não há um consenso e os resultados variam entre diferentes coortes (Chen et al. 2016). Dessa forma, este aspecto clínico, embora prático, não é uma opção interessante de classificação definitiva para o prognóstico. Os subtipos moleculares do TCGA (GS, CIN, MSI e EBV) (Bass et al. 2014) por sua vez, apresentam associações mais sólidas sendo os subtipos MSI e EBV com melhor prognóstico (Sohn et al. 2017; Bass et al. 2014). Uma das características destes subtipos é serem considerados "tumores quentes"ou imunogênicos e apresentarem maior quantidade de infiltrado inflamatório (Wang et al. 2016; Panda et al. 2018).

No presente estudo, observamos que há algumas tendências na composição do infiltrado inflamatório entre os subtipos histológicos de Lauren. A única diferença significativa foi a menor proporção de neutrófilos no subtipo difuso. Entre as tendências, o subtipo difuso apresentou maior proporção de linfócitos B e o subtipo intestinal, maior proporção de macrófagos (Figura 31).

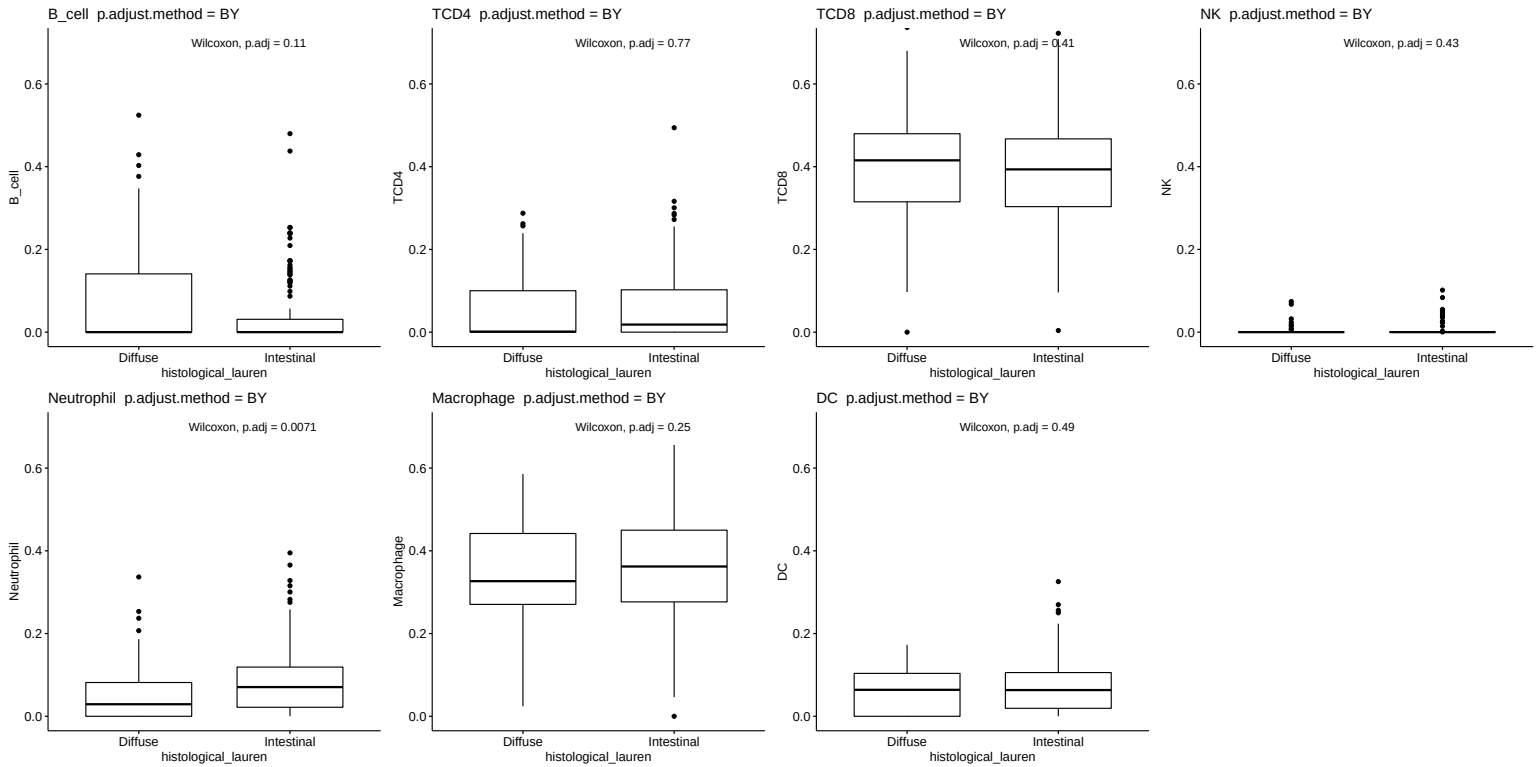


Figura 31 – *Boxplots* das proporções de células estimadas pelo CIBERSORT-Newman2015 nos subtipos histológicos de AdG.
Comparação entre os subtipos intestinal e difuso; teste *U de Mann-Whitney*.

Em relação aos subtipos moleculares, o subtipo GS (*genomically stable*), que está mais associado ao subtipo histológico difuso, apresentou uma maior quantidade de linfócitos B e T CD4+ em comparação com os demais subtipos moleculares (Figura 32). Em contraste, os subtipos MSI (*microsatellite instable*) e EBV apresentaram maior proporção de linfócitos T CD8+ ($p<0,05$). Esse cenário pode ser um indicativo de melhor resposta ao tratamento com imunoterapia além de poder estar associado a melhores taxas de sobrevida, principalmente no subtipo EBV (Figura 33).

O subtipo molecular CIN (*chromosomal instability*) apresentou enriquecimento de macrógafos e neutrófilos. Os neutrófilos também se apresentam em maior proporção no subtipo MSI (Figura 32). Estes resultados são concordantes com os resultados encontrados para os subtipos histológicos, considerando que o subtipo MSI está associado ao subtipo intestinal de Lauren. No caso do subtipo EBV, esté é o subtipo molecular com menos neutrófilos, o que o diferencia do subtipo MSI (Figura 32).

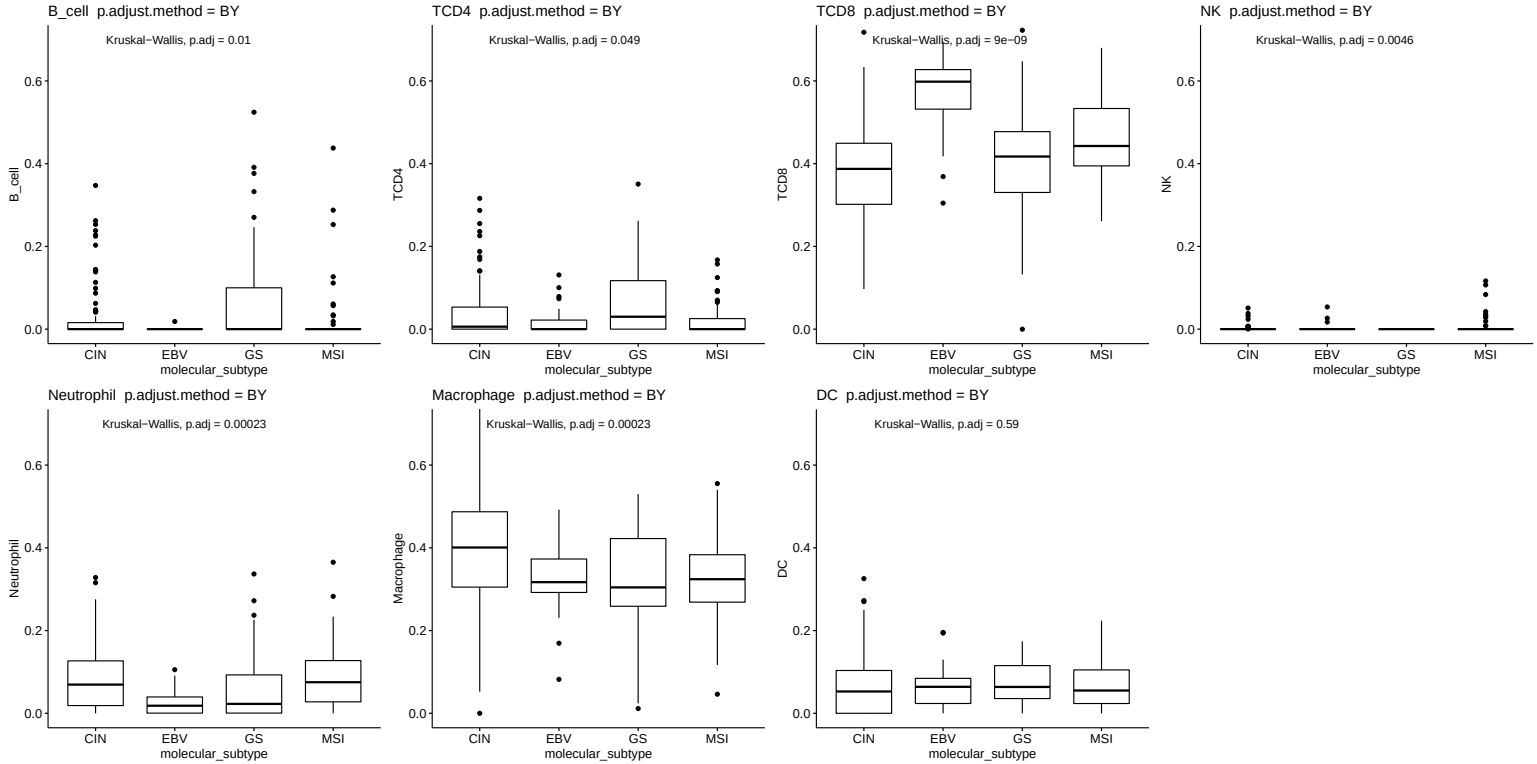


Figura 32 – *Boxplots* das proporções de células estimadas com CIBERSORT-Newman2015 nos subtipos moleculares de AdG.
Comparação entre os subtipos moleculares; teste *Kruskal-Wallis*.

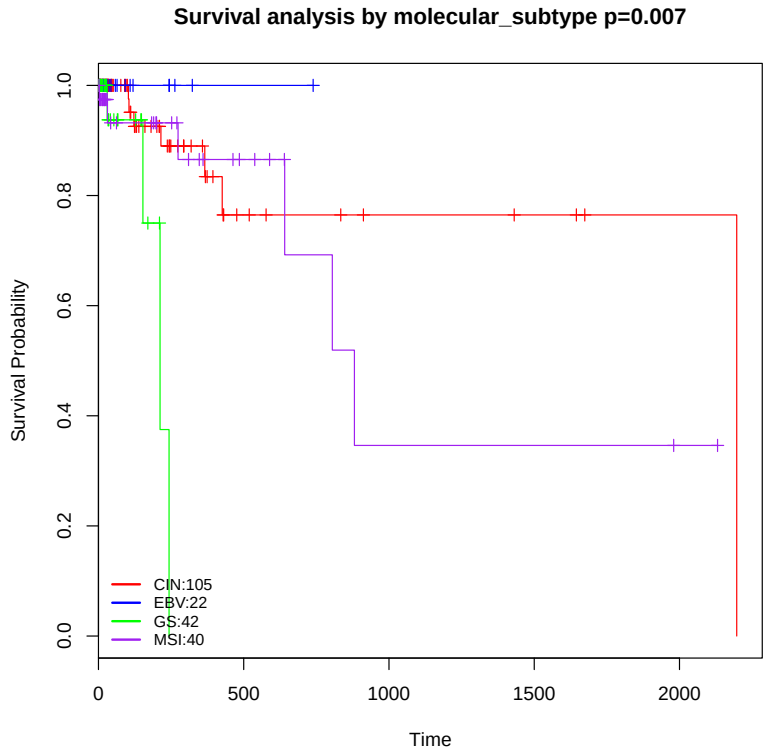
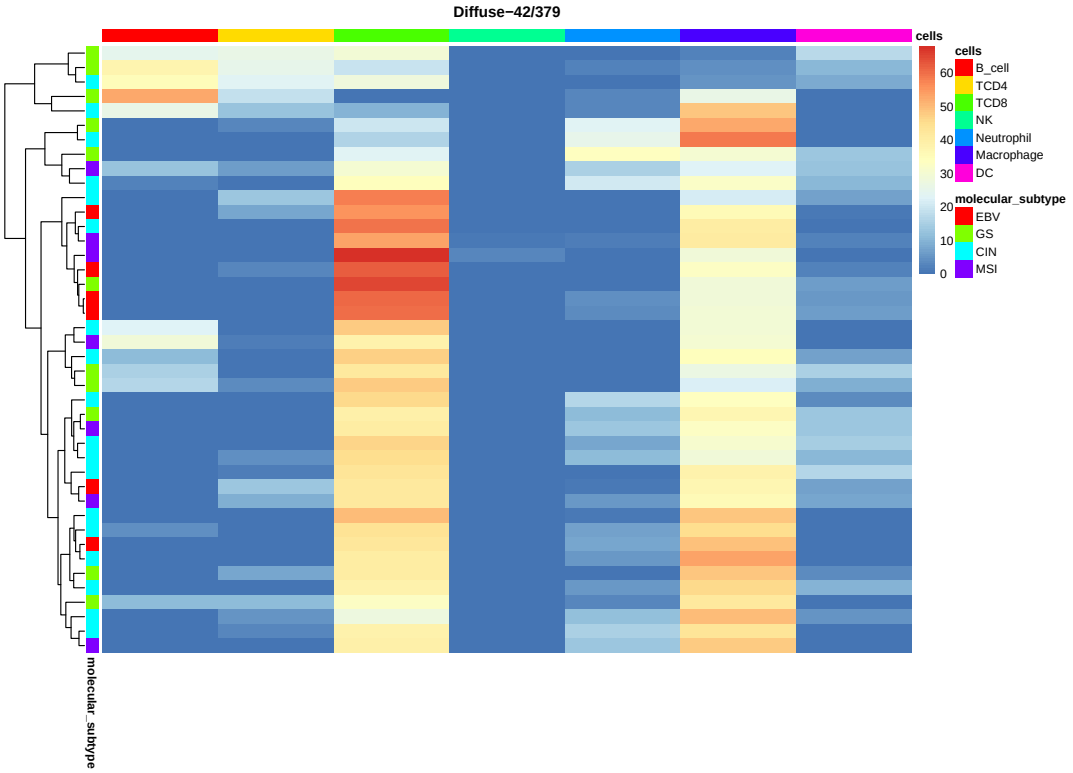


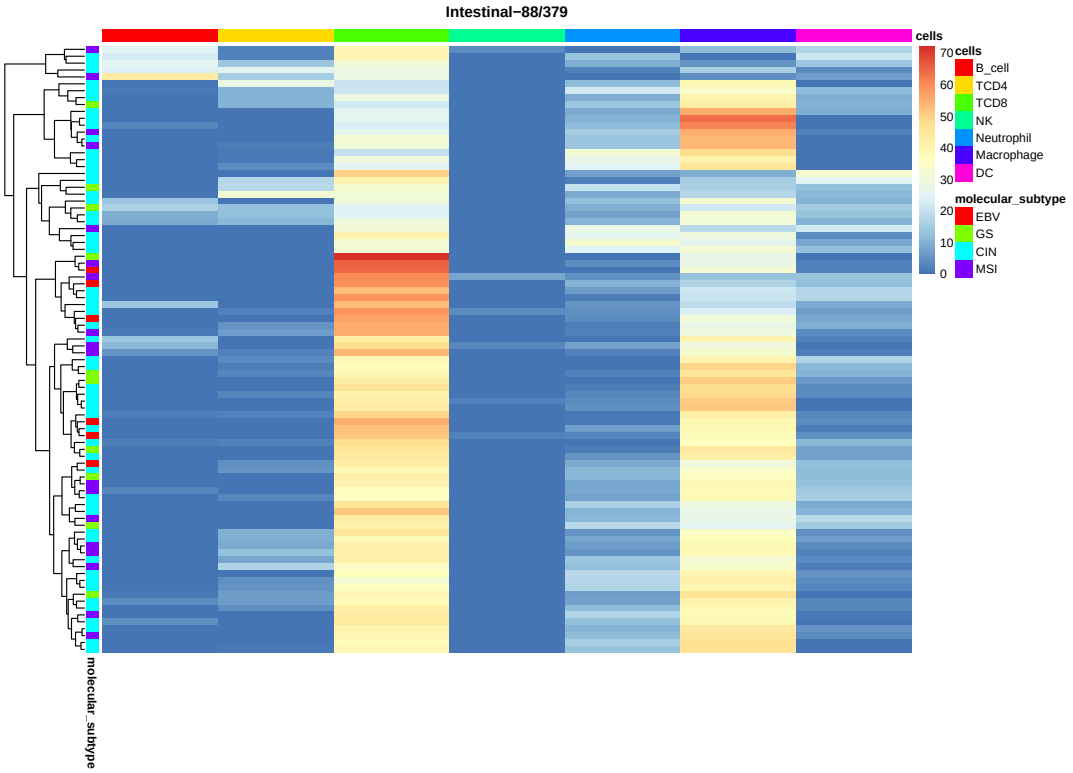
Figura 33 – Sobrevida global entre os subtipos moleculares de AdG. Os pacientes metastáticos foram desconsiderados. Tempo em dias.

A composição do infiltrado inflamatório considerando os 7 tipos básicos de células imunes não estratifica os subtipos histológicos ou moleculares, assim como ambos tipos de classificação se misturam entre si: todos os subtipos moleculares possuem representantes de ambos subtipos histológicos (Figura 34). No entanto, independente do subtipo histológico, as amostras MSI e EBV tendem a ser o grupo com maior proporção de linfócitos T CD8+, por exemplo. Isto sugere que há outros fatores influenciando a composição do infiltrado inflamatório em AdG que poderiam explicar melhor estes grupos de pacientes.

Além das comparações da composição do infiltrado inflamatório entre os principais subtipos de AdG, o valor prognóstico das principais variáveis clínicas associadas a AdG e da composição do infiltrado inflamatório foi avaliado. Para isso, foi construído um modelo simples de Cox (Figura 35). Os resultados não apontaram nenhuma associação significativa com exceção das variáveis com valor



(a) Subtipo difuso



(b) Subtipo intestinal

Figura 34 – *Heatmaps* da proporção estimada das 7 células nos subtipos intestinal juntamente com a classificação molecular.

prognóstico já conhecido como a idade ao diagnóstico e estadiamento.

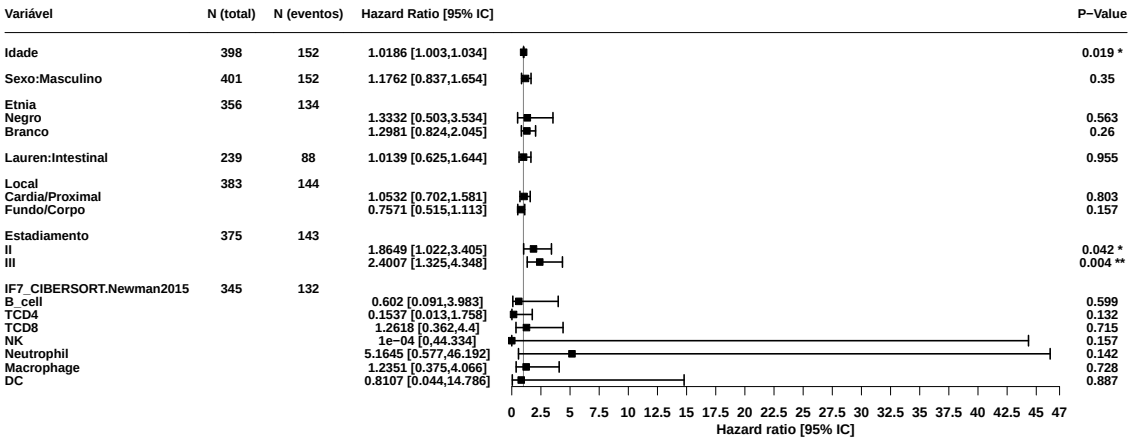


Figura 35 – Modelo simples de Cox de sobrevida global em amostras de AdG (TCGA-STAD).

Estes resultados preliminares foram apresentados e discutidos em eventos científicos (Apêndices **??** e **??**). De importância, estas análises apontaram que, mais do que a comparação entre os subtipos celulares individualmente, a combinação das proporções de células linfóides e mielóides no infiltrado tumoral é que pode estar associada a diferentes prognósticos (Stockmann et al. 2014; Becht et al. 2016; Gajewski et al. 2013; Wang et al. 2016; Whiteside 2008). Além disso, é importante considerar o estado funcional destas células: mais imunosupressor/tolerogênico ou mais efetor/inflamatório/citotóxico. Para entender melhor o contexto imunológico em câncer gástrico como fator prognóstico, a análise foi expandida considerando mais subtipos celulares com a assinatura de LM22 e utilizando quantidades absolutas da composição de células imunes, correlação dos resultados com a expressão de genes de moléculas efetoras, além de relacionar o perfil clínico e molecular dos pacientes de AdG aplicando uma nova classificação baseada em assinaturas mutacionais.

5.2.2 Características clínico-patológicas, assinaturas mutacionais e interação com contexto imunológico

Encontra-se bem documentado que mutações somáticas no genoma do câncer são o resultado de processos mutacionais endógenos e exógenos. Por exemplo, exposição a raios ultra violeta (UV), hábito de fumar cigarro, mutações induzidas pelos genes AICDA e da família APOBEC, responsável pela edição de ácidos nucléico, e a deficiência em mecanismos de reparo do DNA, entre outros mecanismos (Alexandrov et al. 2020). Particularmente, a deficiência na via de reparo de mal pareamento do DNA (MMR, *mismatch repair*) encontra-se associada a um fenótipo conhecido com instabilidade de instabilidade (MSI, *microsatellite instability*) que, por sua vez, é um biomarcador preditivo para resposta à imunoterapia (Zhang et al. 2021; Zhao et al. 2019; Abida et al. 2019).

Embora 4 assinaturas sejam associadas à deficiência de MMR (dMMR), permanece ainda a ser estabelecido sua relevância prognóstica bem como sua etiologia. Nesse contexto, investigamos uma coorte com 787 amostras de adenocarcinoma gástrico disponíveis em bancos de dados públicos (TCGA, ICGC e cbioPortal) e do ACCCC a fim de avaliar o valor prognóstico das assinatura mutacionais associadas com a falha nos mecanismos de reparo e ao fenótipo MSI (detalhes da metodologia e coortes selecionadas descritos no trabalho publicado Buttura et al. (2021), Apêndice A).

Ao total, foram detectadas (*de novo*) 7 assinaturas mutacionais, sendo três relacionadas à dMMR. Das sete assinaturas mutacionais encontradas em câncer gástrico, três tiveram associação com melhor prognóstico considerando a sobrevida global na análise de modelo simples de Cox: S1 associada ao processo de idade, e S4 e S5 associadas à falha no mecanismo de MMR (Figura 36; Apêndice A). Dentre essas assinaturas assoaciadas à dMMR, a assinatura S4 foi a única associada ao melhor prognóstico no modelo múltiplo de Cox (HR=0,59; CI[95%]: 0,37–0,96; p=0,033) (Figura 37, Apêndice A).

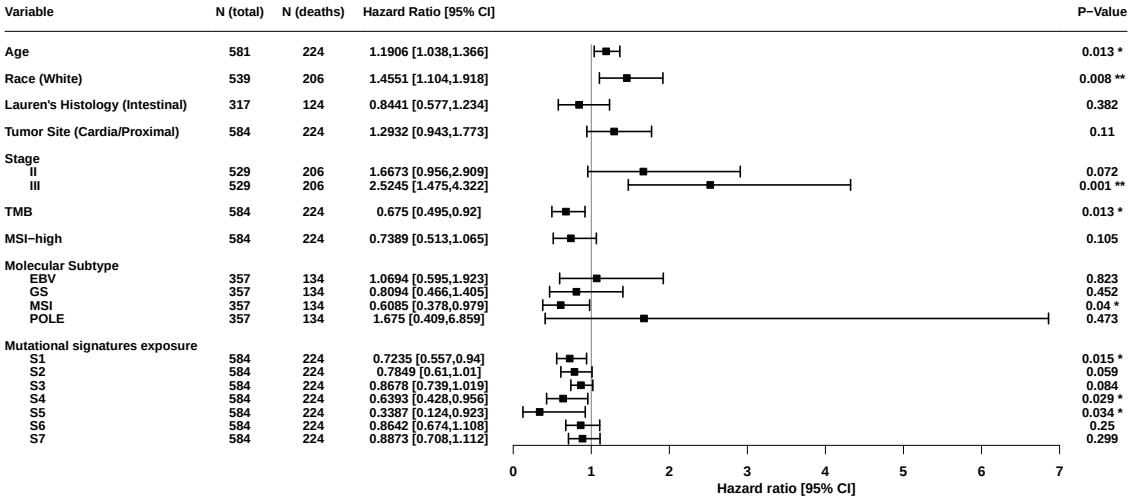


Figura 36 – Modelo de regressão simples de Cox para sobrevida global considerando variáveis clínicas.* p<0,05; ** p<0,01. (Buttura et al. 2021) (Apêndice A)

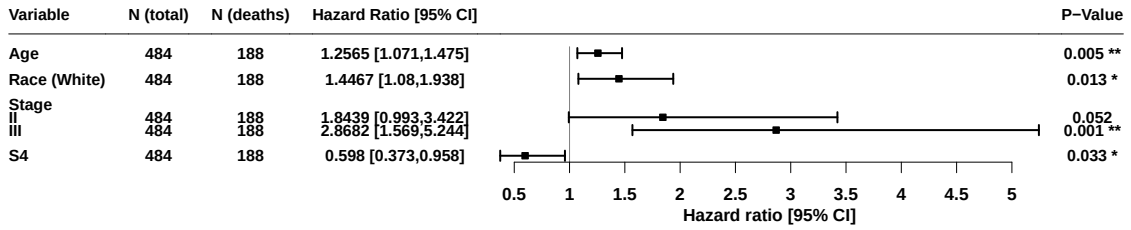


Figura 37 – Modelo de regressão múltiplo considerando as variáveis clínicas significativas e exposição a assinatura S4.* p<0,05; ** p<0,001. (Buttura et al. 2021) (Apêndice A).

Os pacientes foram então classificados em S4-high e S4-low de acordo com a exposição da assinatura S4. O ponto de corte para esta classificação foi o valor do terceiro quartil. Este ponto foi escolhido considerando que o melhor ponto de corte encontrado pela função "maxstat", que define a separação da curva de sobrevida de Kapplan-Meier, encontrava-se no terceiro quartil. Assim, pacientes com os 25% valores maiores de exposição da assinatura S4 foram classificados como S4-high e o restante como S4-low.

De importância, revelamos que essa classificação baseada na exposição da assinatura S4 encontra-se associada ao melhor prognóstico (Figura 38), agru-

pando pacientes com diferentes características clínicas e moleculares, como subtipo histológico de Lauren, local do tumor, subtipos moleculares, carga mutacional e heterogeneidade tumoral, sugerindo que esta classificação pode estratificar melhor os pacientes. A Figura 39 ilustra algumas associações clínicas e moleculares encontradas com esta estratificação (S4-high/S4-T e S4-low/S4-O).

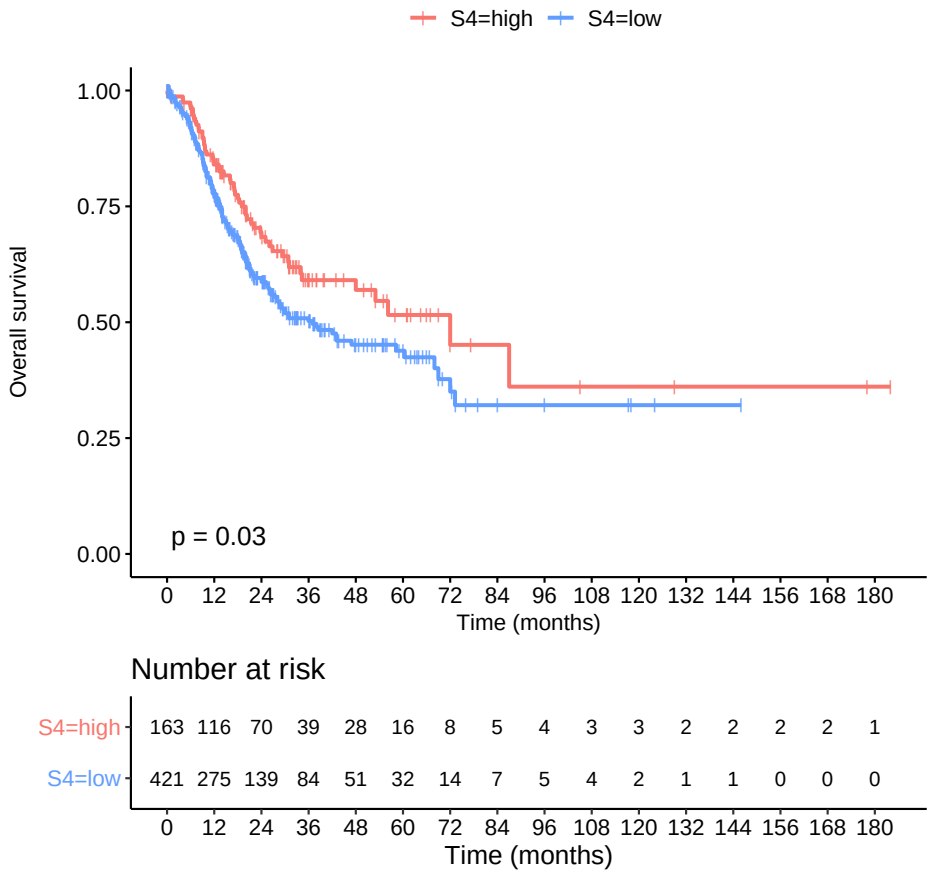


Figura 38 – Sobrevida global entre os grupos S4-high e S4-low (Buttura et al. 2021) Teste *log-rank*; p<0,05. (Apêndice A)

Foi possível observar que os pacientes S4-high possuem em sua maioria tumores do subtipo molecular MSI, intestinal, localizados na região mais distal, com estadiamentos mais precoces e carga mutacional (TMB) elevada (Figura 39; Apêndice A). Por outro lado, as características clínicas de subtipos difuso e GS, localização mais proximal e estadiamentos mais avançados foram associadas ao grupo de pacientes S4-low. Além disso, no grupo S4-low concentrou pacientes

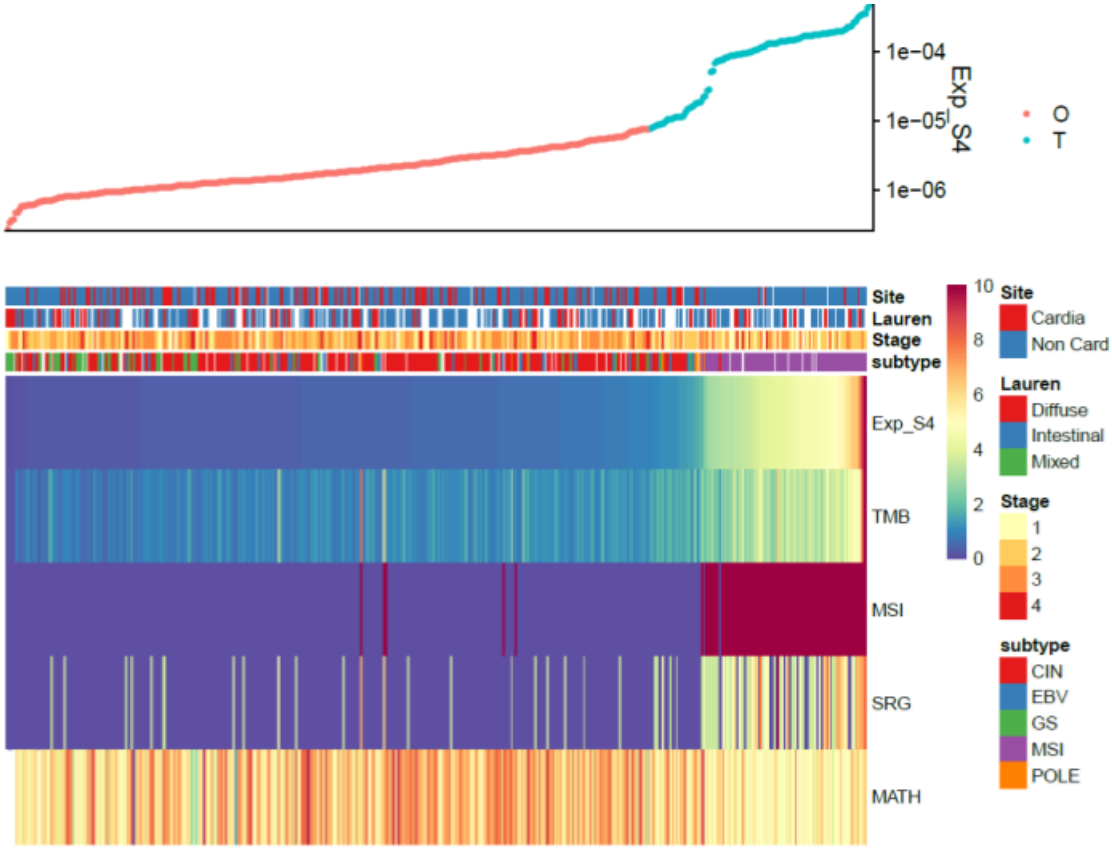
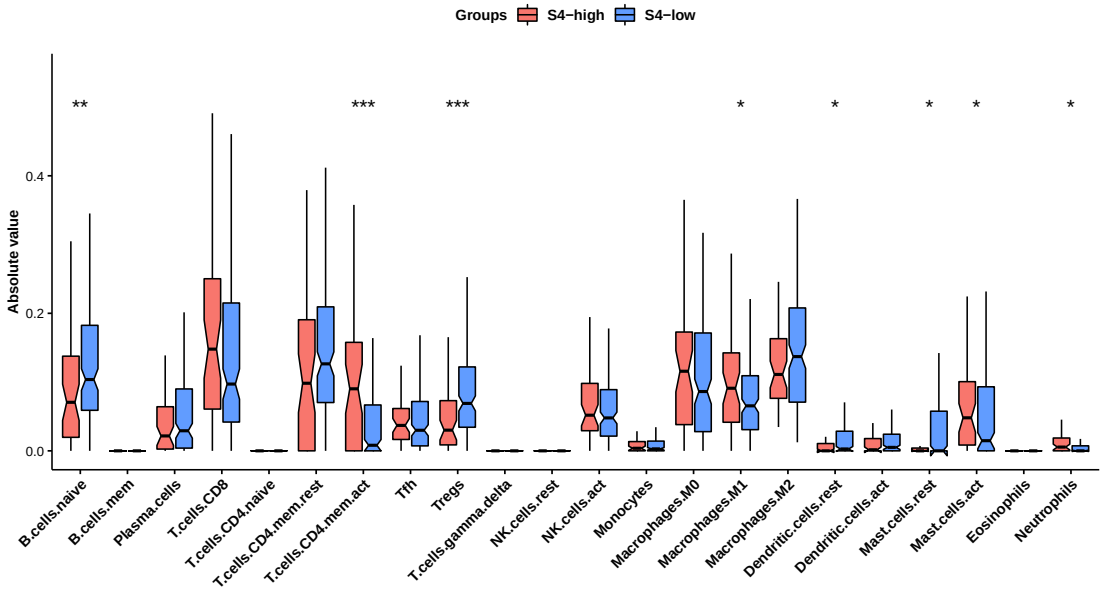


Figura 39 – *Heatmap* com valores normalizados para comparação entre os grupos S4-T e S4-O mostrando a exposição da assinatura mutational S4 (MMR), carga mutacional (TMB), status MSI, genes MMR com mutação somática (SRG) e heterogeneidade tumoral, bem como anotação de variáveis clínicas (sítio anatômico, classificação histológica de Lauren e subtipo molecular).

com tumores de carga mutacional (TMB) de baixa a intermediária, e tumores mais heterogêneos (indicado pelo MATH na Figura 39) do que no grupo S4-high.

A avaliação da composição do infiltrado inflamatório nas amostras do TCGA, também indicou um microambiente tumoral mais responsivo nos pacientes S4-high (Figura 40). As análises de composição do infiltrado para os 22 tipos celulares e resultados do *xCell* mostraram que o grupo S4-high apresentou significativamente maiores proporções de células citotóxicas e pró inflamatórias, enquanto o grupo S4-low apresentou maiores quantidades de células imaturas e imunorregulatórias (Figura 40 e Apêndice A). O aspecto de um microambiente tumoral mais responsivo nos pacientes S4-high foi reforçado pela correlação positiva do grupo S4-high com a expressão de genes efetores da resposta imune

citotóxica e *checkpoints* imunes, enquanto S4-low a correlação positiva foi com genes associados a resposta de células T regulatórias (Figura 41).



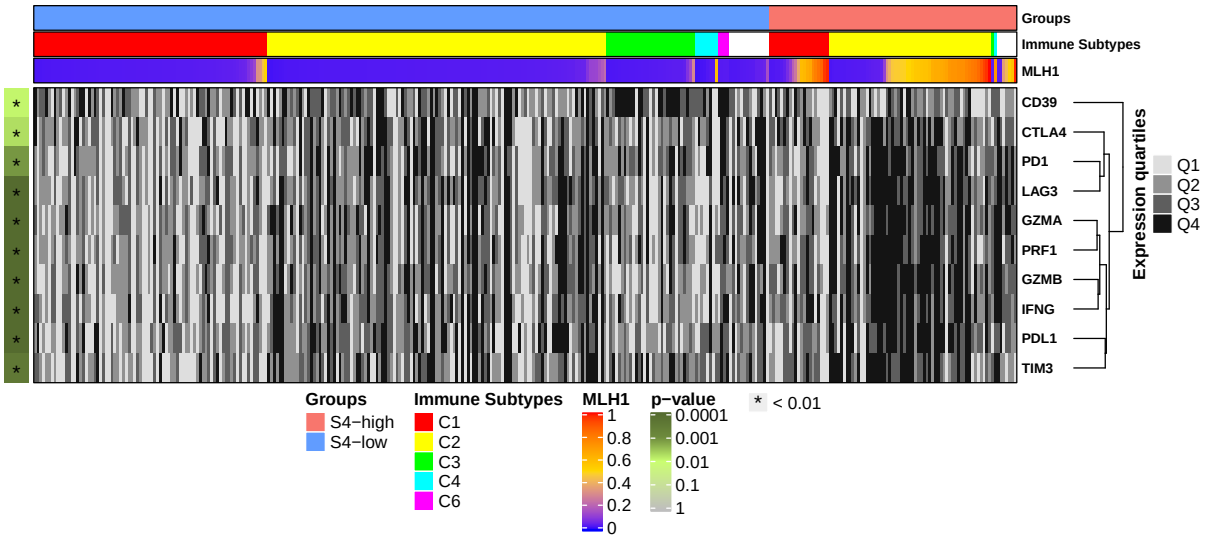
Fonte: Buttura et al. (2021)

Figura 40 – Comparação das quantidades de células imunes estimadas pelo método CIBERSORT entre os grupos S4-high e S4-low. * p<0,05; ** p<0,001; ***p<0,0001

A lista desses genes efetores utilizada para análise de correlações e avaliação da resposta imunológica efetora está descrita na Tabela 5. Este foi um compilado dos principais genes efetores e marcadores utilizados em citometria de fluxo e associados à imunoterapia.

Concluimos então, que o grupo S4-high foi associado com melhor prognóstico, sugerindo que a classificação baseada na exposição de uma assinatura mutacional conseguiu resgatar, independente das características clínicas e moleculares, os pacientes com melhor prognóstico (Apêndice A).

A fim de entender melhor os aspectos imunes em câncer gástrico, complementando os resultados encontrados e detalhando o possível valor prognóstico do contexto imunológico nestes pacientes, foi feito um novo modelo simples de Cox



Fonte: Buttura et al. (2021)

Figura 41 – Expressão gênica de genes marcadores de resposta imune com diferença significativa entre os grupos S4-high e S4-low. A expressão gênica foi classficada pelos quartis, sendo Q1 valores menores que o primeiro quartil, Q2, entre o primeiro quartil e a mediana, Q3, entre a mediana e terceio quartil e Q4, valores maiores que o terceiro quartil.

com a inclusão da composição de células e expressão de genes efetores (Figura 42). Avaliando primeiramente as características clínicas de AdG de um modo geral (coorte TCGA-STAD, n=439 amostras), pode-se observar a idade e estadiamento como variáveis clínicas importantes ao prognóstico (p<0,05). Além disso, a composição estromal do infiltrado tumoral (estimativas pelas ferramentas xCell (Aran et al. 2017) e ESTIMATE (Yoshihara et al. 2013), bem como o subtipo MSI apresentam uma tendência de relação com melhor prognóstico. Considerando a coorte maior (Buttura et al. 2021), o subtipo MSI também foi considerada como uma variável importante para o prognóstico, diminuindo o risco de óbito.

Em seguida, investigamos a associação prognóstica da expressão de genes da resposta imune. De uma lista de 85 genes relacionados à resposta imune (Tabela 5), 6 tiveram a expressão gênica associada significativamente ao prognóstico (Figura 43, p<0,05). A maior expressão de genes das citocinas *IL9* e *IL4*, rela-

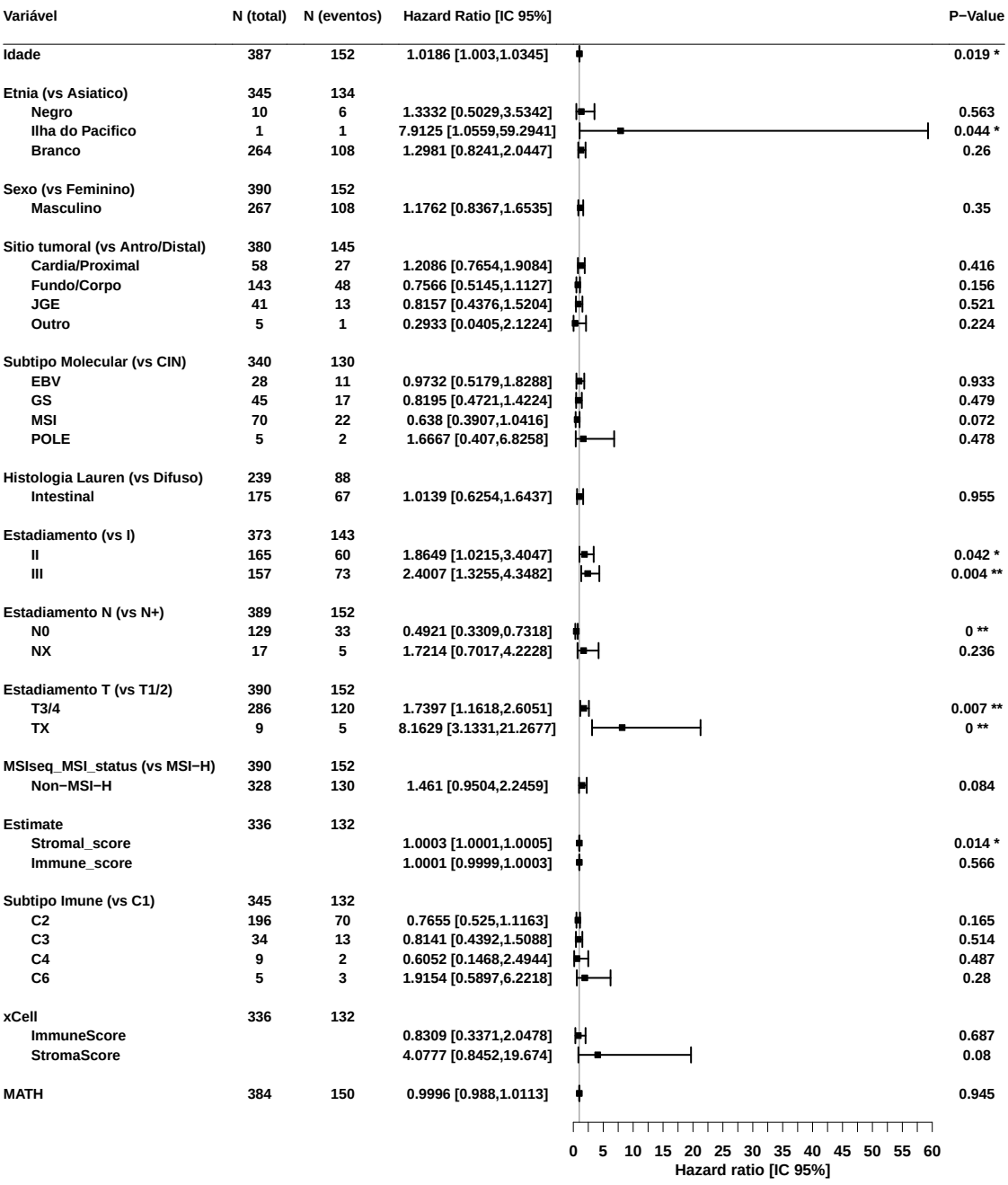


Figura 42 – Modelo simples de Cox para sobrevida global considerando as variáveis clínicas, moleculares e imunes nas amostras de AdG (TCGA-STAD).

Tabela 5 – Revisão de métodos de avaliação da composição de infiltrado inflamatório utilizando dados de expressão gênica.

Genes associados à resposta inflamatória/citotóxica:
IL1A,*IL1B*, *IL2*, *IL6*, *IL8*, *IL9*, *IL12A*, *IL12B*, *IL17A*, *IL17F*, *IL23A*, *TNF* (TNF-alfa), *IFNG* (IFN-gama), *GZMA*, *GZMB*, *PRF1*, *LAMP1* (CD107)
Genes associados à resposta supressora/exaustão:
IL4, *IL5*, *IL10* , *IL13*, *TGFB1* (TGF-beta), *PDCD1* (PD-1), *CD274* (PD-L1), *PDCD1LG2* (PD-L2), *CTLA4*, *LAG3*, *HAVCR2* (TIM3), *ITGAE* (CD103), *ENTPD1* (CD39), *FOXP3*
Genes associados à resposta coestimulatória/apresentação de antígenos:
CD40, *CD80*, *CD86*, *ICOS*, *B2M*, *TAP1*, *TAP2*, *HLA-A*, *HLA-B*, *HLA-C*, *HLA-E*, *HLA-G*, *HLA-F*, *HLA-DOA*, *HLA-DOB*, *HLA-DMA*, *HLA-DMB*, *HLA-DPA1*, *HLA-DPB1*, *HLA-DPB2*, *HLA-DQA1*, *HLA-DQA2*, *HLA-DQB1*, *HLA-DQB2*, *HLA-DRA*, *HLA-DRB1*, *HLA-DRB5*, *HLA-DRB6*
Genes marcadores básicos de células imunes:
CD19 (linfócitos B), *MS4A1* (CD20, linfócitos B), *IL2RA* (CD25, Treg), *FOXP3* (Treg), *CD3E* (linfócitos T), *CD3D* (linfócitos T), *CD4* (linfócitos T), *CD8A* (linfócitos T), *KLRD1* (CD94, células NK), *KLRC1* (NKG2A, células NK), *KLRC2* (NKG2C, células NK), *KLRK1* (NKG2D, células NK), *KIR2DL4* (células NK), *KIR2DL3* (células NK), *KIR3DL2* (células NK), *NCAM1* (CD56, células NK), *CXCR1* (neutrófilos), *CXCR2* (neutrófilos), *ITGAM* (CD11b, células mielóides), *ITGAX* (CD11c, células mielóides), *CD14* (células mielóides), *FCGR3A* (CD16, células mielóides), *CD1A* (células dendríticas), *CD163* (macrófagos), *CD68* (macrófagos), *MS4A6A* (macrófagos)

cionadas a respostas imunes do tipo Th2, foi associada ao maior risco de óbito, enquanto maior expressão de *PDCD1* (molécula PD1) foi associada a menor risco de óbito (Figura 43).

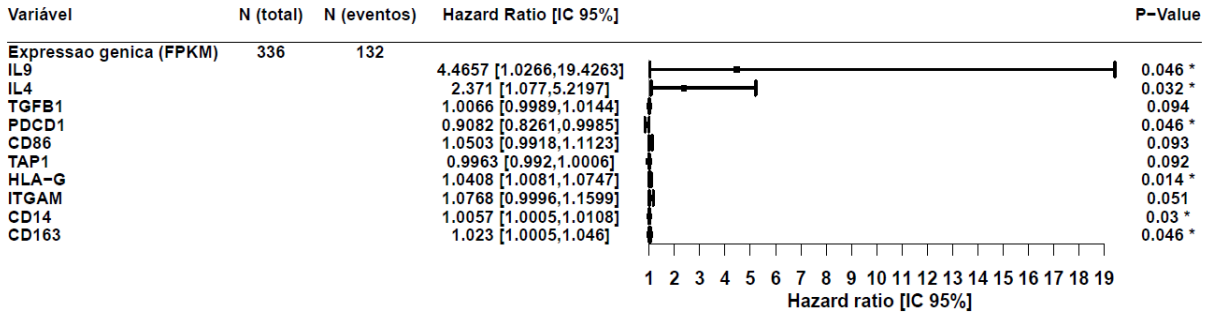


Figura 43 – Modelo simples de Cox para sobrevida global considerando expressão de genes efetores da resposta imune em amostras de AdG (TCGA-STAD).

Para avaliação prognóstica da composição das células imunes, foi determi-

nado um ponto de corte para cada tipo celular com a função *maxstat*, que define o ponto de corte associado ao menor valor de p na análise de sobrevida. Menores quantidades de células B naive, linfócitos T CD4 desativados (*resting*), monócitos, macrófagos (M1 e M2) e mastócitos ativados foram associadas significativamente ao melhor prognóstico (Figuras 44). Menores quantidades de macrófagos M1 e mastócitos ativados sugerem que o ambiente mais inflamatório de resposta imune do tipo inata teria pior prognóstico. Já maiores quantidades de células B de memória, plasmócitos, linfócitos T CD8+ e helper foliculares (Tfh) e células dendríticas ativadas tiveram associação (ou tendência) com melhor prognóstico, diminuindo o risco de óbito (Figura 44). Além disso, menores quantidades de linfócitos T reguladores (Treg) foram associados ao menor risco de óbito (Figura 44).

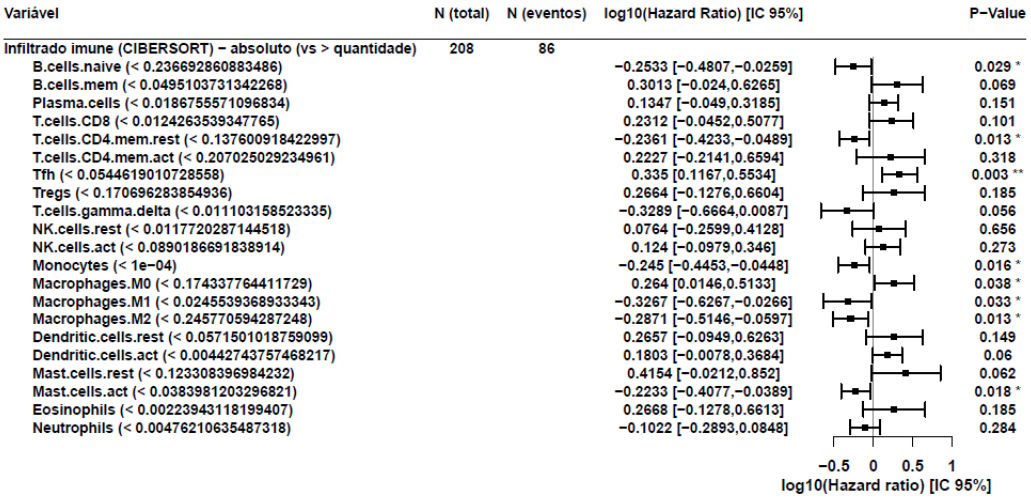


Figura 44 – Modelo simples de Cox para sobrevida global considerando quantidade estimada de células imunes em amostras de AdG (TCGA-STAD).

Comparada à análise com 7 tipos celulares básicos, a avaliação da composição do infiltrado inflamatório em AdG utilizando mais subtipos celulares relacionados com o aspecto funcional da resposta imune no microambiente tumoral é mais informativa e permitiu determinar algumas associações prognósticas significativas. Há indícios de que a formação de estruturas terciárias linfóides, com base na composição e relação prognóstica de linfócitos Tfh, células B ativadas e

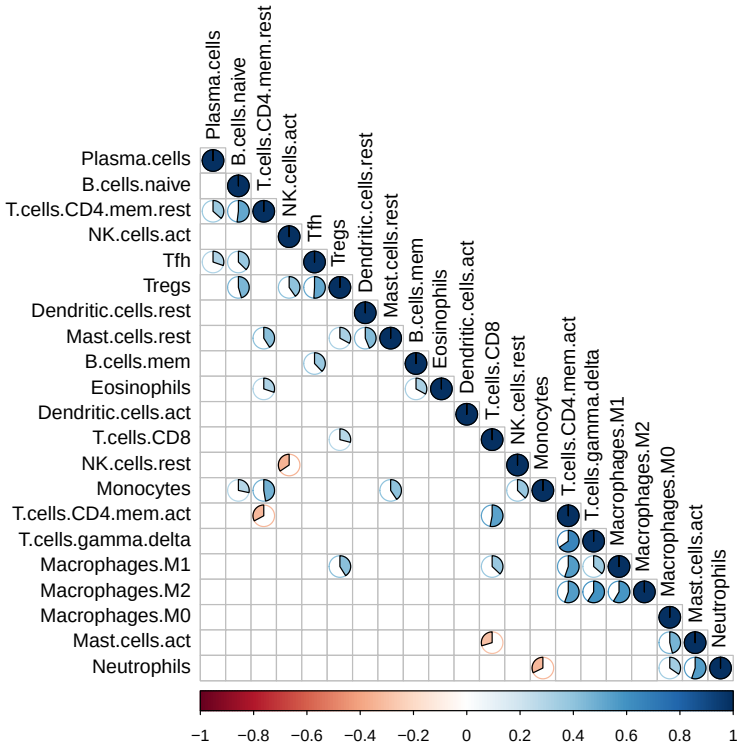
linfóctos T CD8+, é importante para o bom prognóstico em AdG. Por outro lado, células e expressão de genes relacionados à resposta inflamatória, como mastócitos, macrófagos M1 e genes de citocinas como *IL4* e *IL9*, foram assoaciados ao pior prognóstico.

Considerando a classificação dos pacientes em S4-high e S4-low, foi feita a correlação das células entre si e, em seguida, com os genes marcadores de resposta imune descritos na Tabela 5. A expressão dos genes foi normalizada com o valores de 1 - pureza do tumor estimados pelo método ESTIMATE (Yoshihara et al. 2013).

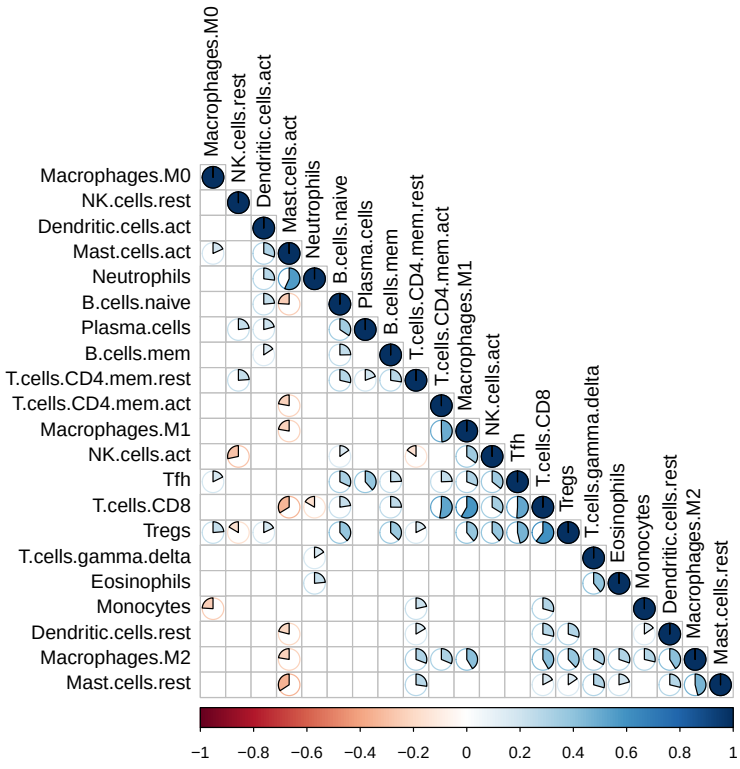
A correlação entre as células sugere dois perfis diferentes de resposta imune associadas aos grupos S4-high e S4-low (Figura 45). No primeiro, a ideia de formação de estruturas terciárias linfóides é reforçada pelas correlações positivas observadas entre os linfócitos B de meméria, Tfh e T CD8. Além disso, há tabém correlações positivas entre as célula citotóxicas, como linfócitos T gama-delta, e macrófagos do tipo M2.

Já o grupo S4-low mostrou um perfil com correlações positivas entre linfócitos Tregs e linfócitos T CD4 *resting*, caracterizando um mecanismo imunorregulatório diferente. Isso fortaleca as primeiras associações encontradas nas comparações das células individuais entre os grupos, de que o grupo S4-high possui um microambiente mais imuno efetor, com assoaciação entre os linfócitos B e T, enquanto que o grupo S4-low possui um microambiente caracterizado pela ação de linfócitos Treg (Figura 45).

A análise de correlação global das células e com a expressão de genes da resposta imune e de genes do repertório de receptores dos linfócitos T e B também mostrou diferentes padrões de correlações funcionais entre os dois grupos. De importância, no grupo S4-high observamos correlações positivas significativas restritas às células T CD8, T CD4 de memória ativadas, T gama delta e macrófagos com genes de *checkpoints* imunes, de moléculas efetoras citotóxicas (*IFNG*, *GZMA*, *GZMB*, *PRF1*), genes relacionados à apresentação de antígenos



(a) Grupo S4-high



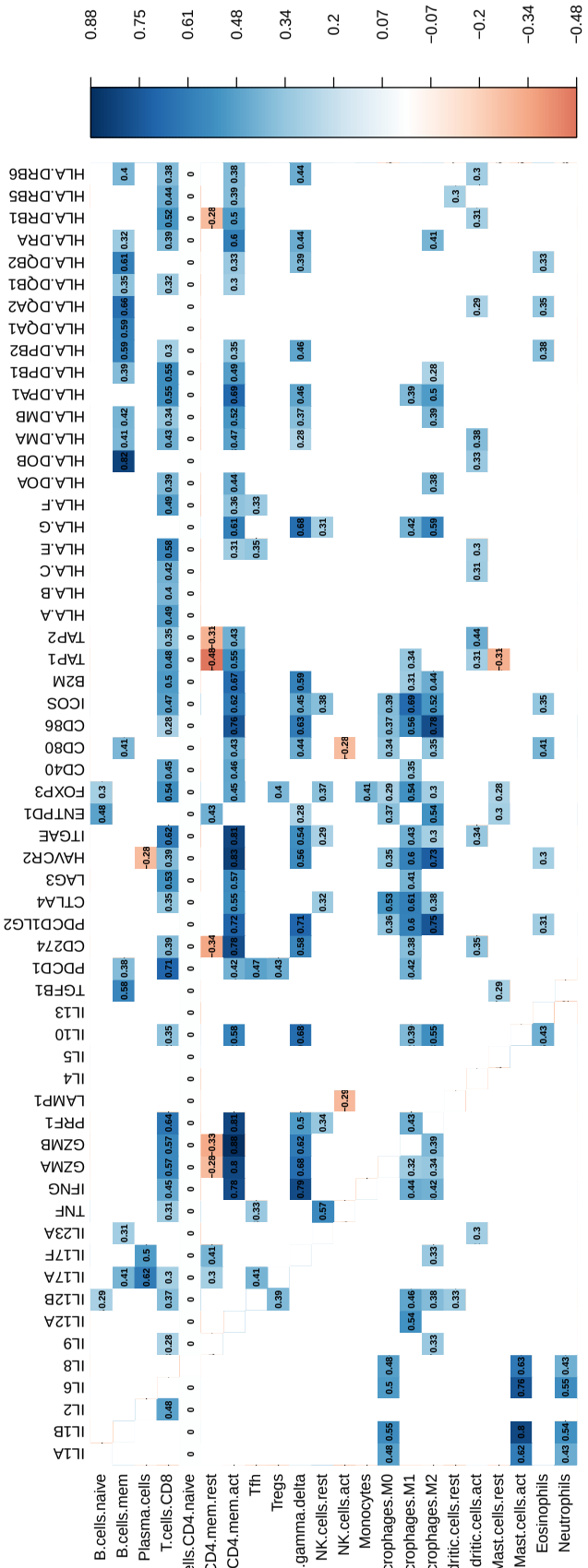
(b) Grupo S4-low

Figura 45 – Correlograma de Pearson entre as quantidades de células estimadas por grupo S4-high e S4-low.

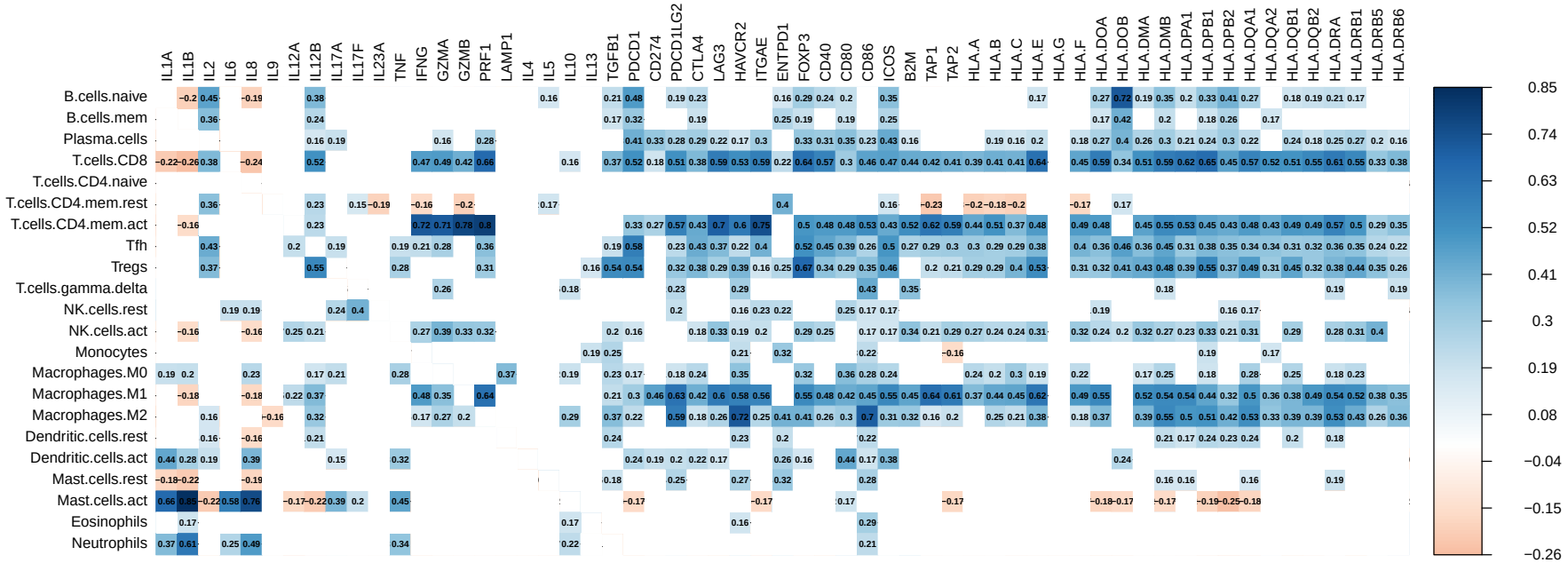
e de moléculas HLA de classe II (Figura 46).

O grupo S4-low, por sua vez, apresentou correlações positivas, de uma maneira generalizada entre várias células com os genes de *checkpoint* imunes e de moléculas HLA de classe I e II, com destaque para correlação menor de linfócitos T gama delta e correlações mais fortes para Tregs, se comparado com grupo S4-high (Figura 46).

Considerando os genes de repertório de receptores de linfócitos T e B, o que observamos foi uma confirmação das correlações positivas mais concentradas nas células efetoras no grupo S4-high (linfóictos B de memória, linfócitos T CD4 e T CD8, e macrófagos) enquanto que no grupo S4-low as correlações não se destacam entre os subtipos celulares. Além disso, ficou evidente uma resposta de oposta entre linfócitos e mastócitos no grupo S4-low (Figura 47) e uma resposta de linfócitos B muito mais forte no grupo S4-high (Figura 48).

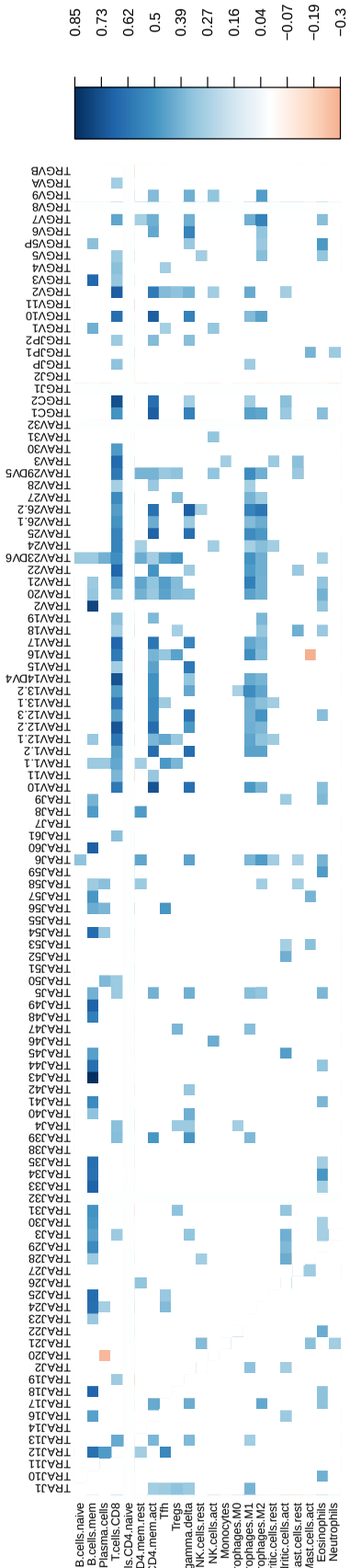


(a) Grupo S4-high

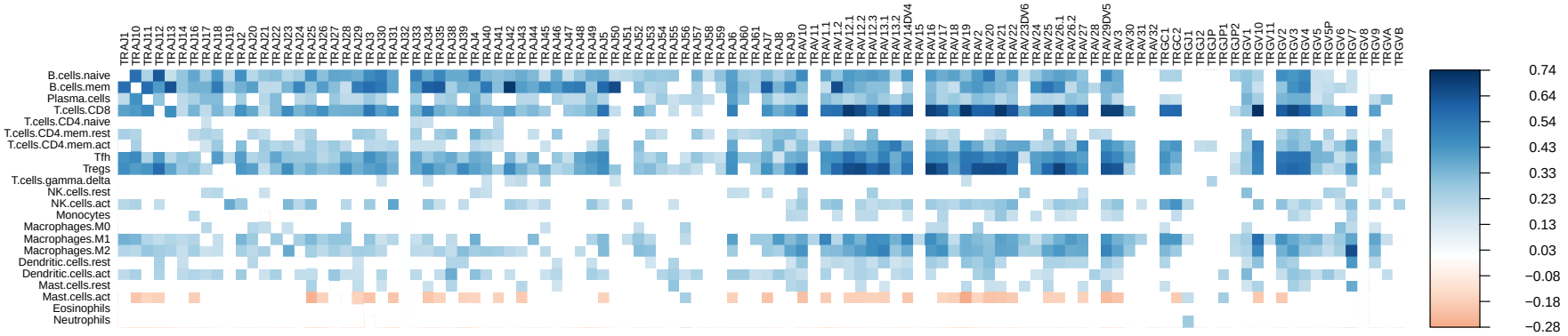


(b) Grupo S4-low

Figura 46 – Correlograma de Pearson da expressão de genes efetores da resposta imune por grupo S4-high e S4-low.

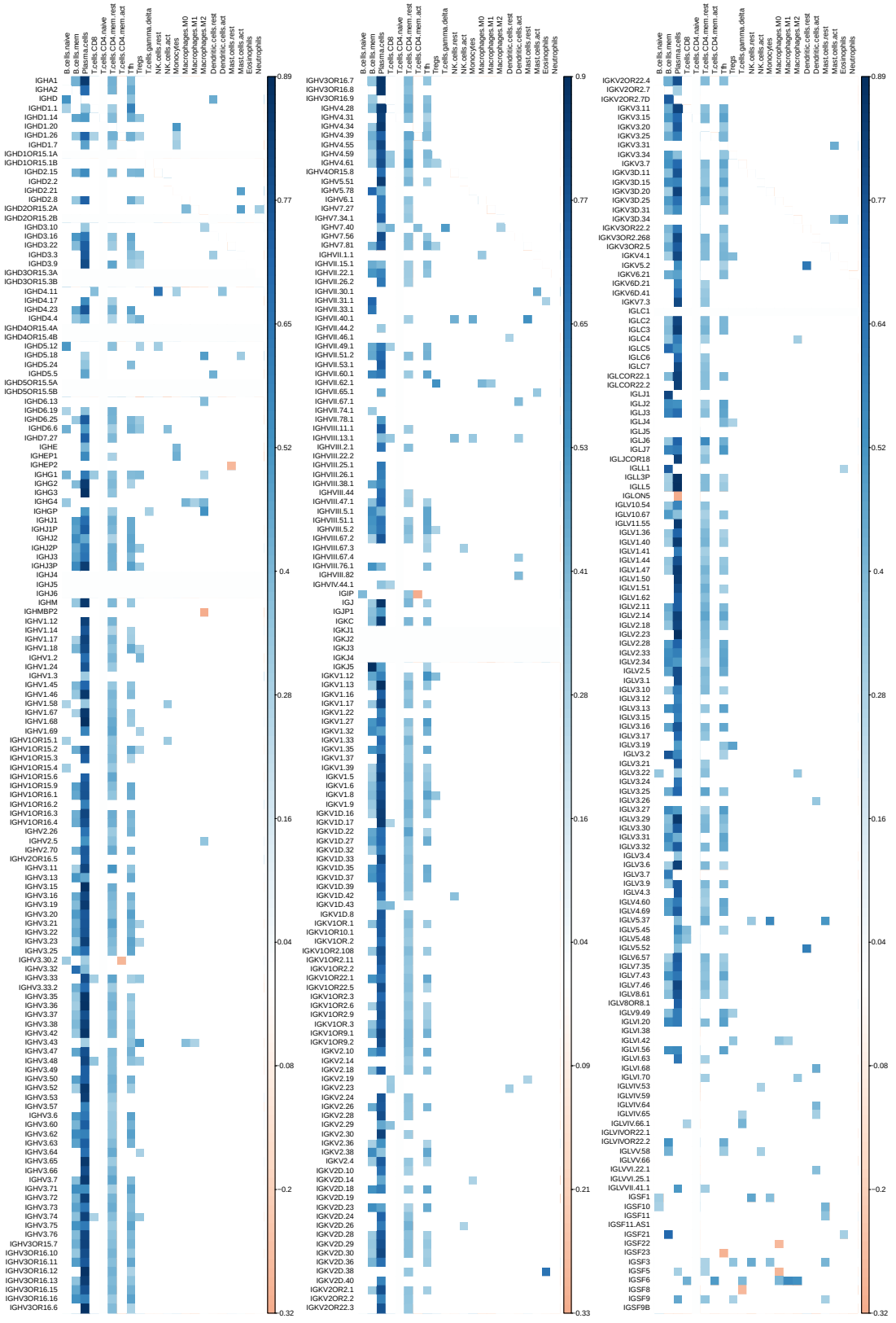


(a) Grupo S4-high.

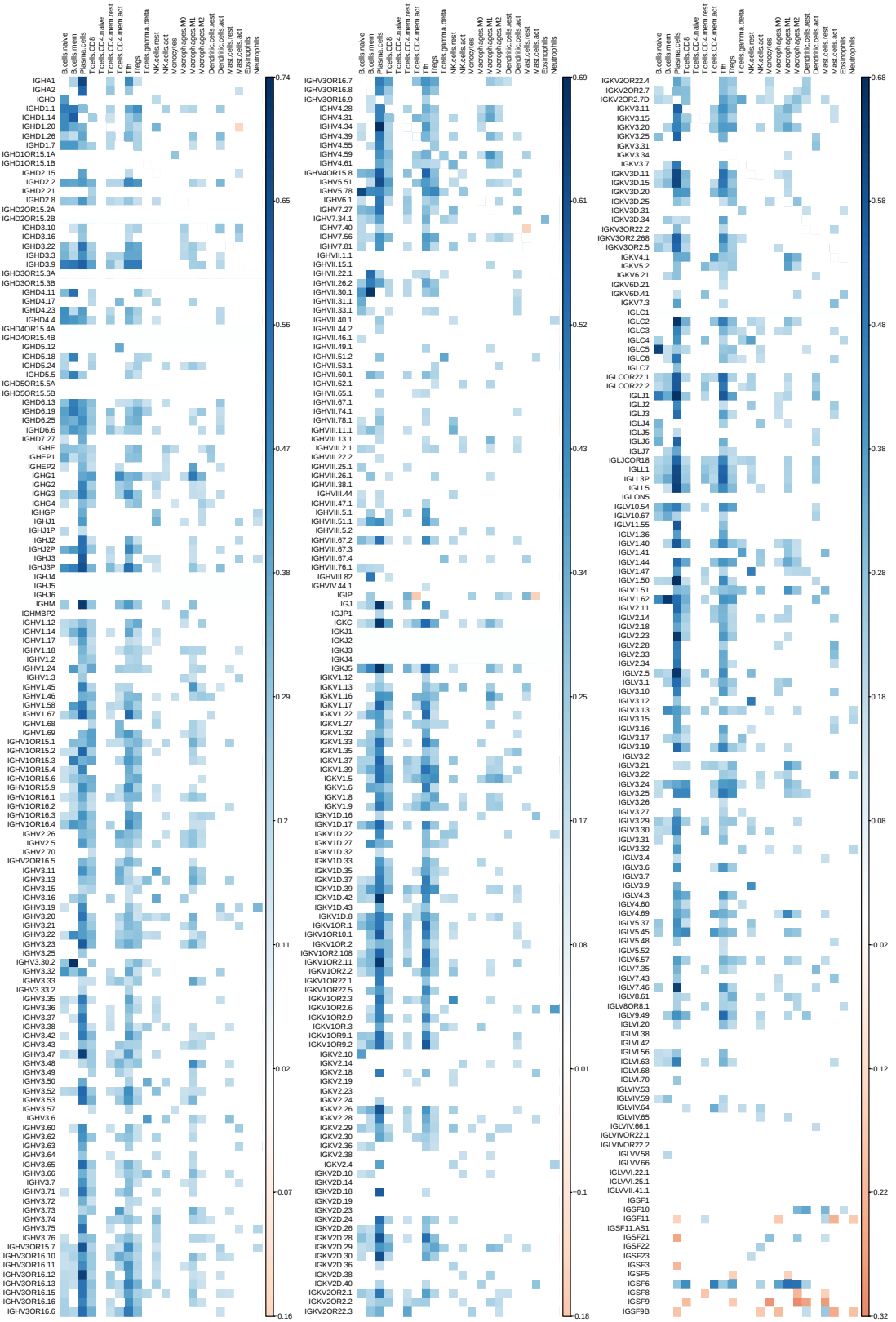


(b) Grupo S4-low.

Figura 47 – Correlograma de Pearson da expressão de genes do repertório de receptores de linfócitos T (TCR) por grupo S4-high e S4-low.



(a) Grupo S4-high.



(b) Grupo S4-low.

Figura 48 – Correlograma de Pearson da expressão de genes do repertório de receptores de células B (Ig) por grupo S4-high e S4-low.

6 CONCLUSÕES

Atualmente, esforços têm sido feitos para elucidar as interações que existem entre tumor e os componentes do sistema imunológico, principalmente a fim de descrever novos preditores prognósticos. Considerando o papel chave do sistema imunológico, por exemplo, em processos inflamatórios, produção e apresentação de citocinas e reconhecimento de antígenos, as células imunes no microambiente tumoral podem influenciar a resposta ao tratamento, seja convencional ou por imunoterapia, o que está diretamente relacionado ao desfecho clínico do paciente.

Este trabalho destaca a importância da avaliação global do microambiente tumoral, seus aspectos moleculares e composição de células, que, em câncer gástrico, se mostraram mais informativas quanto ao prognóstico dos pacientes. Nossos achados reforçam o valor prognóstico das assinaturas mutacionais na estratificação de pacientes com melhor prognóstico, independente das atuais características clínicas de subtipo histológico e sítio do tumor. Concluímos ainda que as características do microambiente tumoral nesses pacientes indicam qual tipo de resposta está ativa.

Nossos achados endossam a relevância da avaliação do contexto imunológico na prática clínica ao considerar seu potencial valor preditivo de resposta às terapias atuais. A avaliação global dos níveis de expressão gênica em amostras tumorais revelam aspectos funcionais e ativos no microambiente com grande pontencial de auxiliar na predição de melhores estratégias de tratamento, novos alvos terapêuticos e biomarcadores prognósticos.

7 REFERÊNCIAS

Abbas AR, Baldwin D, Ma Y, Ouyang W, Gurney A, Martin F, et al. Immune response in silico (IRIS): immune-specific genes identified from a compendium of microarray expression data. Genes & Immunity. 2005 jun; 6:319–331.

Abbas AR, Wolslegel K, Seshasayee D, Modrusan Z, Clark HF. Deconvolution of Blood Microarray Data Identifies Cellular Activation Patterns in Systemic Lupus Erythematosus. PLoS ONE. 2009 jul; 4:e6098.

Abida W, Cheng ML, Armenia J, Middha S, Autio KA, Vargas HA, et al. Analysis of the Prevalence of Microsatellite Instability in Prostate Cancer and Response to Immune Checkpoint Blockade. JAMA Oncology. 2019 apr; 5:471–478.

Alexandrov LB, Kim J, Haradhvala NJ, Huang MN, Tian Ng AW, Wu Y, et al. The repertoire of mutational signatures in human cancer. Nature 2020 578:7793. 2020 feb; 578:94–101.

Alexandrov LB, Nik-Zainal S, Siu HC, Leung SY, Stratton MR. A mutational signature in gastric cancer suggests therapeutic strategies. Nature Communications. 2015 oct; 6:.

Allantaz F, Cheng DT, Bergauer T, Ravindran P, Rossier MF, Ebeling M, et al. Expression Profiling of Human Immune Cell Subsets Identifies miRNA-mRNA Regulatory Relationships Correlated with Cell Type Specific Expression. PLoS ONE. 2012 jan; 7:e29979.

Anderson WF, Camargo MC, Fraumeni JF, Correa P, Rosenberg PS, Rabkin CS. Age-Specific Trends in Incidence of Noncardia Gastric Cancer in US Adults. JAMA. 2010 may; 303:1723.

Angell HK, Bruni D, Barrett JC, Herbst R, Galon J. The Immunoscore: Colon Cancer and Beyond. Clinical Cancer Research. 2020 jan; 26:332–339.

Angelova M, Charoentong P, Hackl H, Fischer ML, Snajder R, Krogsdam AM, et al. Characterization of the immunophenotypes and antigenomes of colorectal cancers reveals distinct tumor escape mechanisms and novel targets for immunotherapy. Genome Biology. 2015 dec; 16:64.

Aran D, Hu Z, Butte AJ. xCell: Digitally portraying the tissue cellular heterogeneity landscape. Genome Biology. 2017 nov; 18:.

Arzimanoglou II, Gilbert F, Barber HR. Microsatellite instability in human solid tumors. Cancer. 1998 May; 82:1808–20.

Bass AJ, Thorsson V, Shmulevich I, Reynolds SM, Miller M, Bernard B, et al. Comprehensive molecular characterization of gastric adenocarcinoma. Nature. 2014 sep; 513:202–209.

Becht E, Giraldo NA, Dieu-Nosjean MC, Sautès-Fridman C, Fridman WH. Cancer immune contexture and immunotherapy. Current Opinion in Immunology. 2016 apr; 39:7–13.

Benatti P, Gafà R, Barana D, Marino M, Scarselli A, Pedroni M, et al. Microsatellite instability and colorectal cancer prognosis. Clinical cancer research : an official journal of the American Association for Cancer Research. 2005 dec; 11: 8332–8340.

Bindea G, Mlecnik B, Tosolini M, Kirilovsky A, Waldner M, Obenauf A, et al. Spatiotemporal Dynamics of Intratumoral Immune Cells Reveal the Immune Landscape in Human Cancer. Immunity. 2013 oct; 39:782–795.

Bobrovnitchaia I, Valieris R, Drummond RD, Lima JP, Freitas HC, Bartelli TF, et al. APOBEC-mediated DNA alterations: A possible new mechanism of carcinogenesis in EBV-positive gastric cancer. International Journal of Cancer. 2020 jan; 146:181–191.

Boland C, Goel A. Microsatellite instability in colorectal cancer. Gastroenterology. 2010; 138:.

Boland CR, Thibodeau SN, Hamilton SR, Sidransky D, Eshleman JR, Burt RW, et al. A National Cancer Institute Workshop on Microsatellite Instability for Cancer Detection and Familial Predisposition: Development of International Criteria for the Determination of Microsatellite Instability in Colorectal Cancer. Cancer Research. 1998; 58:.

Bonnal RJ, Ranzani V, Arrigoni A, Curti S, Panzeri I, Gruarin P, et al. De novo transcriptome profiling of highly purified human lymphocytes primary cells. Scientific Data 2015 2:1. 2015 sep; 2:1–10.

Bray F, Ferlay J, Soerjomataram I, Siegel RL, Torre LA, Jemal A. Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. CA: A Cancer Journal for Clinicians. 2018 nov; 68: 394–424.

Bruni D, Angell HK, Galon J. The immune contexture and Immunoscore in cancer prognosis and therapeutic efficacy. Nature Reviews Cancer 2020 20:11. 2020 aug; 20:662–680.

Buckowitz A, Knaebel H, Benner A, Bläker H, Gebert J, Kienle P, et al. Microsatellite instability in colorectal cancer is associated with local lymphocyte infiltration and low frequency of distant metastases. British journal of cancer. 2005 may; 92:1746–1753.

Buttura J, Provisor Santos M, Valieris R, Drummond R, Defelicibus A, Lima J, et al. Mutational Signatures Driven by Epigenetic Determinants Enable the Stratification of Patients with Gastric Cancer for Therapeutic Intervention. Cancers. 2021 feb; 13:1–21.

Cannataro VL, Gaffney SG, Sasaki T, Issaeva N, Grewal NKS, Grandis JR, et al. APOBEC-induced mutations and their cancer effect size in head and neck squamous cell carcinoma. Oncogene 2019 38:18. 2019 jan; 38:3475–3487.

Cesano A. nCounter(®) PanCancer Immune Profiling Panel (NanoString Technologies, Inc., Seattle, WA). Journal for immunotherapy of cancer. 2015; 3:42.

Chao J. *Stomach Cancer Immunotherapy - Cancer Research Institute (CRI)*. Disponível em: https://www.cancerresearch.org/en-us/immunotherapy/cancer-types/stomach-cancer .

Charoentong P, Finotello F, Angelova M, Mayer C, Efremova M, Rieder D, et al. Pan-cancer Immunogenomic Analyses Reveal Genotype-Immunophenotype Relationships and Predictors of Response to Checkpoint Blockade. Cell reports. 2017 jan; 18:248–262.

Chen YC, Fang WL, Wang RF, Liu CA, Yang MH, Lo SS, et al. Clinicopathological Variation of Lauren Classification in Gastric Cancer. Pathology and Oncology Research. 2016 oct; 22:197–202.

Ciriello G, Miller ML, Aksoy BA, Senbabaoglu Y, Schultz N, Sander C. Emerging landscape of oncogenic signatures across human cancers. Nature Genetics 2013 45:10. 2013 sep; 45:1127–1133.

Craven KE, Gökmen-Polar Y, Badve SS. CIBERSORT analysis of TCGA and METABRIC identifies subgroups with better outcomes in triple negative breast cancer. Scientific Reports 2021 11:1. 2021 feb; 11:1–19.

Cristescu R, Lee J, Nebozhyn M, Kim KM, Ting JC, Wong SS, et al. Molecular analysis of gastric cancer identifies subtypes associated with distinct clinical outcomes. Nature Medicine. 2015 apr; 21:449–456.

Danaher P, Warren S, Dennis L, D'Amico L, White A, Disis ML, et al. Gene expression markers of Tumor Infiltrating Leukocytes. Journal for ImmunoTherapy of Cancer. 2017 dec; 5:18.

Davies H, Glodzik D, Morganella S, Yates LR, Staaf J, Zou X, et al. HRDetect is a predictor of BRCA1 and BRCA2 deficiency based on mutational signatures. Nature Medicine. 2017 apr; 23:517–525.

Fang W, Chang S, Lan Y, Huang K, Chen J, Lo S, et al. Microsatellite instability is associated with a better prognosis for gastric cancer patients after curative surgery. World journal of surgery. 2012 sep; 36:2131–2138.

FDA. *FDA Approves First Immunotherapy for Initial Treatment of Gastric Cancer / FDA*. Disponível em: https://www.fda.gov/news-events/press-announcements/fda-approves-first-immunotherapy-initial-treatment-gastric-cancer .

FDA. *FDA Approves First-Line Immunotherapy for Patients with MSI-H/dMMR Metastatic Colorectal Cancer | FDA*. Disponível em: https://www.fda.gov/news-events/press-announcements/fda-approves-first-line-immunotherapy-patients-msi-hdmmr-metastatic-colorectal-cancer .

Ferlay J, Shin HR, Bray F, Forman D, Mathers C, Parkin DM. Estimates of worldwide burden of cancer in 2008: GLOBOCAN 2008. International Journal of Cancer. 2010 dec; 127:2893–2917.

Fernández EA, Mahmoud YD, Veigas F, Rocha D, Balzarini M, Lujan HD, et al. MIXTURE: an improved algorithm for immune tumor microenvironment estimation based on gene expression data. bioRxiv. 2019 aug;:726562.

Finotello F, Rieder D, Hackl H, Trajanoski Z. Next-generation computational tools for interrogating cancer immunity. Nature reviews Genetics. 2019 dec; 20: 724–746.

Fridman WH, Pagès F, Sautès-Fridman C, Galon J. The immune contexture in human tumours: impact on clinical outcome. Nature Reviews Cancer. 2012 mar; 12:298–306.

Furgason JM, Bahassi EM. Targeting DNA repair mechanisms in cancer. Pharmacology & Therapeutics. 2013 mar; 137:298–308.

Gajewski TF, Schreiber H, Fu YX. Innate and adaptive immune cells in the tumor microenvironment. Nature Immunology. 2013 sep; 14:1014–1022.

Galon J, Costes A, Sanchez-Cabo F, Kirilovsky A, Mlecnik B, Lagorce-Pagès C, et al. Type, Density, and Location of Immune Cells Within Human Colorectal Tumors Predict Clinical Outcome. Science. 2006 sep; 313:1960–1964.

Galon J, Lanzi A. Immunoscore and its introduction in clinical practice. The quarterly journal of nuclear medicine and molecular imaging : official publication of the Italian Association of Nuclear Medicine (AIMN) [and] the International Association of Radiopharmacology (IAR), [and] Section of the Society of. 2020 jun; 64:152–161.

Galon J, Pagès F, Marincola FM, Angell HK, Thurin M, Lugli A, et al. Cancer classification using the Immunoscore: a worldwide task force. Journal of Translational Medicine. 2012 oct; 10:205.

Gaujoux R, Seoighe C. Semi-supervised Nonnegative Matrix Factorization for gene expression deconvolution: a case study. Infection, genetics and evolution : journal of molecular epidemiology and evolutionary genetics in infectious diseases. 2012 jul; 12:913–921.

Gaujoux R, Seoighe C. CellMix: a comprehensive toolbox for gene expression deconvolution. Bioinformatics. 2013 sep; 29:2211–2212.

Giraldo NA, Becht E, Vano Y, Sautès-Fridman C, Fridman WH. The immune response in cancer: from immunology to pathology to immunotherapy. Virchows Archiv. 2015 aug; 467:127–135.

Gong T, Hartmann N, Kohane IS, Brinkmann V, Staedtler F, Letzkus M, et al. Optimal Deconvolution of Transcriptional Profiling Data Using Quadratic Programming with Application to Complex Clinical Blood Samples. PLoS ONE. 2011 nov; 6:e27156.

Greenman C, Stephens P, Smith R, Dalgliesh GL, Hunter C, Bignell G, et al. Patterns of somatic mutation in human cancer genomes. Nature. 2007 mar; 446: 153.

Hanahan D, Weinberg R. The hallmarks of cancer. Cell. 2000 jan; 100:57–70.

Hanahan D, Weinberg R. Hallmarks of Cancer: The Next Generation. Cell. 2011 mar; 144:646–674.

Helleday T, Eshtad S, Nik-Zainal S. *Mechanisms underlying mutational signatures in human cancers*. [S.l.]: Nature Publishing Group, 2014. 585–598 .

Hugo W, Zaretsky JM, Sun L, Song C, Moreno BH, Hu-Lieskovan S, et al. Genomic and Transcriptomic Features of Response to Anti-PD-1 Therapy in Metastatic Melanoma. Cell. 2016 mar; 165:35.

Hunt G, Freytag S, Bahlo M, Gagnon-Bartsch J. dtangle: accurate and robust cell type deconvolution. Bioinformatics (Oxford, England). 2019 jun; 35:2093–2099.

Jager M, Blokzijl F, Kuijk E, Bertl J, Vougioukalaki M, Janssen R, et al. Deficiency of nucleotide excision repair is associated with mutational signature observed in cancer. Genome Research. 2019; 29:1067–1077.

Jiménez-Sánchez A, Cast O, Miller ML. Comprehensive Benchmarking and Integration of Tumor Microenvironment Cell Estimation Methods. Cancer Research. 2019 dec; 79:6238–6246.

Kakiuchi M, Nishizawa T, Ueda H, Gotoh K, Tanaka A, Hayashi A, et al. Recurrent gain-of-function mutations of RHOA in diffuse-type gastric carcinoma. Nature Genetics 2014 46:6. 2014 may; 46:583–587.

Kim J, Mouw KW, Polak P, Braunstein LZ, Kamburov A, Tiao G, et al. Somatic ERCC2 mutations are associated with a distinct genomic signature in urothelial tumors. Nature Genetics. 2016 jun; 48:600–606.

Le DT, Durham JN, Smith KN, Wang H, Bartlett BR, Aulakh LK, et al. Mismatch repair deficiency predicts response of solid tumors to PD-1 blockade. Science. 2017 jul; 357:409–413.

Leung W, Kim J, Kim J, Graham D, Sepulveda A. Microsatellite instability in gastric intestinal metaplasia in patients with and without gastric cancer. The American journal of pathology. 2000; 156:537–543.

Li B, Severson E, Pignon JC, Zhao H, Li T, Novak J, et al. Comprehensive analyses of tumor immunity: implications for cancer immunotherapy. Genome Biology. 2016 dec; 17:174.

Li T, Fan J, Wang B, Traugh N, Chen Q, Liu JS, et al. TIMER: A Web Server for Comprehensive Analysis of Tumor-Infiltrating Immune Cells. Cancer research. 2017 nov; 77:e108–e110.

Liu R, Peng K, Yu Y, Liang L, Xu X, Li W, et al. Prognostic Value of Immunoscore and PD-L1 Expression in Metastatic Colorectal Cancer Patients with Different RAS Status after Palliative Operation. BioMed Research International. 2018; 2018:.

Lordick F, Janjigian YY. Clinical impact of tumour biology in the management of gastroesophageal cancer. Nature Reviews Clinical Oncology. 2016 mar; 13: 348–360.

Ma J, Setton J, Lee NY, Riaz N, Powell SN. The therapeutic significance of mutational signatures from DNA repair deficiency in cancer. Nature Communications 2018 9:1. 2018 aug; 9:1–12.

Mabbott NA, Baillie J, Brown H, Freeman TC, Hume DA. An expression atlas of human primary cells: inference of gene function from coexpression networks. BMC Genomics. 2013 sep; 14:632.

Maby P, Bindea G, Mlecnik B, Galon J. License to kill: microsatellite instability and immune contexture. OncoImmunology. 2021; 10:1905935.

McLean MH, El-Omar EM. Genetics of gastric cancer. Nature Reviews Gastroenterology & Hepatology. 2014 aug; 11:664–674.

Ministério da Saúde, Instituto Nacional de Câncer José Alencar Gomes da Silva (INCA). *Estimativa/2016: Incidência de Câncer no Brasil*. Rio de Janeiro: INCA, 2015. ISBN 978-85-7318-283-5. Disponível em: http://www.inca.gov.br/estimativa/2016/estimativa-2016-v11.pdf.

Ministério da Saúde, Instituto Nacional de Câncer José Alencar Gomes da Silva (INCA). *Estimativa 2018 - Incidência de Câncer no Brasil*. [s.n.], 2017. ISBN 978-85-7318-284-2. Disponível em: http://www1.inca.gov.br/estimativa/2018/estimativa-2018.pdf.

Ministério da Saúde, Instituto Nacional de Câncer José Alencar Gomes da Silva (INCA). *Estimativa 2020: incidência de câncer no Brasil*. [s.n.], 2019. Disponível em: https://www.inca.gov.br/publicacoes/livros/estimativa-2020-incidencia-de-cancer-no-brasil.

Monaco G, Lee B, Xu W, Mustafah S, Hwang Y, Carré C, et al. RNA-Seq Signatures Normalized by mRNA Abundance Allow Absolute Deconvolution of Human Immune Cell Types. Cell reports. 2019 feb; 26:1627–1640.e7.

Newman AM, Liu CL, Green MR, Gentles AJ, Feng W, Xu Y, et al. Robust enumeration of cell subsets from tissue expression profiles. Nature Methods. 2015 apr; 12:453–457.

Newman AM, Steen CB, Liu CL, Gentles AJ, Chaudhuri AA, Scherer F, et al. Determining cell type abundance and expression from bulk tissues with digital cytometry. Nature Biotechnology 2019 37:7. 2019 may; 37:773–782.

Nik-Zainal S, Alexandrov LB, Wedge DC, Loo PV, Greenman CD, Raine K, et al. Mutational Processes Molding the Genomes of 21 Breast Cancers. 2012;:.

Novershtern N, Subramanian A, Lawton LN, Mak RH, Haining WN, McConkey ME, et al. Densely interconnected transcriptional circuits control cell states in human hematopoiesis. Cell. 2011 jan; 144:296–309.

Paglia LL, Laugé A, Weber J, Champ J, Cavaciuti E, Russo A, et al. ATM germline mutations in women with familial breast cancer and a relative with haematological malignancy. Breast Cancer Research and Treatment 2009 119:2. 2009 apr; 119:443–452.

Panda A, Mehnert JM, Hirshfield KM, Riedlinger G, Damare S, Saunders T, et al. Immune Activation and Benefit From Avelumab in EBV-Positive Gastric Cancer. JNCI: Journal of the National Cancer Institute. 2018 mar; 110:316–320.

Parkin DM, Bray F, Ferlay J, Pisani P. Global cancer statistics, 2002. CA: a cancer journal for clinicians. 2005; 55:74–108.

PDQ Adult Treatment Editorial Board. *Gastric Cancer Treatment (PDQ®): Health Professional Version*. 2002. Disponível em: http://www.ncbi.nlm.nih.gov/pubmed/26389209 .

Racle J, de Jonge K, Baumgaertner P, Speiser DE, Gfeller D. Simultaneous enumeration of cancer and immune cell types from bulk tumor gene expression data. eLife. 2017 nov; 6:.

Reichert ZR, Urrutia J, Alumkal JJ. *Microsatellite Instability as an Emerging Biomarker for Checkpoint Inhibitor Response in Advanced Prostate Cancer*. [S.l.]: American Medical Association, 2019. 478–479 .

Revathidevi S, Murugan AK, Nakaoka H, Inoue I, Munirajan AK. APOBEC: A molecular driver in cervical cancer pathogenesis. Cancer Letters. 2021 jan; 496: 104.

Saltz J, Gupta R, Hou L, Kurc T, Singh P, Nguyen V, et al. Spatial Organization and Molecular Correlation of Tumor-Infiltrating Lymphocytes Using Deep Learning on Pathology Images. Cell Reports. 2018 apr; 23:181–193.e7.

Samstein RM, Lee CH, Shoushtari AN, Hellmann MD, Shen R, Janjigian YY, et al. Tumor mutational load predicts survival after immunotherapy across multiple cancer types. Nat Genet. 2019 fev.; 51:202–206.

Sehgal R, Sheahan K, O'Connell PR, Hanly AM, Martin ST, Winter DC. Lynch Syndrome: An Updated Review. Genes. 2014; 5:497.

Shen-Orr SS, Gaujoux R. Computational deconvolution: extracting cell type-specific information from heterogeneous samples. Current Opinion in Immunology. 2013 oct; 25:571–578.

Smid M, Rodríguez-González FG, Sieuwerts AM, Salgado R, Prager-Van der Smissen WJC, van der Vlugt-Daane M, et al. Breast cancer genome and transcriptome integration implicates specific mutational signatures with immune cell infiltration. Nature Communications 2016 7:1. 2016 sep; 7:1–9.

Sohn B, Hwang J, Jang H, Lee H, Oh S, Shim J, et al. Clinical Significance of Four Molecular Subtypes of Gastric Cancer Identified by The Cancer Genome Atlas Project. Clinical cancer research : an official journal of the American Association for Cancer Research. 2017 aug; 23:4441–4449.

Staff A. *First Immunotherapy for Initial Treatment of Gastric Cancer Approved*. Disponível em: https://www.ajmc.com/view/first-immunotherapy-for-initial-treatment-of-gastric-cancer-approved .

Stockmann C, Schadendorf D, Klose R, Helfrich I. The Impact of the Immune System on Tumor: Angiogenesis and Vascular Remodeling. Frontiers in Oncology. 2014 apr; 4:69.

Stratton MR, Campbell PJ, Futreal PA. *The cancer genome*. [S.l.]: Nature Publishing Group, 2009. 719–724 .

Sturm G, Finotello F, List M. Immunedeconv: An R Package for Unified Access to Computational Methods for Estimating Immune Cell Fractions from Bulk RNA-Sequencing Data. Methods in Molecular Biology. 2020; 2120:223–232.

Sugio T, Miyawaki K, Kato K, Sasaki K, Yamada K, Iqbal J, et al. Microenvironmental immune cell signatures dictate clinical outcomes for PTCL-NOS. Blood advances. 2018 sep; 2:2242–2252.

Tang H, Qiao J, Fu YX. Immunotherapy and tumor microenvironment. Cancer Letters. 2016 jan; 370:85–90.

Thompson D, Duedal S, Kirner J, McGuffog L, Last J, Reiman A, et al. Cancer Risks and Mortality in Heterozygous ATM Mutation Carriers. JNCI: Journal of the National Cancer Institute. 2005 jun; 97:813–822.

Thorsson V, Gibbs DL, Brown SD, Wolf D, Bortone DS, Ou Yang TH, et al. The Immune Landscape of Cancer. Immunity. 2018 apr; 48:812–830.e14.

Torgovnick A, Schumacher B. DNA repair mechanisms in cancer development and therapy. Frontiers in Genetics. 2015; 0:157.

Torre LA, Bray F, Siegel RL, Ferlay J, Lortet-Tieulent J, Jemal A. Global cancer statistics, 2012. CA: A Cancer Journal for Clinicians. 2015 mar; 65:87–108.

Van Hoeck A, Tjoonk NH, Van Boxtel R, Cuppen E. *Portrait of a cancer: Mutational signature analyses for cancer diagnostics*. [S.l.]: BioMed Central Ltd., 2019. 1–14 .

Volkov NM, Yanus GA, Ivantsov AO, Moiseenko FV, Matorina OG, Bizin IV, et al. Efficacy of immune checkpoint blockade in MUTYH-associated hereditary colorectal cancer. Investigational New Drugs. 2019 aug;:.

Wadhwa R, Song S, Lee JS, Yao Y, Wei Q, Ajani JA. Gastric cancer—molecular and clinical dimensions. Nature Reviews Clinical Oncology. 2013 sep; 10:643–655.

Wang K, Kan J, Yuen ST, Shi ST, Chu KM, Law S, et al. Exome sequencing identifies frequent mutation of ARID1A in molecular subtypes of gastric cancer. Nature Genetics 2011 43:12. 2011 oct; 43:1219–1223.

Wang K, Yuen ST, Xu J, Lee SP, Yan HHN, Shi ST, et al. Whole-genome sequencing and comprehensive molecular profiling identify new driver mutations in gastric cancer. Nature Genetics 2014 46:6. 2014 may; 46:573–582.

Wang M, Busuttil RA, Pattison S, Neeson PJ, Boussioutas A. *Immunological battlefield in gastric cancer and role of immunotherapies*. [S.l.]: Baishideng Publishing Group Co., Limited, 2016. 6373–6384 .

Wang S, Jia M, He Z, Liu X. APOBEC3B and APOBEC mutational signature as potential predictive markers for immunotherapy response in non-small cell lung cancer. Oncogene. 2018 jul; 37:3924–3936.

Watkins NA, Gusnanto A, de Bono B, De S, Miranda-Saavedra D, Hardie DL, et al. A HaemAtlas: characterizing gene expression in differentiated human blood cells. Blood. 2009 may; 113:e1–e9.

Whiteside TL. The tumor microenvironment and its role in promoting tumor growth. Oncogene. 2008 oct; 27:5904–5912.

Yates LR, Campbell PJ. Evolution of the cancer genome. Nature reviews Genetics. 2012 nov; 13:795.

Ye L, Zhang T, Kang Z, Guo G, Sun Y, Lin K, et al. Tumor-infiltrating immune cells act as a marker for prognosis in colorectal cancer. Frontiers in Immunology. 2019; 10:.

Yoshihara K, Shahmoradgoli M, Martínez E, Vegesna R, Kim H, Torres-Garcia W, et al. Inferring tumour purity and stromal and immune cell admixture from expression data. Nature Communications. 2013 oct; 4:289–300.

Zaitsev K, Bambouskova M, Swain A, Artyomov MN. Complete deconvolution of cellular mixtures based on linearity of transcriptional signatures. Nature Communications 2019 10:1. 2019 may; 10:1–16.

Zhang P, Liu M, Cui Y, Zheng P, Liu Y. Microsatellite instability status differentially associates with intratumoral immune microenvironment in human cancers. Briefings in bioinformatics. 2021 may; 22:.

Zhang SC, Hu ZQ, Long JH, Zhu GM, Wang Y, Jia Y, et al. Clinical implications of tumor-infiltrating immune cells in breast cancer. J Cancer. 2019; 10:6175–6184.

Zhang X, Lan Y, Xu J, Quan F, Zhao E, Deng C, et al. CellMarker: a manually curated resource of cell markers in human and mouse. Nucleic Acids Research. 2019 jan; 47:D721.

Zhao J, Zhao J, Du F, Zhang Y, Shen G, Zhu H, et al. Cardia and Non-Cardia Gastric Cancer Have Similar Stage-for-Stage Prognoses After R0 Resection: a Large-Scale, Multicenter Study in China. Journal of gastrointestinal surgery : official journal of the Society for Surgery of the Alimentary Tract. 2016 apr; 20: 700–707.

Zhao P, Li L, Jiang X, Li Q. Mismatch repair deficiency/microsatellite instability-high as a predictor for anti-PD-1/PD-L1 immunotherapy efficacy. Journal of hematology & oncology. 2019 may; 12:.

Zhong QY, Fan EX, Feng GY, Chen QY, Gou XX, Yue GJ, et al. A gene expression-based study on immune cell subtypes and glioma prognosis. BMC Cancer 2019 19:1. 2019 nov; 19:1–10.

Zhong Y, Wan YW, Pang K, Chow LM, Liu Z. Digital sorting of complex tissues for cell type-specific gene expression profiles. BMC Bioinformatics 2013 14:1. 2013 mar; 14:1–10.

Zuo S, Wei M, Wang S, Dong J, Wei J. Pan-cancer analysis of immune cell infiltration identifies a prognostic immune-cell characteristic score (iccs) in lung adenocarcinoma. Frontiers in Immunology. 2020; 11:.

APÊNDICES

Apêndice A – Publicação: Mutational Signatures Driven by Epigenetic Determinants Enable the Stratification of Patients with Gastric Cancer for Therapeutic Intervention

cancers

MDPI

Article

Mutational Signatures Driven by Epigenetic Determinants Enable the Stratification of Patients with Gastric Cancer for Therapeutic Intervention

Jaqueline Ramalho Buttura [1,†], Monize Nakamoto Provisor Santos [1,2,†], Renan Valieris [1], Rodrigo Duarte Drummond [1], Alexandre Defelicibus [1], João Paulo Lima [1], Vinicius Fernando Calsavara [3], Helano Carioca Freitas [4,5], Vladmir C. Cordeiro de Lima [4,6], Thais Fernanda Bartelli [5], Marc Wiedner [7], Rafael Rosales [8], Kenneth John Gollob [6], Joanna Loizou [7,9], Emmanuel Dias-Neto [5,10], Diana Noronha Nunes [5] and Israel Tojal da Silva [1,*]

1 Laboratory of Bioinformatics and Computational Biology, A.C. Camargo Cancer Center, São Paulo 01508-010, Brazil; jaqueline.ramalho@accamargo.org.br (J.R.B.); monize.santos@grupofleury.com.br (M.N.P.S.); renan.valieris@accamargo.org.br (R.V.); rdrummond@accamargo.org.br (R.D.D.); alexandre.defelicibus@accamargo.org.br (A.D.); joao.lima@accamargo.org.br (J.P.L.)
2 Department of Genomics, Fleury Group, São Paulo 04344-070, Brazil
3 Department of Statistics and Epidemiology, A.C. Camargo Cancer Center, São Paulo 01508-010, Brazil; vinicius.calsavara@cshs.org
4 Medical Oncology Department, A.C. Camargo Cancer Center, São Paulo 01508-010, Brazil; helano.freitas@accamargo.org.br (H.C.F.); vladmir.lima@oncologiador.com.br (V.C.C.d.L.)
5 Laboratory of Medical Genomics, A.C. Camargo Cancer Center, São Paulo 01508-010, Brazil; thais.bartelli@accamargo.org.br (T.F.B.); emmanuel@accamargo.org.br (E.D.-N.); dnoronha@accamargo.org.br (D.N.N.)
6 Translational Immuno-Oncology Group, A.C. Camargo Cancer Center, São Paulo 01508-010, Brazil; kenneth.gollob@accamargo.org.br
7 CeMM Research Center for Molecular Medicine of the Austrian Academy of Sciences, 1090 Vienna, Austria; mwiedner@cemm.oeaw.ac.at (M.W.); jloizou@cemm.oeaw.ac.at (J.L.)
8 Department of Mathematics and Computer Science, University of São Paulo, Ribeirão Preto 14049-900, Brazil; rrosales@usp.br
9 Department of Medicine, Institute of Cancer Research, Medical University of Vienna and Comprehensive Cancer Center, 1090 Vienna, Austria
10 Laboratory of Neurosciences, Institute of Psychiatry, University of São Paulo, São Paulo 05403-903, Brazil
* Correspondence: itojal@accamargo.org.br; Tel.: +55-11-2189-5000 (ext. 2941)
† These authors contributed equally to this work.

Simple Summary: Mutational signatures due to DNA mismatch repair deficiency (dMMR) is common in many cancers. However, the prognostic value of dMMR-associated mutational signatures remains to be assessed. Here, we performed a *de novo* extraction of mutational signatures in a cohort of 787 patients with gastric cancer. We detected three dMMR-related signatures, one of which clearly discriminates tumors with *MLH1* gene silencing through hypermethylation of its promoter. We showed evidence that classification based on mutational signature exposure can be used to identify groups of patients with common clinical, immunological, and mutational features related directly to better prognosis.

Abstract: DNA mismatch repair deficiency (dMMR) is associated with the microsatellite instability (MSI) phenotype and leads to increased mutation load, which in turn may impact anti-tumor immune responses and treatment effectiveness. Various mutational signatures directly linked to dMMR have been described for primary cancers. To investigate which mutational signatures are associated with prognosis in gastric cancer, we performed a *de novo* extraction of mutational signatures in a cohort of 787 patients. We detected three dMMR-related signatures, one of which clearly discriminates tumors with *MLH1* gene silencing caused by promoter hypermethylation (area under the curve = 98%). We then demonstrated that samples with the highest exposure of this signature share features related to better prognosis, encompassing clinical and molecular aspects and altered immune infiltrate

check for updates

Citation: Buttura, J.R.; Provisor Santos, M.N.; Valieris, R.; Drummond, R.D.; Defelicibus, A.; Lima, J.P.; Calsavara, V.F.; Freitas, H.C.; Cordeiro de Lima, V.C.; Fernanda Bartelli, T.; et al. Mutational Signatures Driven by Epigenetic Determinants Enable the Stratification of Patients with Gastric Cancer for Therapeutic Intervention. *Cancers* **2021**, *13*, 490. https://doi.org/10.3390/cancers13030490

Received: 1 December 2020
Accepted: 20 December 2020
Published: 27 January 2021

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Copyright: © 2021 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (https://creativecommons.org/licenses/by/4.0/).

composition. Overall, the assessment of the prognostic value and of the impact of modifications in MMR-related genes on shaping specific dMMR mutational signatures provides evidence that classification based on mutational signature exposure enables prognosis stratification.

Keywords: mutational signature; gastric cancer; DNA mismatch repair; prognosis

1. Introduction

Cancer results from the sequential accumulation of DNA alterations, including single-nucleotide mutations [1] that arise from various endogenous and exogenous processes [2]. Distinct DNA-damaging processes leave characteristic nucleotide base-change footprints known as mutational signatures [3]. Researchers have extracted distinct mutational signatures by examining large sets of human cancer genomes, and some of these signatures have been registered in the COSMIC (CS) database (http://cancer.sanger.ac.uk/cosmic/signatures) [4]. This pan-cancer analysis revealed the significant heterogeneity of operational mutational processes, which encompass mutation-triggering events as diverse as the off-target activity of the *AID/APOBEC* family of cytidine deaminases, exposure to ultraviolet light, tobacco smoking, and defective DNA mismatch repair (dMMR) [5,6].

Collectively, the understanding of the mechanistic basis and etiology of mutational signatures may provide clues for cancer diagnosis and have prognostic value [7]. For example, six mutational signatures have been associated with the *BRAC1/2* gene dysfunction, and most likely are predictive of the response to treatment with poly-ADP ribose polymerase (PARP) inhibitors [8]. Homologous recombination repair (HRR) deficiency features based on these signatures allowed the prediction of BRCAness in patients with breast cancer with a 98.7% sensitivity [8]. Additionally, given that nucleotide excision repair (NER)-deficient tumors are more sensitive to certain treatments, somatic variations in the *ERCC2* gene, which encodes a key protein of the NER pathway, have also been linked to characteristic mutational signatures [9,10]. Other mutational processes have been associated with the harboring of biallelic *MUTYH* germline mutations [11], which may indicate deficient base excision repair (BER). Affected patients are eligible for genetic counseling [12] and might benefit from immunotherapy [13].

In addition to the HRR, NER, and BER repair pathways, another mechanism underlying oncogenic genomic variations occurs in tumors with impaired DNA mismatch repair (MMR) which harbor elevated frequencies of single-nucleotide variants (SNVs) and exceptionally high indel rates [14]. Recent studies have demonstrated that various tumors with mismatch repair deficiency (dMMR; glioblastomas and gastrointestinal, endometrial, and prostate tumors) are more responsive to programmed cell death protein 1 (PD1) immune checkpoint inhibitors than MMR-proficient tumors are [15–17]. The microsatellite instability (MSI) phenotype arises mainly because of the dMMR mechanism [18], and thus serves together with the immunohistochemical detection of MMR genes (e.g., *MLH1* and *MSH2*) as a biomarker for the identification of patients with MSI/dMMR and a guide for decisions about their clinical treatment.

A set of four mutational signatures (CS-6, CS-15, CS-20, and CS-26) has been associated with the dMMR and/or MSI phenotypes, mainly in the context of colorectal cancer and Lynch syndrome. CS-6 and CS-15 have also been described in the context of gastric cancer [19]. Furthermore, an improved prognosis of gastric cancer has been associated with dMMR/MSI status, without consideration of the mutational signature landscape [18,19]. It is thought that mutational signatures arise from multiple changes in pathway component events [3,4], and thus its evaluation as a classifier may be more informative than a unique clinical or molecular feature, which allows grouping patients with similar phenotypes based on their mutational profiles.

In this study, we sought to identify mutational signatures associated with the prognosis of gastric cancer and determine the significance of molecular events in MMR genes that

shape these signatures in MMR-deficient gastric adenocarcinomas. The prognostic value of the presence of these signatures was evaluated in a cohort of 787 patients with gastric cancer, including 439 patients whose data is registered in The Cancer Genome Atlas (TCGA) and a validation cohort of 170 patients with gastric cancer [20]. We further investigated whether local tumor immune responses and prognoses varied according to the dMMR exposure load. The stratification of patients with gastric cancer by dMMR mutational signature appears to be related to tumor microenvironment and molecular features predictive of the responsiveness to immune checkpoint blockade. Further studies are recommended to confirm this in clinical practice.

2. Results

2.1. Mutational Signatures

Using signer [21] analysis to estimate de novo mutational signatures across three gastric cancer cohorts, we identified seven mutational signatures (S1–7; Figure 1A) which are related to signatures described in the CS database (CSs), as reflected by cosine similarity scores (Figure 1B). S1 (related to CS-1) is associated with endogenous mutational processes initiated by the spontaneous deamination of 5-methylcytosine; S2, S4, and S5 (related to CS-6 and CS-15, CS-20, and CS-21 and CS-26, respectively) are associated with dMMR and/or MSI; S3 (related to CS-3) is associated with the failure of DNA double-strand break repair by homologous recombination; S6 (related to CS-17) is of unknown etiology; and S7 (related to CS-10) is associated with error-prone polymerase activity in the catalytic subunit of DNA polymerase epsilon (POLE).

CS-3 (S3; Figure 1 and Figure S1) was the predominant signature, and the observations in this study support previous characterizations of it in gastric cancer samples as having a very high prevalence of small indels and base substitutions due to the failure of DNA double-strand break repair by homologous recombination [19]. This finding suggests that 7–12% of patients with gastric cancers could benefit from platinum or PARP inhibitor therapy. Notably, however, some patients not exposed to CS-3 were found to be highly exposed to signatures associated with dMMR (S2, S4, and S5; Figure S1). This distinct group of patients harbors features that might have prognostic relevance; we investigated this possibility further.

2.2. dMMR Signatures and Prognostic Features

To assess whether dMMR signature exposure had prognostic value in patients with gastric cancer, we first evaluated the influences of each mutational signature exposure and its main possible clinical and molecular prognostic features—such as age at diagnosis, ethnicity, tumor pathological stage, Lauren classification, anatomic site, tumor mutational burden (TMB), and MSI status—on overall survival (OS) by fitting a simple Cox regression model (Figure S2). We then fitted a multiple Cox regression model to the dataset using prognostic features with significant p values ($p < 0.05$) in the simple model (Figure S2). Data from 584 patients with gastric cancer and no metastasis at diagnosis for whom vital status information (alive/dead) was available were included in simple and multiple Cox regression models. The median and mean follow-up durations for these patients were 28.9 months (95% confidence interval (CI), 25.8–32.1 months) and 36.2 months (95% CI, 32.9–39.5 months), respectively.

Relative to other dMMR signatures, S4 exposure burden was associated with an improved OS (hazard ratio (HR) = 0.59; 95% CI, 0.37–0.96; Figure 2A and Figure S3). Thus, we focused on S4, which has the potential to offer important, clinically actionable information for prognosis.

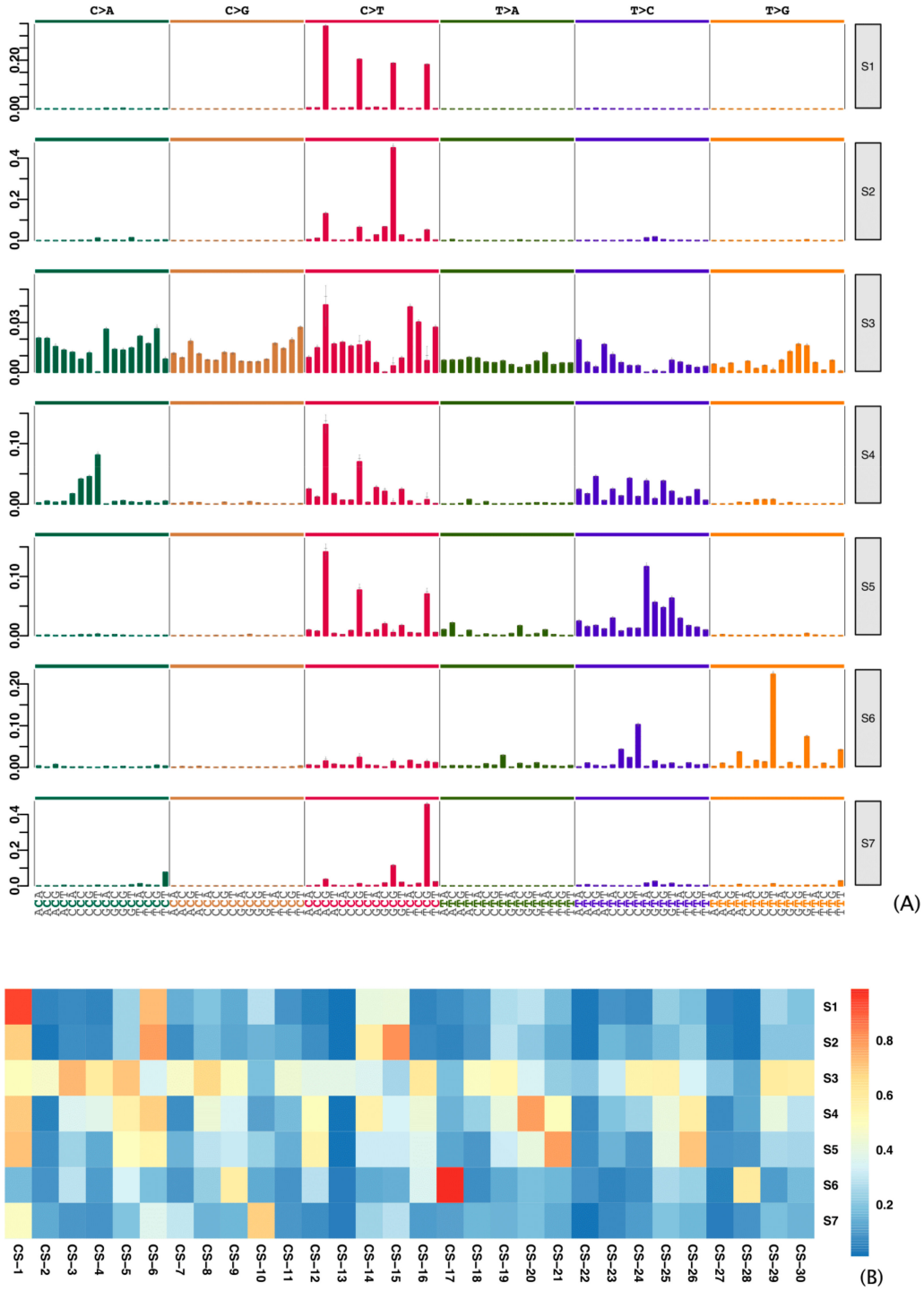


Figure 1. De novo mutational signatures (Ss) in gastric cancer samples. (**A**) Mutational signatures in gastric cancer samples from 787 patients, extracted from The Cancer Genome Atlas, International Cancer Genome Consortium, and cbioPortal cohorts. (**B**) Heatmap of cosine similarities between de novo mutational signatures and COSMIC signatures (CSs).

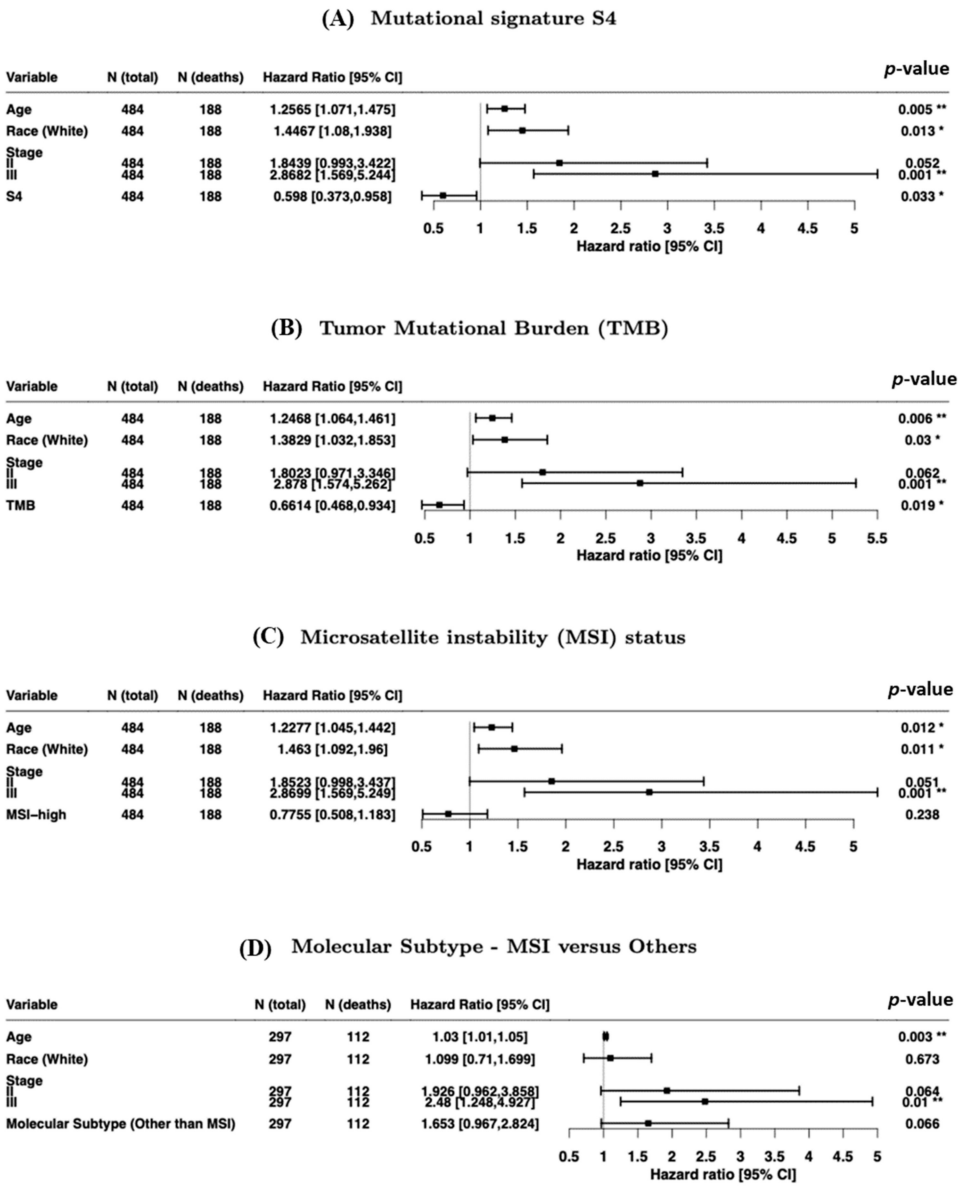


Figure 2. Forest plots of hazard ratios for overall survival (multiple Cox regression models) for (**A**) mutational signature S4, (**B**) the tumor mutational burden (TMB), and (**C**) microsatellite instability (MSI) status and (**D**) MSI molecular subtype. CI, confidence interval. * $p < 0.05$, ** $p < 0.01$.

Our analysis also revealed that higher TMBs were associated with improved OS (HR = 0.66; 95% CI, 0.46–0.93; Figure 2B), consistent with previous findings [22]. OS was not associated with high microsatellite instability (MSI-H) status or MSI molecular subtype (Figure 2C,D). The calibration curves for 2-year OS indicated that all the models were adequate (Figure S4).

We used the "maxstat" function in R to define groups according to S4 exposure. The optimal cut-point was in the highest quartile (Q3); thus, patients with S4 exposure $\geq$ the Q3 level were allocated to the $S4^{high}$ status, and all other patients (<Q3 level) were allocated to the $S4^{low}$ group. Survival curves differed significantly between these groups ($p = 0.03$, Log-rank test); median OS durations were 72 months (95% CI, 48.0 months–∞) in the $S4^{high}$ group and 37 months (95% CI, 28.0–68.0 months) in the $S4^{low}$ group considering the whole follow-up (Figure S5, throughout 5-years overall survival). Next, we used an independent gastric cancer cohort to validate that $S4^{high}$ conferred a survival benefit. In the Kaplan–Meier analysis, the median OS duration had not been reached at 5 years (95% CI, 38.2 months–

∞) in the $S4^{high}$ group and was 48 months (95% CI, 21.3 months–∞) in the $S4^{low}$ group, supporting the presence of a trend toward a survival benefit for $S4^{high}$ (Figure S6).

2.3. Clinical and Molecular Features in the S4 dMMR Groups

We next further examined the clinical and molecular features of the $S4^{high}$ and $S4^{low}$ groups. We first determined whether S4 exposure was represented equally in the three cohorts. To avoid performing statistical tests with different numbers of samples, a subsampling procedure was applied; 24 samples were selected randomly from each cohort and the Kruskal–Wallis test was applied to this sub cohort. The procedure was repeated 1000 times, and all the p values were >0.05, indicating that the S4 exposure was similar in all cohorts. These results reinforce the prognostic value of S4 signature independent of ethnic background.

The clinical features associated previously with the improved prognosis of gastric cancer, such as distal anatomic site and intestinal histology [23], were also enriched in the $S4^{high}$ group (Table 1). This group also showed a significantly greater occurrence of clinical variables known to be associated with worse gastric cancer prognosis, such as cardia/proximal anatomic site, diffuse histology, positive lymph-node metastasis (stage N+), and advanced pathological stage (III and IV) [23] (Table 1). In addition, the predicted MSI-H status, MSI, and POLE molecular subtypes were enriched in the $S4^{high}$ group, whereas the genomically stable (GS) and chromosomal instability (CIN) molecular subtypes were enriched in the $S4^{low}$ group (Table 1). Most (n = 119/160 (74%)) of the MSI-H cases were in the $S4^{high}$ group, but the $S4^{low}$ group unexpectedly contained a small proportion of such cases and the $S4^{high}$ group contained non-MSI-H cases (Table 1). Survival curves revealed a trend for worse prognosis for MSI-H cases in the $S4^{low}$ group (median OS duration, 9.07 months; 95% CI, 9.0 months–∞) than non-MSI-H cases in the $S4^{high}$ group (median OS, 53 months; 95% CI, 20.0 months–∞; Figure 3A). The prognosis seems to be better for cases with the diffuse histological subtype in the $S4^{high}$ group (median OS not reached; 95% CI, 24.0 months–∞) than the intestinal histological subtype in the $S4^{low}$ group (median OS 43.1 months; 95% CI, 28.0 months–∞; Figure 3B). These results indicate that classification according to mutational signatures improved the stratification of patients within the prognostic groups, independent of their previous clinical or molecular classification.

We also examined the tumor heterogeneity, TMB, and neoantigen count. Tumor heterogeneity was evaluated quantitatively by the examination of the distribution of allele frequencies and the calculation of mutant-allele tumor heterogeneity (MATH) scores [24]. We then performed a correlation analysis that incorporated the MATH score, S4 exposure, and TMB (Figure 4). Correlations of the MATH score with TMB and S4 exposure were negative in the $S4^{high}$ group and positive in the $S4^{low}$ group. In addition, the MATH scores were higher in the $S4^{low}$ group than in the $S4^{high}$ group ($p = 3.711 \times 10^{-12}$, Mann–Whitney U test). The TMB and neoantigen load correlated positively with S4 exposure in both groups (Figure 4).

These findings suggest that tumors with high S4 exposure were more homogeneous in the $S4^{high}$ group, and that reduced tumor heterogeneity together with a high TMB and high neoantigen load is determinant of a good prognosis.

Table 1. Clinicopathological features of gastric cancer samples according to mutational signature 4 DNA mismatch repair deficiency groups.

Variable	All (n = 787)		$S4^{low}$ (n = 590)		$S4^{high}$ (n = 197)		p Value
	N	%	N	%	N	%	
Age (mean ± SD)	64.17 ± 11.73		64.17 ± 11.73		66.25 ± 10.68		0.0024
Gender	767	97	571	91	196	99	
Female	274	0.36	192	0.34	82	0.42	0.0473
Male	493	0.64	379	0.66	114	0.58	
Race	726	92	546	93	180	91	
White	275	0.38	203	0.37	72	0.40	0.7427
Black	13	0.02	11	0.02	2	0.01	
Asian	437	0.60	331	0.61	106	0.59	
Other	1	0.00	1	0.00	0	0.00	
Anatomic Site	627	80	495	84	132	67	
Cardia/Proximal	168	0.27	148	0.30	20	0.15	0.0015
Fundus/Body	212	0.34	166	0.34	46	0.35	
Antrum/Distal	242	0.39	178	0.36	64	0.48	
Other	5	0.01	3	0.01	2	0.02	
Histology Lauren	467	59	383	65	84	43	
Diffuse	150	0.32	132	0.34	18	0.21	0.0041
Intestinal	301	0.64	235	0.61	66	0.79	
Mixed	16	0.03	16	0.04	0	0.00	
Stage T	712	90	526	89	186	94	
T1-T2	181	0.25	132	0.25	49	0.26	0.8116
T3-T4	531	0.75	394	0.75	137	0.74	
Stage N	712	90	526	59	186	94	
N0	173	0.24	115	0.22	58	0.31	0.0144
N+	539	0.76	411	0.78	128	0.69	
Stage M	707	90	524	89	183	93	
M0	623	0.88	461	0.88	162	0.89	0.9422
M1	62	0.01	47	0.09	15	0.08	
MX	22	0.03	16	0.03	6	0.03	
Pathological Stage	715	91	546	93	169	86	
I	85	0.12	58	0.11	27	0.16	0.0386
II	220	0.31	160	0.29	60	0.36	
III	289	0.40	228	0.42	61	0.36	
IV	121	0.15	100	0.18	21	0.12	
Molecular Subtype	403	51	289	49	114	58	
CIN	223	0.55	206	0.71	17	0.15	<0.0001
GS	50	0.12	47	0.16	3	0.03	
EBV	38	0.09	33	0.11	5	0.04	
MSI	85	0.21	0	0.00	85	0.75	
POLE	7	0.02	3	0.01	4	0.04	
MSIseq Status	787	100	590	100	197	100	
MSI-H	160	0.20	41	0.07	119	0.60	<0.0001
Non MSI-H	627	0.80	549	0.93	78	0.40	
Immune Subtypes	388	49	285	48	103	52	
C1	128	0.33	101	0.35	27	0.26	<0.0001
C2	209	0.54	135	0.47	74	0.72	
C3	35	0.09	34	0.12	1	0.01	
C4	9	0.02	8	0.03	1	0.01	
C6	7	0.02	7	0.02	0	0.00	

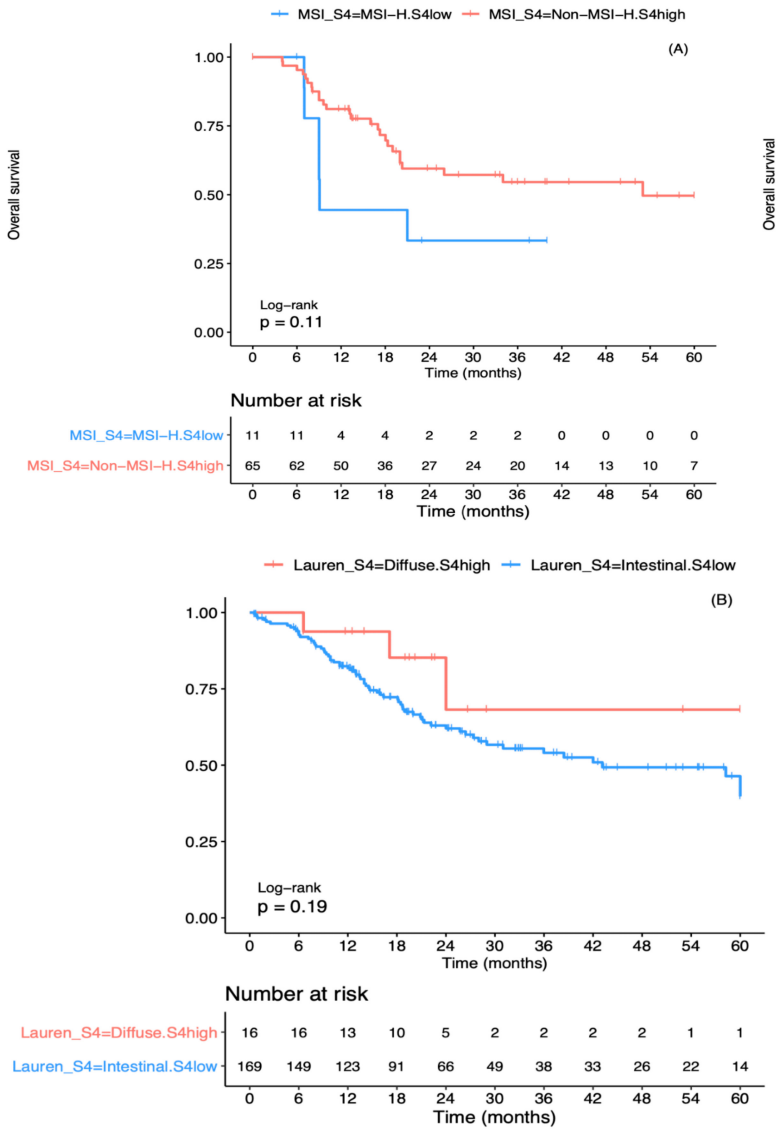


Figure 3. Five-year overall survival for subgroups of the mutational signature $S4^{high}$ and $S4^{low}$ groups in the public cohort. (**A**) Comparison of cases with high microsatellite instability (MSI-H) and in the $S4^{low}$ group with non-MSI-H cases in the $S4^{high}$ group. (**B**) Comparison of cases with diffuse histology in the $S4^{high}$ group with those with intestinal histology in the $S4^{low}$ group. Log rank test was used to compare the survival distributions of two groups.

2.4. S4 is Associated with Epigenetic Changes and Mutational Background in MMR Genes

To investigate the mechanisms underlying S4, we evaluated the epigenetic context and mutational landscape of MMR genes in the $S4^{high}$ and $S4^{low}$ groups. Although the genes associated with dMMR have been identified, the underlying main genetic modifications that lead to each dMMR signature remain poorly characterized.

To improve our understanding of the determinant changes that influence the dMMR signatures detected in this study, we first searched for epigenetic changes in the MMR genes. In line with previous findings [25,26], we observed the down-regulation of *MLH1* expression, driven by the hypermethylation of its promoter (Figure S7, rho = −0.81, *p*-value < 0.001). To further assess the manner in which mutational exposure is associated with epigenetic changes in the *MLH1* gene, simple and multiple logistic regression models were fitted to the dataset (Table 2). S4 exposure burden was associated with an increased chance of *MLH1* promoter methylation (h*MLH1*; odds ratio (OR) = 22.561; 95% CI, 7.909–64.353) and S5 exposure burden was associated with a decreased chance of such

methylation (OR = 0.107; 95% CI, 0.048–0.238); no such association was found for the S2 exposure burden. This model showed adequate to good performance (Hosmer–Lemeshow goodness-of-fit test: $\chi^2(8) = 10.257$, $p = 0.247$; Brier score = 0.0364; Figure 5A) and excellent discrimination power (area under the receiver operating characteristic curve (AUC) = 0.982; 95% CI, 0.971–0.994; Figure 5B). Using the Youden index, we determined that the best cutoff value was 0.125, which yielded a 95.45% sensitivity and a 95.82% specificity (Figure 5B).

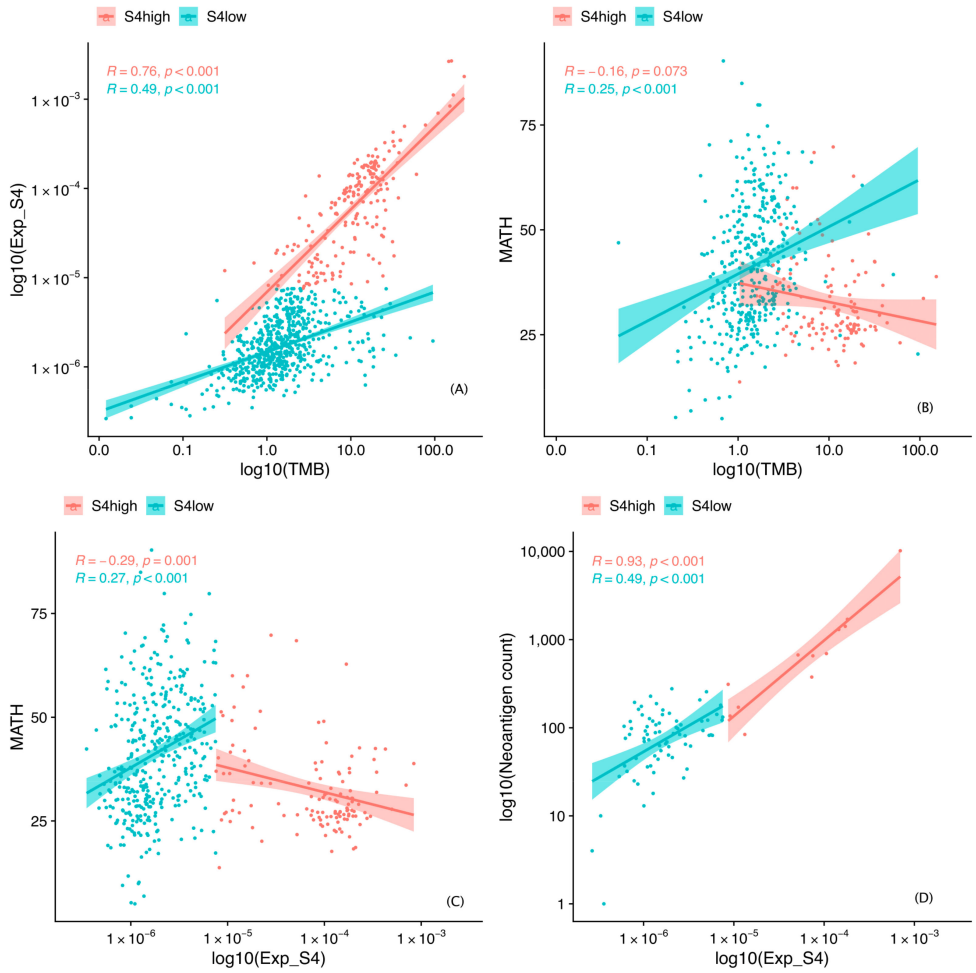


Figure 4. Scatter plots of Spearman correlation between molecular features. S4, mutational signature 4. TMB, tumor mutational burden; MATH, mutant-allele tumor heterogeneity.

Table 2. Simple and multiple logistic regression results for *MLH1* methylation and DNA mismatch repair deficiency mutational signatures.

Variable	Coefficient	Standard Error	CI (95%) for Coefficient		*p*-Value	OR	CI (95%) for Coefficient	
			Lower	Upper			Lower	Upper
Simple logistic regression model								
ExpS2	4.772	0.5226	3.748	5.796	<0.0001	118.155	42.424	329.078
ExpS4	3.2424	0.3469	2.562	3.922	<0.0001	25.595	12.968	50.518
ExpS5	0.2725	0.1674	−0.056	0.601	0.104	1.313	0.946	1.823
Multiple logistic regression model								
Intercept	−2.640	0.300	−3.227	−2.052	<0.0001			
ExpS2	1.304	0.730	−0.126	2.733	0.074	3.682	0.881	15.386
ExpS4	3.1162	0.5348	2.068	4.164	<0.0001	22.561	7.909	64.353
ExpS5	−2.231	0.4067	−3.028	−1.434	<0.0001	0.107	0.048	0.238

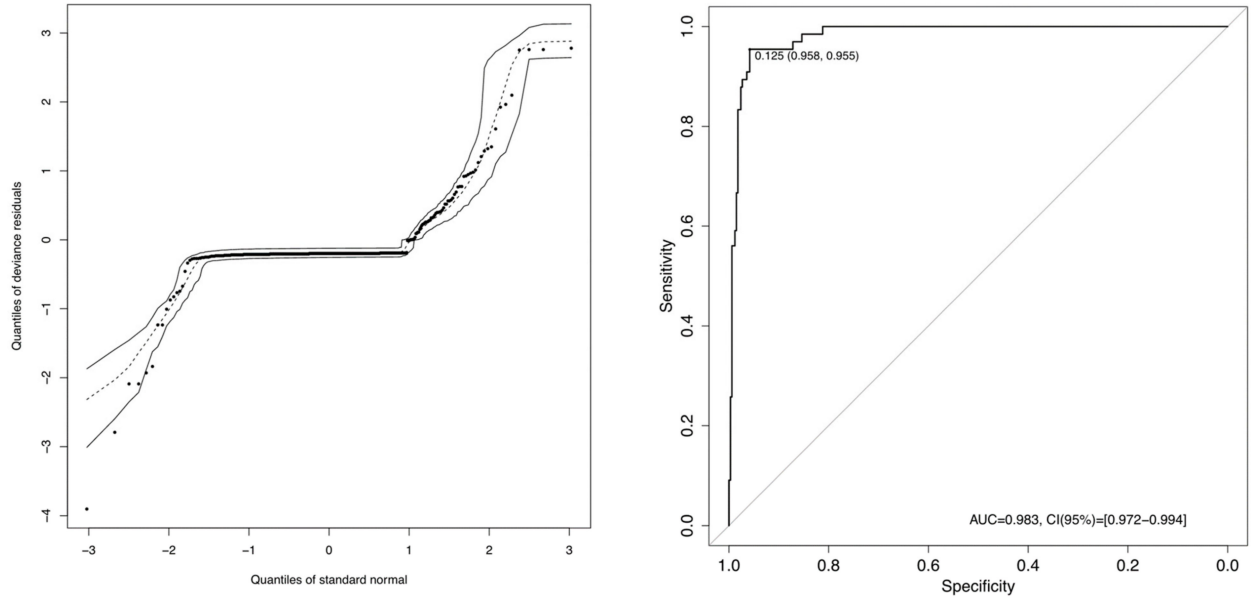


Figure 5. Performance and power discrimination of the logistic model. (**A**) Q–Q plot showing that the propensity score model was adequately calibrated, according to the Hosmer–Lemeshow test. (**B**) Summary of the receiver operating characteristic (ROC) curve showing the power of discrimination for *MLH1* methylation in mutational signature S4. AUC, area under the receiver operating characteristic curve; CI, confidence interval.

We next looked for somatic and germline SNVs and indels in MMR genes (*LIG1*, *POLE*, *EXO1*, *MLH1*, *MLH3*, *MSH2*, *MSH3*, *MSH5*, *MSH6*, *PCNA*, *PMS1*, *PMS2*, *PMS2L3*, *PMS2L4*, *POLD1*, *POLD2*, *POLD3*, *POLD4*, and *SSBP1*). Six percent (n = 12/197) of patients harbored somatic variations in the *MLH1* gene; these patients were allocated to the $S2^{high}$ and $S4^{high}$ groups, and no such mutation was observed among patients in the respective S^{low} groups. Nine percent (n = 17/197) of patients in the $S2^{high}$ group harbored somatic variations in the *MLH3* gene. We also found that only 8% of patients in the $S5^{high}$ group (n = 8/100 considering TCGA cohort) harbored *MSH5* germline mutations. Overall, we observed few cases of mutated MMR genes in the S^{high} groups.

These observations suggest that germline SNVs, somatic SNVs, and indels are not the major modifications affecting MMR gene expression levels. hMLH1 was observed in almost 60% of individuals in the S4 group ($\beta \geq 0.3$), considering the TCGA dataset. To validate this mechanism as a causative factor for S4, we examined genomic sequencing data from three HAP1 cell samples (two MLH1 knockout and one MLH1 wild type). We observed greater S4 exposure in the MLH1 knockout cell lines and greater S5 exposure in the parental cell line (Figure S8), which identifies the absence of MLH1 as a strong cause of the S4 mutational signature.

2.5. Somatic Changes Associated with S4

We further evaluated significantly mutated genes other than those related to MMR in the $S4^{high}$ and $S4^{low}$ groups. We observed increased numbers of SNVs and indels in the $S4^{high}$ group; most mutations in the $S4^{low}$ group were SNVs. These findings were expected, considering that MSI/dMMR harbors a mutator phenotype [14].

The significantly mutated [27] gene set in the $S4^{high}$ group consisted of 102 genes. The most commonly mutated genes in this group were *ARID1A* (42%), *KMT2D* (35%), and *TP53* (31%), in accordance with previous findings [28]; 56 other genes presented mutations in at least 10% of patients (Table S1a). Somatic mutations in chromatin-regulating genes, such as *KMT2D* (also known as *MLL2*) and *ARID1A*, may be associated with improved survival [26]. Of the 24 significantly mutated genes identified in the $S4^{low}$ group, 12 were oncogenes associated with tumor progression or tumor suppressor genes (*PIK3CA*,

KRAS, *RHOA*, *CDH1*, *CTNNB1*, *ITGAV*, *SMAD4*, *TP53*, *CDKN2A*, *APC*, *PTEN*, and *PIK3R1*; Table S1c). These findings support previous reports of the occurrence of *APC*, *CTNNB1*, *SMAD4*, and *SMAD2* mutations among 215 non-hypermutated tumors from the TCGA cohort, with somatic mutations in *CDH1* and *RHOA* enriched in the GS and/or diffuse histologic subtypes [28], as seen in our $S4^{low}$ group.

To assess other mechanisms underlying the remaining ~40% of mutations in the $S4^{high}$ samples, we divided this group into hypermethylated and hypomethylated subgroups according to the *MLH1*-based methylation levels. The *TP53* gene was associated with hypomethylation in the $S4^{high}$ group (OR = 0.314; 95% CI, 0.131–0.736; $p = 0.006$). However, in a comparison of all significantly mutated genes (MutSigCV analysis, $p < 0.05$) between the $S4^{high}$ and $S4^{low}$ groups, *TP53* was the only mutated gene associated with the $S4^{low}$ group (OR = 0.428; 95% CI, 0.293–0.622; $p = 3.542 \times 10^{-6}$). An additional analysis of the somatic interactions ("somaticInteractions" function in the "maftools" package, available in R) revealed exclusive interactions of *TP53* in both groups, indicating different pathways underlying tumorigenesis for each group [29]. Additional studies are necessary to indicate a biomarker that encompasses both hyper and hypomethylated $S4^{high}$ patients.

2.6. Tumor Microenvironment in S4 Groups

To investigate the association of the tumor microenvironment with the improved clinical outcomes seen in the $S4^{high}$ group relative to the $S4^{low}$ group, we determined the immune cell infiltrate and stromal cell compositions in the groups using biomarker gene expression methodologies. We found significantly greater proportions of infiltrating cytotoxic and pro-inflammatory immune cells, such as CD8+ central and effector memory T cells, CD4+ memory T cells, T helper 1 cells, gamma/delta T cells, natural killer (NK) cells, M1 macrophages, and plasmacytoid dendritic cells, in the $S4^{high}$ group than in the $S4^{low}$ group (Figure 6A and Figure S9). In contrast, immature and immune regulatory dendritic cells were more common in the $S4^{low}$ group (Figure 6A and Figure S9).

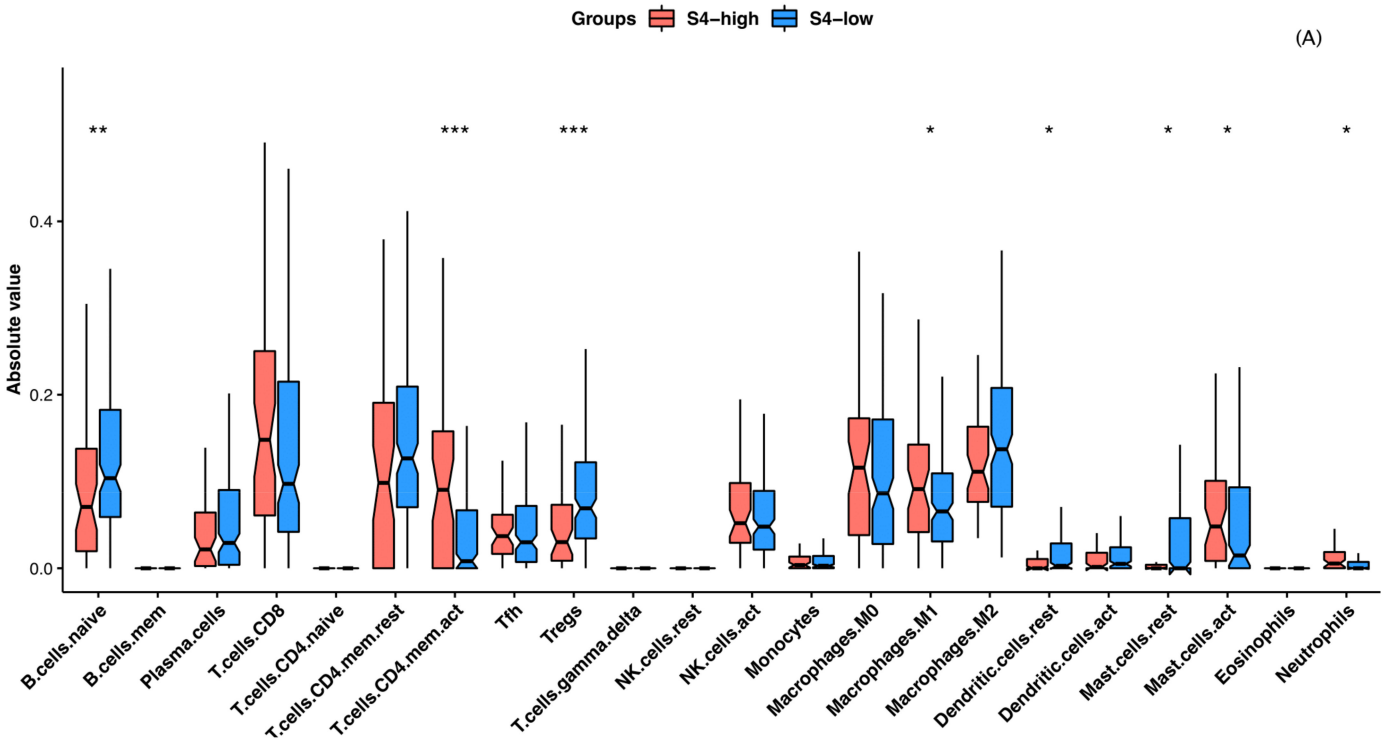


Figure 6. *Cont.*

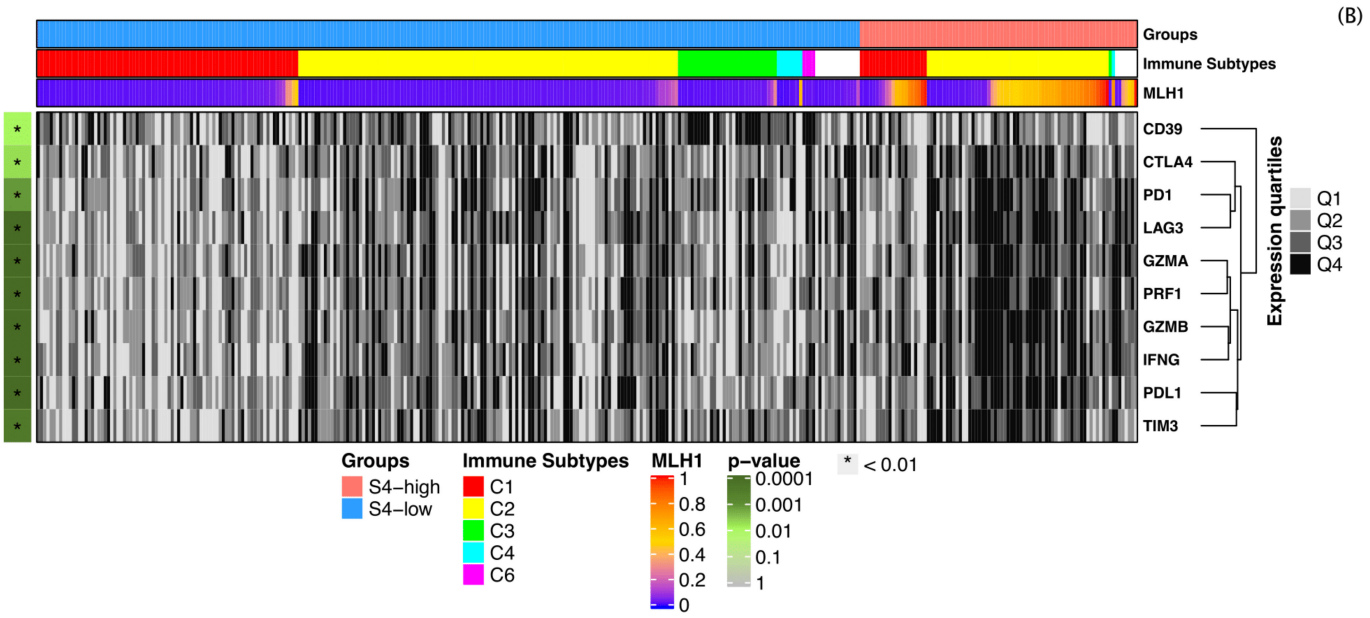


Figure 6. The main immunological features associated with the mutational signature (S)4^{high} and S4^{low} groups. (**A**) Boxplots showing the absolute quantification of immune infiltrate cells estimated by CIBERSORT. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$, Mann–Whitney *U* test. (**B**) Heatmap of immune effector gene (cytotoxic and immune checkpoints) expression. Normalized gene expression levels for each marker gene were classified in quartiles. Q1, 0–25%; Q2, 25–50% (median); Q3, 50–75%; Q4, 75–100%. P values to the left of the heatmap for S4^{high} vs. S4^{low} expression were obtained using the Mann–Whitney *U* test. These comparisons are also shown as boxplots in Figure S9.

To further characterize the immune regulatory environment associated with clinical outcomes, we compared the expression levels of key genes coding for immunoregulatory and effector molecules that have proven to be important for the control of many cancers [30,31]. Genes encoding the CD8+ T-cell–related cytolytic molecules granzyme A/B and perforin-1 (*GZMA* and *GZMB* and *PRF1*, respectively), the inflammatory T-cell response-related cytokine interferon (IFN)-γ (*IFNG*), other proinflammatory cytokines (*IL1B*, *IL6*, and *IL8*), and the NK cell killer-cell immunoglobulin-like receptor family, as well as T-cell activation marker genes (*IL2RA* and *ICOS*), showed greater expression in the S4^{high} than in the S4^{low} group (Figure 6B and Figure S10). No difference in the immunosuppression-related genes *TGFB1*, *IL-10*, and *FOXP3* was observed between groups, whereas the expression of *ENTPD1* (encoding CD39, a protein associated with T regulatory cell (Treg) immunosuppression activity) [32] was greater in the S4^{low} than in the S4^{high} group (Figure 6B and Figure S10). Importantly, the expression of immune checkpoint inhibitor genes (*PDCD1* for the PD1 receptor; *CD274* for programmed death ligand (PDL)-1; *PDCD1LG2* for PD-L2; *HAVCR2* for TIM3, *LAG3*, and *CTLA4*) was also greater in the S4^{high} group (Figure 6B and Figure S10), suggesting a relationship with a more immunologically activated tumor microenvironment [33]. The expression of *HLA*, antigen-processing, and presentation-related genes (e.g., *CD86, B2M, HLA* class II genes, *HLA-E, HLA-C, TAP1*, and *TAP2*) was also greater in the S4^{high} group (Figure S10). The observation of previously characterized immune subtypes [34] reinforces the finding that the tumor microenvironment in the S4^{high} group is more immunologically active, composed predominantly of the C2 (IFN-γ) immune subtype, with a significantly greater proportion of this subtype than that observed in the S4^{low} (Table 1, Figure 6B). This subtype has been associated with highly mutated tumors [34]. On the other hand, the S4^{low} group displayed greater proportions of the C3 (inflammatory) and C1 (wound healing) immune subtypes [34] (Table 1, Figure 6B). The C2 immune subtype appeared to be less activated in the S4^{low} group than in the S4^{high} group, with a reduced relative gene expression of immune effector molecules (Figure 6B). Together, these findings indicate that the immune microenvironment in the S4^{high} group is strongly activated relative to that in the S4^{low} group.

In the evaluation of stromal cells by xCell analysis, scores for myocytes, chondrocytes, hematopoietic stem cells, endothelial cells (microvascular and lymphatic), and fibroblasts (pericytes and mesangial cells) were higher in $S4^{low}$ group (Figure S9). Pericytes and endothelial cells are important cellular components of the tumor microenvironment that have been associated with the worst prognosis, considering the high risks of angiogenic events and metastasis [35–37]. Overall, the stromal score in the $S4^{low}$ group was higher than that in the $S4^{high}$ group, demonstrating an association with the worst prognosis (Figure S9).

3. Discussion

In this study, we conducted a comprehensive and integrated analysis of the impacts of MMR-related gene alterations on specific mutational signatures associated with gastric cancer. We present evidence that the identification of dMMR can be used not only for MSI phenotype classification, but also as a potential indicator of prognosis besides assembling potential patients with a peculiar tumor microenvironment that would respond to immune checkpoint inhibitors.

We performed a de novo extraction of mutational signatures based on somatic SNVs across four whole-exome sequencing cohorts, encompassing 787 gastric cancer samples derived mainly from populations of European and Asian descent. We found seven mutational signatures, three of which were related to dMMR. S4, related to the previously described CS-20, was the only dMMR signature with significant prognostic value; this value was validated in a local cohort of patients with gastric cancer, which was distinct in terms of molecular ancestry and some clinical and molecular features (e.g., Lauren's histology and tumor heterogeneity). Diffuse/mixed histology was predominant in this cohort, whereas the public cohorts were enriched in the intestinal subtype. Furthermore, the $S4^{low}$ group in this independent cohort was less heterogeneous than the $S4^{low}$ group from the public cohorts and the $S4^{high}$ groups from both cohorts. Nevertheless, patients in the $S4^{high}$ groups, including that of this cohort, had a better prognosis.

The main mechanism associated with MMR impairment in samples with S4 exposure seemed to be h*MLH1*. The disruption of *MLH1* in vitro using a CRISPR/Cas9 assay reproduced CS-20 [38], which resembles S4. In this study, we identified an endogenous epigenetic mechanism for this signature in patients with gastric cancer. We also reproduced S4 in an isogenic cell model in which *MLH1* knockout cells had a high S4 exposure. Thus, we conclude that the loss of *MLH1* gene expression due to promoter hypermethylation and mutagenesis loss of function result in the same mutational signature.

The CpG island methylator phenotype is a well-documented early event in tumorigenesis, preceding h*MLH1* in solid tumors, which in turn drives the MSI-H phenotype [25,26,28]. In contrast, the low MSI and microsatellite stability gastric carcinoma subtypes have unmethylated *MLH1* promoters and regular MLH1 activity [25]. In this study, we observed that most MSI-H cases were in the $S4^{high}$ group, but that a small fraction of these cases were in the $S4^{low}$ group. This observation is in line with previous reporting that about one fourth of MSI-H cases have distinct molecular features and poor prognoses [39]. In addition, the presence of non–MSI-H cases in our $S4^{high}$ group shows that mutational signature exposure can be used to cluster samples independently of their MSI status. Furthermore, we identified a few cases (4%) in the $S4^{high}$ group that did not present h*MLH1,* but carried somatic mutations in the *MLH1* gene that apparently lead to a loss of function of the encoded protein. Thus, we demonstrated that h*MLH1* is the main mechanism driving S4 (CS-20) in gastric cancer. Furthermore, we observed that patients without h*MLH1* harbor *TP53* mutations, suggesting another mechanism associated with dMMR signatures in this subset of patients.

We also demonstrated a strong correlation between S4 exposure and the TMB, which showed significant prognostic value for survival. Hypermutated tumors have been shown to have a better prognoses and good responses to immunotherapy, apparently due to neoantigen enrichment and intrinsic antitumor immune responses [40,41]. However, most thresholds for the identification of samples with high TMBs vary with tumor type,

and some do not necessarily predict a better treatment response due to intratumoral heterogeneity [41–43]. Here, we observed that most mutations associated with the $S4^{high}$ signature are clonal, which is important for the prediction of better treatment.

High degrees of intratumoral heterogeneity have been associated with incomplete responses to therapy, higher relapse rates, and poor clinical outcomes [44,45]. The increased genomic instability observed in MSI/dMMR and CIN tumors is the major driver of such heterogeneity [44,46]. However, the most unstable tumors (those with the greatest somatic SNV or copy number alteration burdens) are not the most intrinsically heterogeneous [46]. The greatest degree of intratumoral heterogeneity was found in tumors with relatively large numbers of somatic mutations and copy number alterations, which can be associated with exogenous mutagens, including those induced by viral infections and tobacco smoking. These tumors have large numbers of subclonal mutations related to late events and exhibit frequent chromosomal instability associated with the CIN subtype, *TP53* mutations, and APOBEC-related mutational signatures (previously related to the Epstein–Barr virus (EBV) gastric cancer subtype) [6,46]. Similarly, we observed greater tumor homogeneity in the $S4^{high}$ than in the $S4^{low}$ group. Thus, antitumor immune responses may be more effective in the $S4^{high}$ group due to the intermediate to high TMBs and MSI and CIN molecular phenotypes, associated with lesser tumor heterogeneity. A recent meta-analysis revealed the importance of the MSI status to the treatment response in patients with gastric cancer, as its findings suggested that patients with MSI-H status may not benefit from perioperative or adjuvant therapy and could undergo surgery without these treatments [47].

Several studies have shown that the tumor microenvironment at diagnosis can be used to predict treatment response and clinical outcome [48,49]. The balance of inflammatory/cytotoxic immune cells with elements of an effective antitumor response, including regulatory cells and suppressor signals, may indicate which patients will intrinsically have such responses and thus a better prognosis.

The EBV and MSI subtypes of gastric cancer have been associated with greater immune infiltration and responsiveness to immunotherapy, as well as better prognoses [49]. In this study, we observed many elements indicating that the tumor microenvironment in the $S4^{high}$ group was more active than that in the $S4^{low}$ group, aggregating both cases of MSI and EBV molecular subtypes. In general, the absolute quantification of immune cell subtypes and differential gene expression using CIBERSORT [50] and the calculation of gene set enrichment analysis (GSEA) scores in xCell [51] revealed a greater activity of proinflammatory and cytotoxic cells, as well as more antigen processing and presentation, in the $S4^{high}$ group. In contrast, although the $S4^{low}$ group contained some immunogenic tumors, the predominant environment in this group was enriched in Treg lymphocytes and M2 macrophages, related to worse prognosis [48,49]. In addition, the $S4^{low}$ group harbored stroma-enriched tumor microenvironments associated with poor prognosis [36].

4. Materials and Methods

4.1. Clinical and Genomic Data from Public Cohorts

Clinical and molecular data for 787 gastric adenocarcinoma samples were extracted from the following non-redundant public cohorts: (i) TCGA (STAD-US cohort, n = 439) (ii) cBioPortal (n = 226; composed by UHK (n = 19) [52], UHK/UHK_Pfizer (n = 100) [52], TMUCIH (n = 17) [53] and GACA-JP (n = 30) [54] cohorts), and (iii) the International Cancer Genome Consortium (ICGC, GACA-CN cohort, n = 122; Table S2 (available at https://dcc.icgc.org/projects/GACA-CN)). TCGA data were accessed on 4 October 2018 and correspond to the MC3 variant calling project, which is a comprehensive effort to detect consensus mutations and forms the basis of the Pan-Cancer Atlas initiative [55]. Data from the cBioPortal and ICGC cohorts, composed of patients of Asian descent, were last assessed on 9 January 2019.

Raw reads from matched non-tumor exomes including MMR genes were downloaded from the TCGA dataset and used to detect germline SNVs following the Genome Analysis Tool Kit's (GATK's) best practices for germline alteration calling. We also used filters to

select mutations with a variant allele frequency $\geq$0.3 and a minimum depth coverage of 10 reads. The "dbNSFP_MetaLR_rankscore" annotation was used to filter out synonymous mutations (dbNSFP_MetaLR_rankscore $\leq$0.6). Fragment per kilobase million values from 380 TCGA stomach adenocarcinoma (STAD) samples were used to normalize the gene expression profiles. Methylation, quantified using β values (range, 0–1), was available for the TCGA cohort. We then used CpG sites to detect hypermethylated and hypomethylated MMR gene promoters (threshold, $\beta \geq 0.3$). The baseline clinical features of the cases from each cohort are summarized in Table S3.

4.2. Clinical and Genomic Data from the Validation Cohort

Patients in the validation cohort were enrolled prospectively in an institutional study of the epidemiology and genomics of gastric adenocarcinomas in Brazil [20]. This study was approved by the local ethics committee, and all the participants provided written informed consent [20]. The clinical and molecular characteristics of the 170 patients in the validation cohort are provided in Table S4.

Genomic DNA was extracted from frozen tissues (n = 165) using the AllPrep DNA/RNA Mini Kit (Qiagen), QIASymphony THC 400 device (Qiagen), or phenol/chloroform/isoamyl alcohol precipitation. Genomic DNA was extracted from formalin-fixed paraffin-embedded tissues (n = 4) using the RecoverAll Total Nucleic Acid Isolation Kit (Thermo Fisher). One sample was from a gastric wash. Exome libraries were prepared using the Agilent SureSelect V6 kit and sequenced using Illumina platforms (HiSeq4000, 100 bp, n = 33; Novaseq, 150 bp, n = 137; paired-end reads for both). Somatic SNVs were called using an in-house pipeline following the GATK's best practice guidelines [56], as described previously [6]. Briefly, the raw reads were aligned using the Burrows–Wheeler aligner with default settings to assemble GRCh38. Next, the alignment files were converted to a binary alignment map (BAM) files, sorted, and filtered to exclude reads with mapping quality scores <15. The retained reads were processed using SAMtools (v1.9) and Picard (v3.8; https://broadinstitute.github.io/picard/), which exclude low-quality reads and polymerase chain reaction duplicates, respectively. Finally, somatic SNV calling was performed for the whole-exome data from analysis-ready BAM files using Mutect2 (v3.8) for tumor samples, followed by filtering with a panel of 16 unmatched non-tumor leukocyte samples. Extensive filtering was applied to remove samples with a low mapping quality, as well as strand and position biases and oxoguanine oxidative artifacts. Residual germline mutations from the Genome Aggregation Database (https://gnomad.broadinstitute.org/) and Online Archive of Brazilian Mutations (http://abraom.ib.usp.br/) were removed. The raw sequencing data (fastq files) were deposited in the Sequence Read Archive (http://www.ncbi.nlm.nih.gov/sra; accession no. PRJNA505810).

4.3. Statistical Analyses

Data on baseline patient characteristics are expressed as absolute and relative frequencies for qualitative variables and as means $\pm$ standard deviations for quantitative variables. Mutational signature exposure and the TMB were considered to be continuous variables. Associations between qualitative variables were evaluated using the chi-squared test or Fisher's exact test, as appropriate. Comparisons of the means of quantitative variables and groups were evaluated using the t-test or Mann–Whitney U test, as appropriate.

Overall survival functions were generated using the Kaplan–Meier estimator and compared between S^{high} and S^{low} groups using the log-rank test. A semiparametric Cox proportional-hazards model was fitted to the dataset to describe relationships between OS and the main clinical features. HRs and 95% CIs were calculated for all variables. A backward stepwise selection algorithm was applied, with different significance levels required to enter (p = 0.10) and remain in (p = 0.05) the model. Variables that acted as confounders (>20% change in coefficient) were also removed from the model. The proportional-hazards assumption was assessed based on the Schoenfeld residuals [57]. The analysis provided evidence that all the covariates had constant effects over time.

Multivariate analyses were performed to examine the main clinical features associated previously with OS (e.g., age, pathological stage, Lauren tumor subtype, and ethnicity), mutational signature exposures associated with dMMR, and molecular features (TMB and MSI status). Forest plots were created based on the results of the final multiple Cox regression model. Patients with metastasis were excluded from these analyses. In addition, we fitted simple and multiple logistic regression models to assess the effects of S2, S4, and S5 exposure on *MLH1* methylation. The overall performance, calibration, and discriminatory power of the final multiple logistic regression model were assessed using the Brier score, the Hosmer–Lemeshow goodness-of-fit test, and the AUC, respectively [58]. We also assessed the goodness of fit using a Q–Q plot. The significance level was fixed at 5% for all the tests (two-sided). Statistical analyses were performed using the R software (v3.5).

4.4. Mutational Signatures in Cell Lines

CRISPR-Cas9 knockout clones for *MLH1* were generated in human HAP1 cells using the guide RNA sequence 5′-AAGACAATGGCACCGGGATC-3′. Clonal populations with *MLH1* frameshift mutations were cultured for 3 months to allow for the accumulation of mutations during cellular division [59]. To identify mutations, genomic DNA was submitted to whole-genome sequencing. De novo somatic mutations, including substitutions, indels, and rearrangements, in subclones were identified by removing all mutations seen in parental clones. Next, SNVs were mapped onto trinucleotide sequences by including the 5′ and 3′ neighboring base contexts, followed by the estimation of the degree of sample exposure to previously detected mutational signatures [21].

4.5. Mutational Signature Estimation

All six classes of SNV (C > A, C > G, C > T, T > A, T > C, and T > G) from the public cohorts were mapped onto trinucleotide sequences by including the 5′ and 3′ neighboring base contexts. Next, the SNV spectrum with 96 trinucleotide mutation types for all samples from the public cohorts was loaded into signer [21] for the estimation of the optimal number of mutational signatures based on the median Bayesian information criterion. We next used cosine similarity (>0.7) to compare the de novo extracted mutational signatures with those described in the CS database (v2).

The SNV spectrum with 96 trinucleotide mutation types for samples from the validation cohort was loaded into signeR to estimate the degree of sample exposure to previously identified mutational signatures from the public cohorts. Samples exhibiting greater exposure ($\geq$the third quartile level) to a given signature were assigned to the S^{high} groups and those with lesser exposure (<the third quartile level) were assigned to the S^{low} groups.

4.6. Molecular Features

We used the MSIseq [21] software for MSI status (MSI-H and non–MSI-H) prediction from whole-exome data. Briefly, this software is based on four machine-learning frameworks, and requires a catalog of sample somatic SNVs and microindels, a file containing the exact locations of mononucleotides (length $\geq$5 bases), and microsatellites consisting of di-, tri-, and tetranucleotide repeats, as annotated in the "simpleRepeats" track (http://hgdownload.cse.ucsc.edu/goldenpath/hg19/database). Consistent with the method proposed by Chalmers et al. [60], the TMB was calculated by dividing the total number of mutations by the length of the target region in megabases. Tumor heterogeneity was estimated using the "math.score" function in the "maftools" package (v3.8) [61]. Higher MATH scores indicate a greater heterogeneity. Available data on neoantigens for 77 TCGA-STAD samples were extracted (https://tcia.at/neoantigens) [62].

4.7. Identification of Significantly Mutated Genes

To assess the somatic mutation profiles of genes throughout the genome, we searched for genes that were mutated more frequently than expected by chance [27]. Pre-processed Mutation Annotation Format files (generated using the "prepareMutSig" function in

"maftools" (v8) [27]) were loaded into the online MutSigCV server (v1.3.4; https://cloud.genepattern.org/gp/pages/index.jsf) [27]. Significantly mutated genes associated with S4 were evaluated to determine their associations with $S4^{high}$, $S4^{low}$, and subgroups defined by *MLH1* methylation status using the "compareMaf" function ("maftools" R package). The "somaticInteractions" function ("maftools" R package) was used to identify sets of genes that were mutually mutated in exclusive and co-occurring manner [29,61].

4.8. Evaluation of Tumor Microenvironment Composition and Immunological Aspects

We estimated cellular compositions from the TCGA bulk expression datasets using two complementary approaches. First, the CIBERSORT software, based on the deconvolution method for the characterization of cell composition in complex tissues from their gene expression profiles, was used [50]. CIBESORT takes advantage of a validated leukocyte gene signature matrix, termed LM22. This matrix contains 547 genes that distinguish 22 human hematopoietic cell phenotypes, including seven T cell types, naïve and memory B cells, plasma cells, NK cells, and myeloid subsets. Simultaneously, a recently introduced GSEA-based technique termed xCell [51] was used to infer 64 immune and stroma cell types. We used this approach to confirm the CIBERSORT findings and evaluate the stromal content.

A CIBERSORT analysis was performed online using a public server (http://cibersort.stanford.edu/) for the characterization of absolute and relative immune cell composition, with 1000 permutations and disabled quantile normalization as set parameters. Of the 380 TCGA-STAD tumor samples, 215 (56%) samples yielded data on infiltrating immune cells (CIBERSORT analysis, $p < 0.05$) and were included in further analysis ($S4^{high}$, $n = 50$; $S4^{low}$, $n = 165$). We also used xCell to confirm the results of the $S4^{high}$ and $S4^{low}$ sample comparison. xCell analysis was performed using the R package with default parameters (https://github.com/dviraran/xCell). To verify the immune effector responses in $S4^{high}$ and $S4^{low}$ samples, the differential expression of key immunoregulatory/inflammatory and cytotoxic markers was found. Group comparisons described in this section were performed using the Mann–Whitney *U* test with the significance level set to $p < 0.05$. We also used the pre-processed immune subtypes described by Thorsson et al. [34] for TCGA samples (available for 103 $S4^{high}$ samples and 285 $S4^{low}$ samples; Table S2).

5. Conclusions

In conclusion, whereas previous studies have aimed to classify patients using molecular and clinical features, such as the MSI status, TMB, and *MLH1* gene expression levels, this study provides evidence that classification based on mutational signature exposure can be used to identify groups of patients with common clinical, immunological, and mutational features related directly to better prognosis.

Supplementary Materials: The following are available online at https://www.mdpi.com/2072-6694/13/3/490/s1, Figure S1. Heatmap clustering of gastric cancer sample signature exposure. Mutational signatures called by signeR are arranged in rows, and samples are arranged in columns, Figure S2. Forest plot of hazard ratios for overall survival from univariate Cox model analysis. TMB, tumor mutational burden; MSI, microsatellite instability, Figure S3. Forest plot of hazard ratios for overall survival from the multivariate Cox model for (a) mutational signature S2 and (b) S5, Figure S4. Calibrated plots of 2-year survival (multiple Cox regression models) for (a) mutational signature S2, (b) S4, (c) S5, (d) high microsatellite instability, and (e) tumor mutational burden, Figure S5. Five-year overall survival for $S4^{high}$ and $S4^{low}$ groups in the public cohort, Figure S6. Five-year overall survival for $S4^{high}$ and $S4^{low}$ groups in the validation cohort, Figure S7. Scatter plots of Sperman's correlation between MMR gene expression levels (FPKM) and methylation loads. Gene expression levels were plotted against average β values for the promoter-related CpG islands of MMR genes, Figure S8. Barplot showing the burdens of previously identified mutational signatures in the public gastric cancer cohort; isogenic wild-type cells are compared with two subclones with mutations in the *MLH1* gene induced by CRISP-Cas9 assay, Figure S9. Boxplots of immune and stroma cells scores, estimated by xCell (Aran et al., 2017), in the mutational signature $S4^{high}$ and $S4^{low}$ groups. * $p < 0.05$, ** $p < 0.01$, t test, Figure S10. Boxplots showing the normalized gene expression

of immune-related genes. Immune gene markers were grouped as (a) inflammatory/cytotoxic, (b) suppressor/exhausted, (c) costimulatory/antigen presentation, and (d) other (natural killer cell receptors and monocyte/macrophage markers). Outlying values were removed. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$, **** $p < 0.0001$, Mann–Whitney U test, Table S1. Significantly mutated genes in the mutational signature (S)4^{high} and S4^{low} groups and associated pathways. (a) Significantly mutated genes, related pathways, and frequency in the S4^{high} group. (b) Description of pathways found for the S4^{high} gene set. (c) Significantly mutated genes, related pathways and frequency in the S4^{low} group. (d) Description of pathways found for the S4^{low} gene set, Table S1 SignificantlyMutatedGenes-Pathways.xlsx, Table S2. Metadata on sample clinical, molecular, and immune features, Table S2 Clinical-Molecular-features.xlsx, Table S3. Clinicopathological features of gastric cancer samples by cohort, Table S4. Clinicopathological features of gastric cancer samples according to mutational signature 4 DNA mismatch repair deficiency in the validation cohort.

Author Contributions: I.T.d.S. designed and supervised the research. J.R.B. and M.N.P.S. performed the downstream bioinformatics analysis and interpretation of data. V.F.C., R.R., and R.D.D. performed the statistical analysis. T.F.B. contributed to generate whole exome data and analysis. A.D. and R.V. helped with the bioinformatics analysis and collected the data. K.J.G. helped to interpret the data from an immunological perspective. J.P.L., H.C.F., and V.C.C.d.L. helped to compile the clinical data. M.W. produced the cell line cultures. J.L., K.J.G., E.D.-N., D.N.N., and I.T.d.S. participated in the discussion of the data. J.R.B., D.N.N., and I.T.d.S. wrote the final version of the manuscript, on which all the authors provided comments. All authors have read and agreed to the published version of the manuscript.

Funding: Fundação de Amparo à Pesquisa do Estado de São Paulo (grant nos. 14/26897-0, 15/19324-6 and 16/11791-7).

Institutional Review Board Statement: The study was conducted according to the guidelines of the Declaration of Helsinki, and approved by the Institutional Review Board (or Ethics Committee) of A.C. Camargo Cancer Center (protocol code 2134/15—9 December 2015).

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

Data Availability Statement: The raw sequencing data (fastq files) are openly and avaliable in the Sequence Read Archive (http://www.ncbi.nlm.nih.gov/sra; accession no. PRJNA505810).

Acknowledgments: This project received financial support from the Fundação de Amparo à Pesquisa do Estado de São Paulo (nos. 14-26897-0, 15/19324-6 and 16/11791-7); E.D.-N. and K.J.G. are research fellows at the Conselho Nacional de Desenvolvimento Científico e Tecnológico. E.D.-N. acknowledges the support given by the Associação Beneficente Alzira Denise Hertzog Silva.

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

Abbreviations

CIN	chromosomal instability
dMMR	DNA mismatch repair deficiency
HRR	homologous recombination repair
GS	genomically stable
GSEA	gene set enrichment analysis
MSI	microsatellite instability
MSI-H	microsatellite instability high
MSI-L	microsatellite instability low
NER	nucleotide excision repair
OS	overall survival
SNV	single-nucleotide variation
TCGA	The Cancer Genome Atlas
TMB	tumor mutational burden

References

1. Stratton, M.R.; Campbell, P.J.; Futreal, P.A. The cancer genome. *Nature* **2009**, *458*, 719–724. [CrossRef]
2. Hakem, R. DNA-damage repair; the good, the bad, and the ugly. *EMBO J.* **2008**, *27*, 589–605. [CrossRef] [PubMed]
3. Helleday, T.; Eshtad, S.; Nik-Zainal, S. Mechanisms underlying mutational signatures in human cancers. *Nat. Rev. Genet.* **2014**, *15*, 585–598. [CrossRef] [PubMed]
4. Alexandrov, L.B.; Nik-Zainal, S.; Wedge, D.C.; Aparicio, S.A.J.R.; Behjati, S.; Biankin, A.V.; Bignell, G.R.; Bolli, N.; Borg, A.; Børresen-Dale, A.L.; et al. Signatures of mutational processes in human cancer. *Nature* **2013**, *500*, 415–421. [CrossRef] [PubMed]

5. Hollstein, M.; Alexandrov, L.B.; Wild, C.P.; Ardin, M.; Zavadil, J. Base changes in tumour DNA have the power to reveal the causes and evolution of cancer. *Oncogene* **2017**, *36*, 158–167. [CrossRef] [PubMed]
6. Bobrovnitchaia, I.; Valieris, R.; Drummond, R.D.; Lima, J.P.; Freitas, H.C.; Bartelli, T.F.; de Amorim, M.G.; Nunes, D.N.; Dias-Neto, E.; da Silva, I.T. APOBEC-mediated DNA alterations: A possible new mechanism of carcinogenesis in EBV-positive gastric cancer. *Int. J. Cancer* **2020**, *146*, 181–191. [CrossRef]
7. Van Hoeck, A.; Tjoonk, N.H.; Van Boxtel, R.; Cuppen, E. Portrait of a cancer: Mutational signature analyses for cancer diagnostics. *BMC Cancer* **2019**, *19*, 457. [CrossRef]
8. Davies, H.; Glodzik, D.; Morganella, S.; Yates, L.R.; Staaf, J.; Zou, X.; Ramakrishna, M.; Martin, S.; Boyault, S.; Sieuwerts, A.M.; et al. HRDetect is a predictor of BRCA1 and BRCA2 deficiency based on mutational signatures. *Nat. Med.* **2017**, *23*, 517–525. [CrossRef]
9. Kim, J.; Mouw, K.W.; Polak, P.; Braunstein, L.Z.; Kamburov, A.; Tiao, G.; Kwiatkowski, D.J.; Rosenberg, J.E.; Van Allen, E.M.; D'Andrea, A.D.; et al. Somatic ERCC2 mutations are associated with a distinct genomic signature in urothelial tumors. *Nat. Genet.* **2016**, *48*, 600–606. [CrossRef]
10. Jager, M.; Blokzijl, F.; Kuijk, E.; Bertl, J.; Vougioukalaki, M.; Janssen, R.; Besselink, N.; Boymans, S.; De Ligt, J.; Pedersen, J.S.; et al. Deficiency of nucleotide excision repair is associated with mutational signature observed in cancer. *Genome Res.* **2019**, *29*, 1067–1077. [CrossRef]
11. Thibodeau, M.L.; Zhao, E.Y.; Reisle, C.; Ch'ng, C.; Wong, H.L.; Shen, Y.; Jones, M.R.; Lim, H.J.; Young, S.; Cremin, C.; et al. Base excision repair deficiency signatures implicate germline and somatic MUTYH aberrations in pancreatic ductal adenocarcinoma and breast cancer oncogenesis. *Cold Spring Harb. Mol. Case Stud.* **2019**, *5*, a003681. [CrossRef] [PubMed]
12. Giráldez, M.D.; Balaguer, F.; Bujanda, L.; Cuatrecasas, M.; Muñoz, J.; Alonso-Espinaco, V.; Larzabal, M.; Petit, A.; Gonzalo, V.; Ocaña, T.; et al. MSH6 and MUTYH deficiency is a frequent event in early-onset colorectal cancer. *Clin. Cancer Res.* **2010**, *16*, 5402–5413. [CrossRef] [PubMed]
13. Volkov, N.M.; Yanus, G.A.; Ivantsov, A.O.; Moiseenko, F.V.; Matorina, O.G.; Bizin, I.V.; Moiseyenko, V.M.; Imyanitov, E.N. Efficacy of immune checkpoint blockade in MUTYH-associated hereditary colorectal cancer. *Investig. New Drugs* **2019**, *38*, 894–898. [CrossRef]
14. Mandal, R.; Samstein, R.M.; Lee, K.W.; Havel, J.J.; Wang, H.; Krishna, C.; Sabio, E.Y.; Makarov, V.; Kuo, F.; Blecua, P.; et al. Genetic diversity of tumors with mismatch repair deficiency influences anti-PD-1 immunotherapy response. *Science* **2019**, *364*, 485–491. [CrossRef] [PubMed]
15. Le, D.T.; Durham, J.N.; Smith, K.N.; Wang, H.; Bartlett, B.R.; Aulakh, L.K.; Lu, S.; Kemberling, H.; Wilt, C.; Luber, B.S.; et al. Mismatch repair deficiency predicts response of solid tumors to PD-1 blockade. *Science* **2017**, *357*, 409–413. [CrossRef] [PubMed]
16. Abida, W.; Cheng, M.L.; Armenia, J.; Middha, S.; Autio, K.A.; Vargas, H.A.; Rathkopf, D.; Morris, M.J.; Danila, D.C.; Slovin, S.F.; et al. Analysis of the Prevalence of Microsatellite Instability in Prostate Cancer and Response to Immune Checkpoint Blockade. *JAMA Oncol.* **2019**, *5*, 471–478. [CrossRef] [PubMed]
17. Reichert, Z.R.; Urrutia, J.; Alumkal, J.J. Microsatellite Instability as an Emerging Biomarker for Checkpoint Inhibitor Response in Advanced Prostate Cancer. *JAMA Oncol.* **2019**, *5*, 478–479. [CrossRef]
18. Svrcek, M.; Lascols, O.; Cohen, R.; Collura, A.; Jonchère, V.; Fléjou, J.F.; Buhard, O.; Duval, A. MSI/MMR-deficient tumor diagnosis: Which standard for screening and for diagnosis? Diagnostic modalities for the colon and other sites: Differences between tumors. *Bull. Cancer* **2019**, *106*, 119–128. [CrossRef]
19. Alexandrov, L.B.; Nik-Zainal, S.; Siu, H.C.; Leung, S.Y.; Stratton, M.R. A mutational signature in gastric cancer suggests therapeutic strategies. *Nat. Commun.* **2015**, *6*, 8683. [CrossRef]
20. Bartelli, T.F.; Senda de Abrantes, L.L.; Freitas, H.C.; Thomas, A.M.; Silva, J.M.; Albuquerque, G.E.; Araújo, L.F.; Branco, G.P.; de Amorim, M.G.; Serpa, M.S.; et al. Genomics and epidemiology for gastric adenocarcinomas (GE4GAC): A Brazilian initiative to study gastric cancer. *Appl. Cancer Res.* **2019**, *39*, 12. [CrossRef]
21. Rosales, R.A.; Drummond, R.D.; Valieris, R.; Dias-Neto, E.; Da Silva, I.T. signeR: An empirical Bayesian approach to mutational signature discovery. *Bioinformatics* **2017**, *33*, 8–16. [CrossRef] [PubMed]
22. Büttner, R.; Longshore, J.W.; López-Ríos, F.; Merkelbach-Bruse, S.; Normanno, N.; Rouleau, E.; Penault-Llorca, F. Implementing TMB measurement in clinical practice: Considerations on assay requirements. *ESMO Open* **2019**, *4*, e000442. [CrossRef] [PubMed]
23. Cammerer, G.; Formentini, A.; Karletshofer, M.; Henne-Bruns, D.; Kornmann, M. Evaluation of Important Prognostic Clinical and Pathological Factors in Gastric Cancer. *Anticancer Res.* **2012**, *32*, 1839–1842. [PubMed]
24. Rajput, A.; Bocklage, T.; Greenbaum, A.; Lee, J.H.; Ness, S.A. Mutant-Allele Tumor Heterogeneity Scores Correlate with Risk of Metastases in Colon Cancer. *Clin. Colorectal Cancer* **2017**, *16*, e165–e170. [CrossRef]
25. Leung, S.Y.; Yuen, S.T.; Chung, L.P.; Chu, K.M.; Chan, A.S.Y.; Ho, J.C.I. hMLH1 Promoter Methylation and Lack of hMLH1 Expression in Sporadic Gastric Carcinomas with High-Frequency Microsatellite Instability. *Cancer Res.* **1999**, *57*, 3370–3374.
26. Hu, W.; Yang, Y.; Ge, W.; Zheng, S. Deciphering molecular properties of hypermutated gastrointestinal cancer. *J. Cell. Mol. Med.* **2019**, *23*, 370–379. [CrossRef]
27. Lawrence, M.S.; Stojanov, P.; Polak, P.; Kryukov, G.V.; Cibulskis, K.; Sivachenko, A.; Carter, S.L.; Stewart, C.; Mermel, C.H.; Roberts, S.A.; et al. Mutational heterogeneity in cancer and the search for new cancer-associated genes. *Nature* **2013**, *499*, 214–218. [CrossRef]

28. Bass, A.J.; Thorsson, V.; Shmulevich, I.; Reynolds, S.M.; Miller, M.; Bernard, B.; Hinoue, T.; Laird, P.W.; Curtis, C.; Shen, H.; et al. Comprehensive molecular characterization of gastric adenocarcinoma. *Nature* **2014**, *513*, 202–209.
29. Yeang, C.-H.; McCormick, F.; Levine, A. Combinatorial patterns of somatic gene mutations in cancer. *FASEB J.* **2008**, *22*, 2605–2622. [CrossRef]
30. Fuertes Marraco, S.A.; Neubert, N.J.; Verdeil, G.; Speiser, D.E. Inhibitory receptors beyond T cell exhaustion. *Front. Immunol.* **2015**, *6*, 310. [CrossRef]
31. Landskron, G.; Fuente MD la Thuwajit, P.; Thuwajit, C.; Hermoso, M.A. Chronic Inflammation and Cytokines in the Tumor Microenvironment. *J. Immunol. Res.* **2014**, *2014*, 149185. [CrossRef] [PubMed]
32. Bastid, J.; Cottalorda-Regairaz, A.; Alberici, G.; Bonnefoy, N.; Eliaou, J.F.; Bensussan, A. ENTPD1/CD39 is a promising therapeutic target in oncology. *Oncogene* **2013**, *32*, 1743–1751. [CrossRef] [PubMed]
33. Binnewies, M.; Roberts, E.W.; Kersten, K.; Chan, V.; Fearon, D.F.; Merad, M.; Coussens, L.M.; Gabrilovich, D.I.; Ostrand-Rosenberg, S.; Hedrick, C.C.; et al. Understanding the tumor immune microenvironment (TIME) for effective therapy. *Nat. Med.* **2018**, *24*, 541–550. [CrossRef] [PubMed]
34. Thorsson, V.; Gibbs, D.L.; Brown, S.D.; Wolf, D.; Bortone, D.S.; Ou Yang, T.-H.; Porta-Pardo, E.; Gao, G.F.; Plaisier, C.L.; Eddy, J.A.; et al. The Immune Landscape of Cancer. *Immunity* **2018**, *48*, 812–830.e14. [CrossRef] [PubMed]
35. Ribeiro, A.L.; Okamoto, O.K. Combined effects of pericytes in the tumor microenvironment. *Stem Cells Int.* **2015**, *2015*, 868475. [CrossRef] [PubMed]
36. Bremnes, R.M.; Dønnem, T.; Al-Saad, S.; Al-Shibli, K.; Andersen, S.; Sirera, R.; Camps, C.; Marinez, I.; Busund, L.T. The role of tumor stroma in cancer progression and prognosis: Emphasis on carcinoma-associated fibroblasts and non-small cell lung cancer. *J. Thorac. Oncol.* **2011**, *6*, 209–217. [CrossRef] [PubMed]
37. Garza Treviño, E.N.; González, P.D.; Valencia Salgado, C.I.; Martinez Garza, A. Effects of pericytes and colon cancer stem cells in the tumor microenvironment. *Cancer Cell Int.* **2019**, *19*, 173. [CrossRef] [PubMed]
38. Drost, J.; van Boxtel, R.; Blokzijl, F.; Mizutani, T.; Sasaki, N.; Sasselli, V.; de Ligt, J.; Behjati, S.; Grolleman, J.E.; van Wezel, T.; et al. Use of CRISPR-modified human stem cell organoids to study the origin of mutational signatures in cancer. *Science* **2017**, *358*, 234–238. [CrossRef]
39. Hu, W.; Yang, Y.; Qi, L.; Chen, J.; Ge, W.; Zheng, S. Subtyping of microsatellite instability-high colorectal cancer. *Cell Commun. Signal.* **2019**, *17*, 79. [CrossRef]
40. Goodman, A.M.; Kato, S.; Bazhenova, L.; Patel, S.P.; Frampton, G.M.; Miller, V.; Stephens, P.J.; Daniels, G.A.; Kurzrock, R. Tumor mutational burden as an independent predictor of response to immunotherapy in diverse cancers. *Mol. Cancer Ther.* **2017**, *16*, 2598–2608. [CrossRef]
41. Mishima, S.; Kawazoe, A.; Nakamura, Y.; Sasaki, A.; Kotani, D.; Kuboki, Y.; Bando, H.; Kojima, T.; Doi, T.; Ohtsu, A.; et al. Clinicopathological and molecular features of responders to nivolumab for patients with advanced gastric cancer. *J. Immunother. Cancer* **2019**, *7*, 24. [CrossRef] [PubMed]
42. Samstein, R.M.; Lee, C.H.; Shoushtari, A.N.; Hellmann, M.D.; Shen, R.; Janjigian, Y.Y.; Barron, D.A.; Zehir, A.; Jordan, E.J.; Omuro, A.; et al. Tumor mutational load predicts survival after immunotherapy across multiple cancer types. *Nat. Genet.* **2019**, *51*, 202–206. [CrossRef] [PubMed]
43. Jia, Q.; Wu, W.; Wang, Y.; Alexander, P.B.; Sun, C.; Gong, Z.; Cheng, J.N.; Sun, H.; Guan, Y.; Xia, X.; et al. Local mutational diversity drives intratumoral immune heterogeneity in non-small cell lung cancer. *Nat. Commun.* **2018**, *9*, 5361. [CrossRef] [PubMed]
44. Stanta, G.; Bonin, S. Overview on clinical relevance of intra-tumor heterogeneity. *Front. Med.* **2018**, *5*, 85. [CrossRef] [PubMed]
45. Bolhaqueiro, A.C.F.; Ponsioen, B.; Bakker, B.; Klaasen, S.J.; Kucukkose, E.; van Jaarsveld, R.H.; Vivié, J.; Verlaan-Klink, I.; Hami, N.; Spierings, D.C.J.; et al. Ongoing chromosomal instability and karyotype evolution in human colorectal cancer organoids. *Nat. Genet.* **2019**, *51*, 824–834. [CrossRef] [PubMed]
46. Raynaud, F.; Mina, M.; Tavernari, D.; Ciriello, G. Pan-cancer inference of intra-tumor heterogeneity reveals associations with different forms of genomic instability. *PLoS Genet.* **2018**, *14*, e1007669. [CrossRef]
47. Pietrantonio, F.; Miceli, R.; Raimondi, A.; Kim, Y.W.; Kang, W.K.; Langley, R.E.; Choi, Y.Y.; Kim, K.M.; Nankivell, M.G.; Morano, F.; et al. Individual patient data meta-analysis of the value of microsatellite instability as a biomarker in gastric cancer. *J. Clin. Oncol.* **2019**, *37*, 3392–3400. [CrossRef] [PubMed]
48. Fridman, W.H.; Zitvogel, L.; Sautès-Fridman, C.; Kroemer, G. The immune contexture in cancer prognosis and treatment. *Nat. Rev. Clin. Oncol.* **2017**, *14*, 717–734. [CrossRef]
49. Wang, M.; Busuttil, R.A.; Pattison, S.; Neeson, P.J.; Boussioutas, A. Immunological battlefield in gastric cancer and role of immunotherapies. *World J. Gastroenterol.* **2016**, *22*, 6373–6384. [CrossRef]
50. Newman, A.M.; Liu, C.L.; Green, M.R.; Gentles, A.J.; Feng, W.; Xu, Y.; Hoang, C.D.; Diehn, M.; Alizadeh, A.A. Robust enumeration of cell subsets from tissue expression profiles. *Nat. Methods* **2015**, *12*, 453–457. [CrossRef]
51. Aran, D.; Hu, Z.; Butte, A.J. xCell: Digitally portraying the tissue cellular heterogeneity landscape. *Genome Biol.* **2017**, *18*, 220. [CrossRef] [PubMed]
52. Wang, K.; Kan, J.; Yuen, S.T.; Shi, S.T.; Chu, K.M.; Law, S.; Chan, T.L.; Kan, Z.; Chan, A.S.; Tsui, W.Y.; et al. Exome sequencing identifies frequent mutation of ARID1A in molecular subtypes of gastric cancer. *Nat. Genet.* **2011**, *43*, 1219–1223. [CrossRef] [PubMed]

53. Chen, K.; Yang, D.; Li, X.; Sun, B.; Song, F.; Cao, W.; Brat, D.J.; Gao, Z.; Li, H.; Liang, H.; et al. Mutational landscape of gastric adenocarcinoma in Chinese: Implications for prognosis and therapy. *Proc. Natl. Acad. Sci. USA* **2015**, *112*, 1107–1112. [CrossRef] [PubMed]
54. Kakiuchi, M.; Nishizawa, T.; Ueda, H.; Gotoh, K.; Tanaka, A.; Hayashi, A.; Yamamoto, S.; Tatsuno, K.; Katoh, H.; Watanabe, Y.; et al. Recurrent gain-of-function mutations of RHOA in diffuse-type gastric carcinoma. *Nat. Genet.* **2014**, *46*, 583–587. [CrossRef] [PubMed]
55. Ellrott, K.; Bailey, M.H.; Saksena, G.; Covington, K.R.; Kandoth, C.; Stewart, C.; Hess, J.; Ma, S.; Chiotti, K.E.; McLellan, M.; et al. Scalable Open Science Approach for Mutation Calling of Tumor Exomes Using Multiple Genomic Pipelines. *Cell Syst.* **2018**, *6*, 271–281. [CrossRef]
56. Depristo, M.A.; Banks, E.; Poplin, R.; Garimella, K.V.; Maguire, J.R.; Hartl, C.; Philippakis, A.A.; Del Angel, G.; Rivas, M.A.; Hanna, M.; et al. A framework for variation discovery and genotyping using next-generation DNA sequencing data. *Nat. Genet.* **2011**, *43*, 491–501. [CrossRef]
57. Schoenfeld, D. Partial Residuals for the Proportonal Hazards Regression Model. *Biometrika* **1982**, *69*, 239. [CrossRef]
58. Hosmer, D.W.; Lemeshow, S.; Sturdivant, R.X. *Applied Logistic Regression*, 3rd ed.; Wiley: Hoboken, NJ, USA, 2013; 528p.
59. Zou, X.; Owusu, M.; Harris, R.; Jackson, S.P.; Loizou, J.I.; Nik-Zainal, S. Validating the concept of mutational signatures with isogenic cell models. *Nat. Commun* **2018**, *9*, 1–16. [CrossRef]
60. Chalmers, Z.R.; Connelly, C.F.; Fabrizio, D.; Gay, L.; Ali, S.M.; Ennis, R.; Schrock, A.; Campbell, B.; Shlien, A.; Chmielecki, J.; et al. Analysis of 100,000 human cancer genomes reveals the landscape of tumor mutational burden. *Genome Med.* **2017**, *9*, 1744. [CrossRef]
61. Charoentong, P.; Finotello, F.; Angelova, M.; Mayer, C.; Efremova, M.; Rieder, D.; Hackl, H.; Trajanoski, Z. Pan-cancer Immunogenomic Analyses Reveal Genotype-Immunophenotype Relationships and Predictors of Response to Checkpoint Blockade. *Cell Rep.* **2017**, *18*, 248–262. [CrossRef]
62. Mayakonda, A.; Lin, D.C.; Assenov, Y.; Plass, C.; Koeffler, H.P. Maftools: Efficient and comprehensive analysis of somatic variants in cancer. *Genome Res.* **2018**, *28*, 1747–1756. [CrossRef] [PubMed]