

IMPACTO DA OBESIDADE NA INSTABILIDADE GENÔMICA

GABRIELA STEFANIE S SANTOS

**Dissertação apresentada à Fundação Antônio
Prudente para a obtenção do título de Mes-
tre em Ciências.**

Área de concentração: Oncologia

Orientador: Dr. Israel Tojal da Silva

São Paulo

2021

FICHA CATALOGRÁFICA

Silva Santos, Gabriela Stefanie

Impacto da obesidade na instabilidade genômica / Gabriela Stefanie Silva Santos - São Paulo, 2021.
75f.

Dissertação (Mestrado)-Fundação Antônio Prudente.

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

Orientador: Dr. Israel Tojal da Silva

1. Carcinoma de esôfago. 2. Obesidade 3. Protocolo para detecção de assinaturas Mutacionais.

CDU 616

"If you can't explain it simply, you don't understand it well enough."

Albert Einstein

AGRADECIMENTOS

Agradeço ao meu orientador Israel, pela oportunidade, pela paciência, por todo apoio e pela dedicação a esse projeto.

Agradeço ao Rodrigo Drummond por toda ajuda, ideias e discussões.

Agradeço ao Alexandre, pela paciência em ensinar, por toda ajuda com os imprevistos e pelos cafés.

Agradeço ao João Paulo pelas discussões e sugestões que foram cruciais para o andamento desse projeto.

Agradeço ao pessoal do LBCB, Renan, Jaqueline, Bruno, Rodrigo Borges e Felipe por todas discussões e aprendizado.

Agradeço ao Dr Emmanuel Dias Neto e à Dra Thaís Fernanda Bartelli por gentilmente nos fornecerem amostras necessárias para o desenvolvimento desse projeto.

Agradeço à CAPES por financiar esse projeto de pesquisa.

Agradeço aos professores e demais profissionais do A.C. Camargo Cancer Center por todo ensinamento, por se esforçarem à adaptação das atividades de pós-graduação durante a pandemia, pela estrutura, recursos e pelos serviços prestados aos alunos.

Agradeço aos voluntários do A.C. Camargo Cancer Center e também do TCGA por gentilmente fornecerem suas amostras tumorais, pelo bem da ciência, tornando possível a realização desse projeto.

Agradeço à minha família por todo apoio e incentivo.

Agradeço à Natália por toda ajuda, dicas e pela oportunidade de colaboração.

Agradeço também ao Victor por não me deixar desanimar.

RESUMO

Santos GSS. **Impacto da obesidade na instabilidade genômica**. [Dissertação]. São Paulo; Fundação Antônio Prudente; 2021.

A obesidade é uma condição associada com o desenvolvimento de diversas doenças, incluindo o câncer. A incidência de sobrepeso e obesidade vem aumentando consideravelmente no mundo, sendo atualmente uma grande preocupação de saúde pública. A obesidade induz um estado inflamatório crônico sistêmico, que subsequentemente leva à produção desregulada de espécies reativas de oxigênio e nitrogênio pelas células do sistema imunológico, causando um desequilíbrio em relação à capacidade antioxidante do organismo. Essas espécies reativas interagem com o DNA e geram modificações oxidativas, podendo afetar mecanismos de reparo do DNA e escapar da ação de enzimas antioxidantes levando ao acúmulo de danos oxidativos, instabilidade genômica e, consequentemente, a tumorigênese.

Com o objetivo de identificar mecanismos relacionados à obesidade e instabilidade genômica associados à tumorigênese, analisamos neste trabalho dados de expressão gênica, metilação e mutações somáticas de 14 tipos tumorais de pacientes do consórcio *The Cancer Genome Atlas* (TCGA) e uma coorte com 28 pacientes com adenocarcinoma gástrico tratados no A.C. Camargo Cancer Center.

Como resultado, observamos uma alteração no perfil do infiltrado inflamatório, incluindo genes marcadores de inflamação e citotoxicidade. De importância, identificamos ainda que alterações epigenéticas via hipo-metilação do promotor de alguns genes inflamatórios levaram a desregulação dos respectivos níveis de expressão.

Também foram identificadas duas assinaturas mutacionais causadas por espécies reativas de oxigênio (SBS17a e SBS17b) que, por sua vez, mostraram-se associadas ao sobrepeso. Por meio de uma regressão de Cox multivariada, mostramos que a atividade mutacional é um biomarcador prognóstico superior ao estadiamento clínico do tumor.

Palavras-chave: Obesidade, inflamação, ROS, câncer, bioinformática.

ABSTRACT

Santos GSS. [**Impact of obesity on genomic instability**]. [Dissertation]. São Paulo; Fundação Antonio Prudente; 2021.

Obesity is a condition associated with the development of several diseases, including cancer. The incidence of overweight and obesity has increased considerably in the world and is currently a major public health concern. Obesity induces a chronic systemic inflammatory state, which subsequently leads to unregulated production of reactive oxygen and nitrogen species by the cells of the immune system, causing an imbalance in the antioxidant capacity of the body. These reactive species interact with DNA and generate oxidative modifications that can affect DNA repair mechanisms and escape the action of antioxidant enzymes, leading to the accumulation of oxidative damage, genomic instability, and consequently tumorigenesis.

In order to identify mechanisms related to obesity and genomic instability associated with tumorigenesis, we analyzed in this work gene expression, methylation and somatic mutation data from 14 tumor types of patients from The Cancer Genome Atlas consortium, as well as mutation data from 28 patients of the AC Camargo Cancer Center with gastric adenocarcinoma.

As a result, we could observe an alteration in the inflammatory infiltrate profile including several inflammation and cytotoxicity marker genes. Importantly, we identified that epigenetic alterations via hypo-methylation of the promoters of some inflammatory genes led to dysregulation of their levels of expression.

We also identified two mutational signatures caused by reactive oxygen species (SBS17a and SBS17b) which, in turn, showed to be associated with overweight. Through a multivariate Cox regression, we have shown that the mutational activity is a prognosis biomarker superior to the clinical staging of the tumor.

Keywords: Obesity, inflammation, ROS, cancer, bioinformatics.

LISTA DE FIGURAS

Figura 1 – Proporção estimada de casos de câncer nos Estados Unidos associados ao estilo de vida	2
Figura 2 – Localizações anatômicas de depósitos de tecido adiposo em humanos	3
Figura 3 – Relação entre a infiltração de macrófagos e ROS em obesidade	6
Figura 4 – Assinatura mutacional SBS1 do COSMIC	8
Figura 5 – Assinatura mutacional DBS1 do COSMIC	9
Figura 6 – Assinatura mutacional ID1 do COSMIC	9
Figura 7 – <i>Workflow</i> com os macro temas propostos no presente trabalho.	13
Figura 8 – <i>Workflow</i> do <i>pipeline</i> de assinaturas mutacionais.	20
Figura 9 – Total de pacientes no TCGA com informação de peso e altura	24
Figura 10 – Correlação entre assinaturas mutacionais e o IMC dos pacientes do TCGA . . .	25
Figura 11 – Níveis de exposição dos pacientes do TCGA-ESCA à assinatura SBS17a do COSMIC v3.2.	26
Figura 12 – Níveis de exposição dos pacientes da coorte do TCGA-ESCA à assinatura SBS17b do COSMIC v3.2.	26
Figura 13 – Correlação entre as exposições à assinaturas obtidas por diferentes métodos de <i>fitting</i> de assinaturas	27
Figura 14 – Similaridade entre as assinaturas extraídas por todos os métodos <i>de novo</i> e todas assinaturas SBS descritas no COSMIC v3.2	28
Figura 15 – Assinatura mutacional S3 extraída <i>de novo</i> pelo método <i>signeR</i>	28
Figura 16 – Assinatura Mutacional SBS17b do COSMIC.	29
Figura 17 – Níveis de exposição dos pacientes do TCGA-ESCA à assinatura S3 extraída <i>de novo</i> pelo <i>signeR</i>	29
Figura 18 – Assinaturas mutacionais DBS extraídas pelo <i>sigProfiler</i>	30
Figura 19 – Assinaturas mutacionais ID extraídas pelo <i>sigProfiler</i>	31
Figura 20 – Correlação entre assinaturas SBS, DBS e ID extraídas pelo <i>sigProfiler</i> , e o IMC dos pacientes da coorte do TCGA-ESCA.	31

Figura 21 – Diferença na densidade mutacional entre as fitas transcrita e não transcrita dos pacientes com IMC normal, sobrepeso e obesidade da coorte do TCGA-ESCA.	33
Figura 22 – Infiltrado inflamatório em pacientes obesos, com sobrepeso e normais em TCGA-ESCA.	34
Figura 23 – Correlação entre os infiltrados inflamatórios e as variáveis BMI, exposição às assinaturas SBS17a e SBS17b.	35
Figura 24 – Expressão normalizada de genes marcadores pró-inflamatórios nos grupos normal, sobrepeso e obesidade da coorte do TCGA-ESCA.	36
Figura 25 – Correlação entre genes pró-inflamatórios e as variáveis BMI, exposição às assinaturas SBS17a e SBS17b	36
Figura 26 – Expressão normalizada de genes marcadores anti-inflamatórios nos grupos normal, sobrepeso e obesidade	37
Figura 27 – Correlação entre genes anti-inflamatórios e as variáveis BMI, exposição às assinaturas SBS17a e SBS17b.	38
Figura 28 – Expressão normalizada de genes marcadores da via BER nos grupos normal, sobrepeso e obesidade.	39
Figura 29 – Correlação entre genes da via BER e as variáveis BMI, exposição às assinaturas SBS17a e SBS17b.	40
Figura 30 – Expressão normalizada de genes marcadores da via NER nos grupos normal, sobrepeso e obesidade.	41
Figura 31 – Correlação entre genes da via NER e as variáveis BMI, exposição às assinaturas SBS17a e SBS17b.	41
Figura 32 – Expressão normalizada de genes marcadores da via MMR nos grupos normal, sobrepeso e obesidade.	42
Figura 33 – Correlação entre genes da via MMR e as variáveis BMI, exposição às assinaturas SBS17a e SBS17b.	43
Figura 34 – Perfil de expressão diferencial entre os grupos normal e obesidade.	44
Figura 35 – Clusterização das amostras de acordo com os grupos normal e obesidade.	45
Figura 36 – <i>Heatmap</i> dos DEGs.	45
Figura 37 – <i>Heatmap</i> do β -value dos DEGs.	46

Figura 38 – Distribuição dos β -values de pacientes obesos e normais.	47
Figura 39 – Comparação dos β -values entre pacientes obesos e normais.	47
Figura 40 – Vias alteradas em pacientes obesos a partir do enriquecimento de vias pelo GO de acordo com MF.	48
Figura 41 – Comparação dos β -values de genes inflamatórios entre pacientes obesos e nor- mais.	49
Figura 42 – Correlação de <i>Pearson</i> entre o β -value e a expressão normalizada do gene NOS2	50
Figura 43 – Correlação de <i>Pearson</i> entre o β -value do gene NOS2 e a exposição à assinatura SBS17b	51
Figura 44 – Curva ROC do modelo de regressão logística	52
Figura 45 – <i>Fold change</i> do gene NOS2 em outros tumores do TCGA.	53
Figura 46 – Média dos β -values do gene NOS2 em outros tumores do TCGA.	54
Figura 47 – Sobrevida global em até 10 anos para pacientes da coorte do TCGA-ESCA e IMC classificado como normal ou obesidade.	55
Figura 48 – Sobrevida global em até 10 anos para pacientes da coorte do TCGA-ESCA e níveis <i>high</i> e <i>low</i> de exposição à assinatura SBS17a.	55
Figura 49 – Sobrevida global em até 10 anos para pacientes da coorte do TCGA-ESCA e níveis <i>high</i> e <i>low</i> de exposição à assinatura SBS17b.	56
Figura 50 – Regressão de Cox múltipla para variáveis clínicas.	57
Figura 51 – Similaridade dos cossenos entre as assinaturas mutacionais extraídas por todos os métodos <i>de novo</i> e as assinaturas SBS descritas no COSMIC v3.2.	59
Figura 52 – Níveis de exposição dos pacientes do A.C. Camargo à assinatura S3	60
Figura 53 – Número de mutações por megabase atribuídas a cada assinatura mutacional onde a assinatura está presente.	62

LISTA DE TABELAS

Tabela 1 – Risco relativo (RR) de câncer <i>per</i> 5 Kg/m ² de aumento no IMC.	2
Tabela 2 – Classificação oficial de IMC definida pela OMS.	14
Tabela 3 – Regressão logística simples para a metilação do NOS2 e a assinatura SBS17b. OR, <i>Odds Ratio</i> (Razão de chances); CI, <i>Confidence Interval</i> (Intervalo de con- fiança).	51
Tabela 4 – Informações clínicas de pacientes com adenocarcinoma gástrico do A.C. Ca- margo Cancer Center.	58

LISTA DE SIGLAS E ABREVIATURAS

8-oxodG	<i>8-oxo-7,8-dihidro-2-desoxiguanosina</i>
APOBEC3	<i>Apolipoprotein B mRNA Editing Enzyme Catalytic Polypeptide-like</i>
BER	<i>Base Excision Repair</i>
BMI	<i>Body Mass Index</i>
CAT	<i>Catalase</i>
CI	<i>Confidence Interval</i>
COSMIC	<i>Catalogue of Somatic Mutations in Cancer</i>
DBS	<i>Doublet base substitutions</i>
DNA	<i>Deoxyribonucleic Acid</i>
EBV	<i>Epstein-Barr Virus</i>
FFA	<i>Free Fatty Acid</i>
FPKM	<i>Fragments per kilo base per million mapped reads</i>
GATK	<i>Genome Analysis Toolkit</i>
GGR	<i>Global Genomic Repair</i>
HR	<i>Homologous Recombination</i>
IBGE	<i>Instituto Brasileiro de Geografia e Estatística</i>
ID	<i>Insertions or Deletion Mutation</i>
IMC	<i>Índice de Massa Corporal</i>
JNK	<i>c-Jun N-terminal kinase</i>
KG	<i>Kilograma</i>

NER	<i>Nucleotide Excision Repair</i>
NFKB	<i>Nuclear Factor Kappa Beta</i>
NHEJ	<i>Non-homologous end joining</i>
NK	<i>Natural Killer</i>
NMF	<i>Non-negative Matrix Factorization</i>
NO	<i>Nitric Oxide</i>
MAF	<i>Mutation Annotation Format</i>
MMR	<i>Mismatch Repair</i>
OS	<i>Oxidative Stress</i>
PNS	<i>Pesquisa Nacional de Saúde</i>
PON	<i>Panel of Normals</i>
RNS	<i>Reactive Nitrogen Species</i>
ROS	<i>Reactive Oxygen Species</i>
RAM	<i>Random Access Memory</i>
RNA-Seq	<i>RNA Sequencing</i>
RR	<i>Relative Risk</i>
SBS	<i>Single base substitutions</i>
SNV	<i>Single Nucleotide Variants</i>
SOD	<i>Superoxide dismutase</i>
TB	<i>Terabase</i>
TCGA	<i>The Cancer Genome Atlas</i>
TCR	<i>Transcription Coupled Repair</i>

Tg	<i>Timina-glicol</i>
TIL	<i>Tumor-Infiltrating Lymphocytes</i>
TLS	<i>Translesion synthesis</i>
TMB	<i>Tumor Mutation Burden</i>
TME	<i>Tumor microenvironment</i>
TPM	<i>Transcripts per Million</i>
Tregs	<i>Regulatory T Cells</i>
UV	<i>Ultravioleta</i>
VAF	<i>Variant Allelic Frequency</i>
VCF	<i>Variant Call Format</i>
Vigitel	<i>Vigilância de Fatores de Risco e Proteção para Doenças Crônicas por Inquérito Telefônico</i>
WAT	<i>White Adipose Tissue</i>
WES	<i>Whole Exome Sequencing</i>

LISTA DE SÍMBOLOS

α	Alpha
β	Beta
κ	Kappa
m^2	Metro Quadrado
H_2O_2	Peróxido de hidrogênio
$ONOO^-$	Peroxinitrito
O_2^-	Superóxido
OH^-	Hidroxila

SUMÁRIO

1	INTRODUÇÃO	1
1.1	Obesidade, uma epidemia crescente	1
1.2	Estresse oxidativo, inflamação e mecanismos de formação de radicais livres	3
1.3	Danos ao DNA e mecanismos de reparo	6
1.4	Assimetria transcricional	7
1.5	Assinaturas Mutacionais	7
1.6	Impacto do sobrepeso nos mecanismos de reparo do DNA	10
1.7	Regulação epigenética	10
1.8	Aspectos relevantes desse projeto	11
2	OBJETIVOS	12
2.1	Objetivo geral	12
2.2	Objetivos específicos	12
3	METODOLOGIA	13
3.1	Obtenção dos dados	13
3.1.1	Amostras públicas	13
3.1.2	Amostras do AC Camargo Cancer Center	16
3.2	Estrutura e <i>workflow</i> do <i>pipeline</i> de assinaturas mutacionais	17
3.2.1	Cálculo da assimetria transcricional	21
3.3	Análise de expressão diferencial	21
3.4	Análise do <i>status</i> de metilação	21
3.5	Enriquecimento de vias	22
3.6	Infiltrado inflamatório	22
3.7	Análises estatísticas	22
4	RESULTADOS	23
4.1	Obtenção dos dados públicos	23

4.2	Análise Mutacional	24
4.2.1	Assinaturas Mutacionais	24
4.3	Infiltrado inflamatório e genes marcadores de inflamação	33
4.4	Genes marcadores de reparo	38
4.5	Identificação dos genes diferencialmente expressos	44
4.6	Análise do <i>status</i> de metilação para DEGs	46
4.7	Enriquecimento de vias para DEGs	48
4.8	Análise do <i>status</i> de metilação para marcadores inflamatórios	49
4.9	Associação entre expressão gênica e metilação do gene NOS2, e a assinatura mutacional SBS17b	50
4.10	Expressão e metilação do gene NOS2 em outros tumores do TCGA	52
4.11	Sobrevida global	54
4.12	Amostras do AC Camargo Cancer Center	57
5	DISCUSSÃO	61
6	CONCLUSÕES	68
7	REFERÊNCIAS	69

APÊNDICES 77

Apêndice A – ANÁLISE DESCRITIVA DOS PACIENTES DA COORTE DO TCGA	78
--	-----------

1 INTRODUÇÃO

1.1 OBESIDADE, UMA EPIDEMIA CRESCENTE

A incidência de sobrepeso e obesidade, definidos em adultos como Índice de Massa Corporal (IMC) ≥ 25 e IMC ≥ 30 respectivamente, aumentou consideravelmente na população mundial durante as últimas décadas. Entre 1980 e 2008, a média do IMC mundial aumentou $0,4 \text{ Kg/m}^2$ por década entre os homens e $0,5 \text{ kg/m}^2$ entre as mulheres (Finucane et al. 2011). Desde 1975 os casos de obesidade no mundo quase triplicaram, tornando-se uma grande preocupação de saúde pública e, em 2016, mais de 1,9 bilhão de adultos (a partir de 18 anos) estavam acima do peso, dos quais mais de 650 milhões eram obesos (World Health Organization 2020).

No Brasil, segundo dados da pesquisa Vigilância de Fatores de Risco e Proteção para Doenças Crônicas por Inquérito Telefônico (Vigitel)¹ foi demonstrado que mais da metade da população (55,7%) apresenta excesso de peso, um aumento de quase 40% quando comparado com o percentual de 42,6% no ano de 2006. Também foi registrado um aumento no índice de obesidade nos últimos treze anos, saindo de 11,8% em 2006 para 19,8% em 2018. O crescimento da obesidade foi maior entre os adultos de 25 a 34 anos (84,2%) e 35 a 44 anos (81,1%) (Ministério da Saúde 2018).

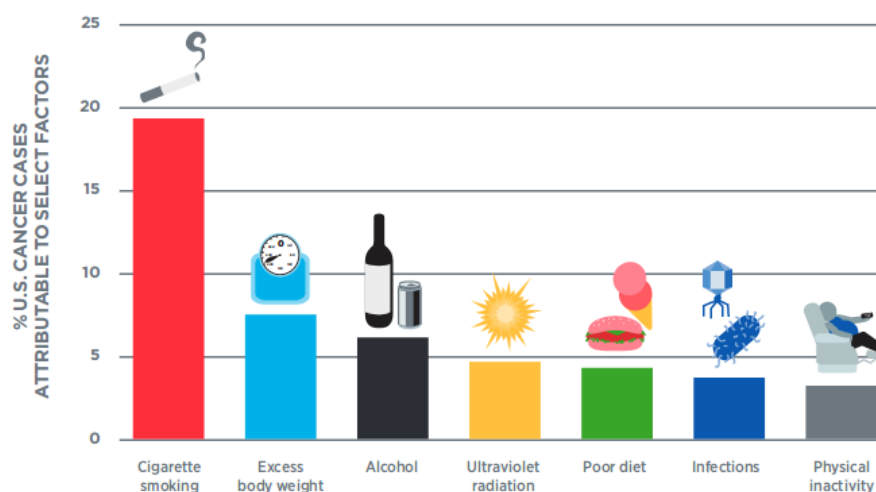
Além disso, de acordo com dados da Pesquisa Nacional de Saúde (PNS), realizada pelo Instituto Brasileiro de Geografia e Estatística (IBGE) em parceria com o Ministério da Saúde em 2019, o percentual de pessoas obesas atingiu a marca de 26,9% (cerca de 41,2 milhões de pessoas), e o de pessoas com sobrepeso 61,7%, representando quase dois terços da população brasileira adulta (IBGE 2020).

Encontra-se bem documentado que diversos tumores (**Tabela 1**) estão associados ao excesso de peso corporal (Renehan et al. 2008; Wolin et al. 2010; Chen et al. 2012; Collaborative Group on Epidemiological Studies of Ovarian Cancer 2012; Castillo et al. 2014; Goodwin e Stambolic 2015; Avgerinos et al. 2019) e representa a maior causa de incidência de câncer na população dos Estados Unidos (**Figura 1**), logo atrás do tabagismo (American Association for Cancer Research 2019).

¹ Pesquisa realizada pelo Ministério da Saúde em 2018

Tabela 1 – Risco relativo (RR) de câncer *per* 5 Kg/m² de aumento no IMC.

Tipo de câncer	Homens		Mulheres	
	RR	95% CI	RR	95% CI
Cardia gástrica	1,82	(1,32 - 2,49)	1,82	(1,32 - 2,49)
Adenocarcinoma de esôfago	1,52	(1,33 - 1,74)	1,51	(1,3 - 1,74)
Vesícula Biliar	–	–	1,59	(1,02 - 2,47)
Endométrio	–	–	1,59	(1,50 - 1,68)
Linfoma difuso de grandes células B	1,34	(1,11 - 1,60)	1,23	(1,09 - 1,39)
Tireoide	1,33	(1,04 - 1,70)	1,14	(1,06 - 1,23)
Reto	1,30	(1,25–1,35)	1,12	(1,07 - 1,18)
Renal	1,24	(1,15 - 1,34)	1,34	(1,25 - 1,43)
Cólon	1,24	(1,20 - 1,28)	1,09	(1,05 - 1,13)
Fígado	1,19	(1,09 - 1,29)	1,10	(1,04 - 1,16)
Melanoma maligno	1,17	(1,05 - 1,30)	–	–
Pâncreas	1,14	(1,01 - 1,29)	1,13	(1,06 - 1,21)
Mama (pós-menopausa)	–	–	1,12	(1,08 - 1,16)
Mieloma múltiplo	1,11	(1,05 - 1,18)	1,11	(1,07 - 1,15)
Ovário	–	–	1,10	(1,07 - 1,13)

**Figura 1** – Proporção estimada de casos de câncer nos Estados Unidos associados ao estilo de vida. As barras fixadas no eixo horizontal representam os fatores de risco associados à tumorigênese e no eixo vertical a porcentagem dos casos de câncer associados com cada fator de risco. Adaptado de (American Association for Cancer Research 2019).

Em particular, o Brasil conta com uma fração de 3,4% de todos os casos de câncer mundialmente atribuídos ao sobrepeso corporal (International Agency for Research On Cancer 2012). Estimou-se que em 2012 o sobrepeso corporal foi responsável por cerca de 4% de todos os casos de câncer diagnosticados no Brasil. Além disso, é esperado que em 2025 o número de casos de câncer associados ao excesso de peso corporal dobre em relação ao de 2012 (Rezende et al. 2018).

1.2 ESTRESSE OXIDATIVO, INFLAMAÇÃO E MECANISMOS DE FORMAÇÃO DE RADICAIS LIVRES

Na maioria dos mamíferos existem dois tipos principais de tecido adiposo: (i) Tecido adiposo branco (WAT, *White Adipose Tissue*), rico em lipídios e responsável por armazenar energia, e o tecido adiposo marrom (BAT, *Brown Adipose Tissue*), rico em mitocôndrias e responsável pelo gasto energético, bem como regulação da termogênese. O WAT é distribuído por diferentes partes anatômicas, sendo que os maiores depósitos de WAT são subcutâneos e viscerais (na cavidade abdominal, e entre os órgãos) e também são encontrados depósitos menores em volta de veias e artérias e até dentro da medula óssea (Quail e Dannenberg 2019). Já o BAT é encontrado em menor proporção em áreas altamente vascularizadas, principalmente entre os omoplatas, pescoço e área supraclavicular e ao longo da medula espinhal (Townsend e Tseng 2012) (**Figura 2**).

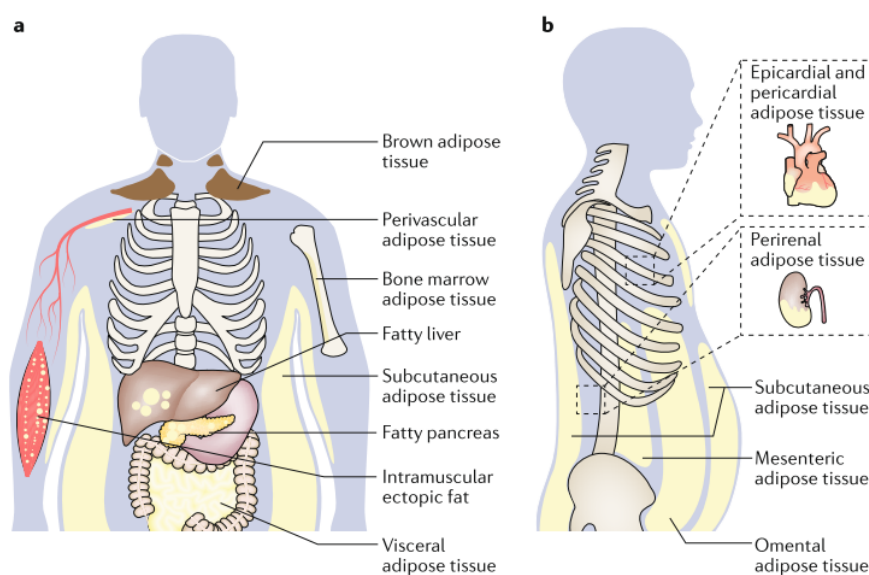


Figura 2 – Localizações anatômicas de depósitos de tecido adiposo em humanos. Adaptado de (Quail e Dannenberg 2019).

Em homeostase metabólica, *i. e.* peso corporal saudável, o microambiente do tecido adiposo (ATME, *Adipose Tissue Microenvironment*) é bem vascularizado e rico em citocinas anti-inflamatórias, incluindo IL-4, IL-10 e IL-13. Consequentemente, também tem abundância de células do sistema imunológico com perfil anti-inflamatório como macrófagos M2, células linfoides inatas, células T *helper* e eosinófilos (Quail e Dannenberg 2019).

A sobrecarga metabólica causada pelo excesso de nutrientes, condição comum em obesi-

dade, aumenta a circulação de ácidos graxos livres (FFA, *free fatty acids*) e o depósito de energia no WAT, resultando em um desequilíbrio energético, bem como hipertrofia, hiperplasia e remodelação do tecido adiposo, levando à hipóxia do tecido e necrose celular (Usman e Volpi 2018). Esses processos desregulam a função normal dos adipócitos, e levam à infiltração de células do sistema imunológico inato, promovendo uma resposta pró-inflamatória.

É bem conhecido que a desregulação dos adipócitos causa alterações na produção de adipocitocinas como TNF- α (*Tumor Necrosis Factor Alpha*), IL-6 (*Interleukin 6*) e IL-1 (*Interleukin 1*), promovendo o recrutamento de células do sistema imunológico marcadoras de inflamação, como neutrófilos, macrófagos, células dendríticas (Ottobelli Chielle et al. 2016) e o acúmulo de células B (Quail e Dannenberg 2019). O aumento de TNF- α é um marcador de inflamação sistêmica (Fernández-Sánchez et al. 2011) e sua sinalização prolonga a sobrevivência e proliferação de células cancerosas pela ativação de NF- κ B (*Nuclear Factor Kappa Beta*) e JNK (*c-Jun N-terminal kinase*) (Karin e Lin 2002). É descrito que sua atividade pode aumentar os danos ao DNA, assim induzindo aumento da carga mutacional (Yan et al. 2006).

A infiltração de macrófagos (principalmente M1), característica da obesidade, pode secretar altas concentrações de óxido nítrico (NO, *Nitric Oxide*), que pode levar à oxidação do DNA (Tamir et al. 1996). Além disso, em indivíduos obesos é observada uma diminuição de células anti-inflamatórias como células T regulatórias (Tregs), bem como células NK (*Natural Killer*) (Font-Burgada et al. 2016), e células T CD8 citotóxicas (Nishimura et al. 2009), reforçando o estado pró-inflamatório associado com a obesidade.

A infiltração de células imune pró-inflamatórias para o ATME causa mais produção de citocinas pró-inflamatórias, favorecendo o recrutamento de mais células imune marcadoras de inflamação, que por sua vez, estimulam a produção de espécies reativas de oxigênio (ROS, *Reactive Oxygen Species*) e nitrogênio (RNS, *Reactive Nitrogen Species*) por meio de reações enzimáticas das oxidases da família NADPH, como NOX1, NOX2 e NOX4 (Mittal et al. 2014), em um ciclo vicioso, caracterizado como inflamação crônica. Tal fator é conhecido por causar danos ao DNA (**Figura 3**) (Surmi e Hasty 2010; Cerdá et al. 2014). Sendo assim, a obesidade está associada com uma atividade imunológica que resulta em dano ao DNA, fator que induz o crescimento do tumor devido à ausência de imunovigilância, que poderia retardar o crescimento de tumores em estágio inicial (Font-Burgada et al. 2016).

O estresse oxidativo (OS, *oxidative stress*) é descrito como um desequilíbrio dos níveis de ROS/RNS em relação à capacidade antioxidante do corpo, resultando no acúmulo de produtos oxidativos (Azzi 2007; Codoñer-Franch et al. 2011).

ROS incluem radicais livres como superóxido (O_2^-), radicais hidroxila (OH^-), e peróxido de hidrogênio (H_2O_2), que são os radicais livres mais tóxicos dos sistemas biológicos (Roberts et al. 2010b). Já as espécies RNS incluem radicais como peroxinitrito ($ONOO^-$), formado pela reação do superóxido com NO (Gomes et al. 2008).

As condições fisiológicas normais requerem a disponibilidade e uma produção rigorosamente regulada de ROS e RNS. A principal fonte celular da produção dessas espécies reativas é a cadeia mitocondrial transportadora de elétrons, como resultado da transferência incompleta de elétrons para o oxigênio, fator que pode ocorrer mesmo em condições fisiológicas normais (Gao et al. 2010). Além de aumentar o consumo de oxigênio pelas mitocôndrias, causando alteração na cadeia respiratória mitocondrial, a obesidade pode induzir OS também por outros mecanismos, incluindo a oxidação mitocondrial de FFA e a infiltração de células do sistema imunológico (Fernández-Sánchez et al. 2011).

Além desses fatores, com o aumento do tecido adiposo, há diminuição da disponibilidade de enzimas antioxidantes como superóxido dismutase (SOD), catalase (CAT) e glutathione peroxidase (GPx) (Fernández-Sánchez et al. 2011), responsáveis por reduzir a ação de ROS e RNS, causando uma perturbação no sistema redox das células (Usman e Volpi 2018; Korac et al. 2021). Como resultado, as espécies ROS/RNS interagem com diversas macromoléculas, como fosfolípidios, carboidratos, proteínas e ácidos nucleicos, que se tornam modificados estruturalmente pelo processo de oxidação, produzindo modificações irreversíveis na célula (Cerdá et al. 2014; Regueira et al. 2008; Shang et al. 2006; Wang et al. 2010).

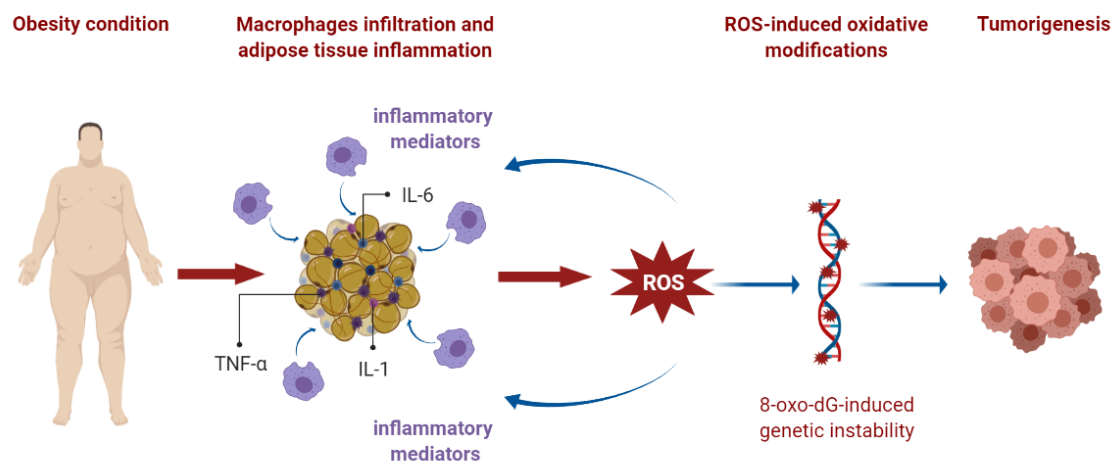


Figura 3 – Relação entre a infiltração de macrófagos e ROS em obesidade. Adaptado de (Cerdá et al. 2014; Roberts et al. 2010b).

1.3 DANOS AO DNA E MECANISMOS DE REPARO

Devido ao estresse oxidativo na célula, a interação de ROS com o DNA gera modificações oxidativas produzindo diferentes subprodutos com importante potencial mutagênico. Mais de 20 tipos diferentes de danos nas bases de DNA foram identificados após exposição ao estresse oxidativo (Slupphaug et al. 2003). O dano mais prevalente nas purinas é o 8-oxo-7,8-dihidro-2-desoxiguanosina (8-oxodG) e nas pirimidinas é a formação da timina-glicol (Tg). A base mutada 8-oxodG pode ter um papel crítico na ligação dos distúrbios metabólicos com a incidência de câncer (Barregard et al. 2013; Oltra et al. 2001; Ribeiro et al. 2008), baseada na habilidade de induzir erros durante a replicação do DNA (Shibutani et al. 1991).

O aumento da base mutagênica 8-oxodG, observada em indivíduos obesos, pode acelerar a taxa mutacional das células e também afetar vários mecanismos de reparo do DNA, levando à acumulação de lesões não reparadas que promovem a instabilidade genômica, subsequentemente criando condições favoráveis à tumorigênese (Roberts et al. 2010a; Lichtman 2010; Mostoslavsky et al. 2006). Esse efeito enfatiza a importância de um mecanismo de reparo de DNA eficiente para manter a estabilidade genômica (Klungland et al. 1999).

Para limitar os diferentes erros que ocorrem no DNA, há uma rede complexa de vias de reparo, incluindo reparo por excisão de base (BER, *base excision repair*), reparo de excisão de nucleotídeo (NER, *nucleotide excision repair*), reparo acoplado à transcrição (TCR, *transcription-coupled repair*), reparo global do genoma (GGR, *global genomic repair*), reparo do pareamento de

bases no DNA (MMR, *mismatch repair*), síntese translesão (TLS, *translesion synthesis*), recombinação homóloga (HR, *homologous recombination*) e junção de extremidades não homólogas (NHEJ, *Non-homologous end joining*). Particularmente, a via BER é a principal via de reparo de danos oxidativos nas bases, com as vias TCR e MMR envolvidas no reparo das fitas transcritas e novas fitas replicadas, respectivamente (Slupphaug et al. 2003; David et al. 2007; Russo et al. 2007).

1.4 ASSIMETRIA TRANSCRICIONAL

Os danos na fita de DNA transcrita potencialmente podem paralisar a progressão da RNA polimerase, levando ao recrutamento do complexo NER para sua correção (Fousteri e Mullenders 2008; Spivak e Ganesan 2014; Jiang e Sancar 2006; Hanawalt e Spivak 2008). É importante ressaltar que altos níveis de transcrição de um gene expõem o DNA a danos mas, por outro lado, essas regiões estão associadas a mais oportunidades de reparo acoplado à transcrição (TCR), levando a uma correlação inversa entre o nível de expressão de um gene e sua densidade mutacional (Pleasant et al. 2010; Lawrence et al. 2013; Chapman et al. 2011).

Inversamente, os danos na fita de DNA não transcrita podem não paralisar a progressão da RNA polimerase e, portanto, escapar do reparo pelo TCR. Além disso, a fita não transcrita permanece como fita simples durante o processo de transcrição e é, portanto, mais vulnerável a danos (Jinks-Robertson e Bhagwat 2014). Conjuntamente, esses mecanismos levam a diferenças nas densidades mutacionais das fitas transcrita e não transcrita (Pleasant et al. 2010), causando uma assimetria mutacional entre as fitas de DNA. Notavelmente, essa assimetria fornece informações sobre os danos no DNA e a ação do TCR (Haradhvala et al. 2016).

1.5 ASSINATURAS MUTACIONAIS

O câncer surge como resultado de alterações no DNA adquiridas ao longo da vida de um indivíduo, sendo que diversos mecanismos, incluindo aqueles relacionados ao estilo de vida, fatores genéticos, tais como falha nos mecanismos de reparo do DNA, dano *off-target* de enzimas de edição de ácidos nucleicos (APOBECs, *Apolipoprotein B mRNA Editing Enzyme Catalytic Polypeptide-like*), tabagismo, exposição à radiação ultravioleta (UV) e componentes da dieta. Es-

sas mutações envolvem substituições de base única (SBS, *Single base substitutions*), de bases duplas (DBS, *Doublet base substitutions*), e também inserções e deleções (ID, *Small insertions and deletions*), com base em seus respectivos contextos de nucleotídeos 5' e 3' imediatamente adjacentes à mutação, dando origem às assinaturas mutacionais (Hainaut e Pfeifer 2001; Brennan e Wild 2015).

A variação de base única (SNV, *Single Nucleotide Variant*) é o tipo de alteração mais comum no DNA das células tumorais (Stratton et al. 2009). Ao considerar que cada uma das SNVs podem ser mapeadas em seis tipos (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), esses conjuntos podem incluir as bases vizinhas 5' e 3' adjacentes ao nucleotídeo alterado. Sendo assim, um total de 96 combinações possíveis de mutações em um contexto de tri-nucleotídeos são geradas (Alexandrov et al. 2013b). Até o momento, foram descritas 72 assinaturas mutacionais de substituição de nucleotídeo único (SBS) no catálogo de mutações somáticas em câncer (COSMIC, *Catalogue of Somatic Mutations in Cancer*)² (Alexandrov et al. 2020). Exemplo de assinatura mutacional SBS: (**Figura 4**).

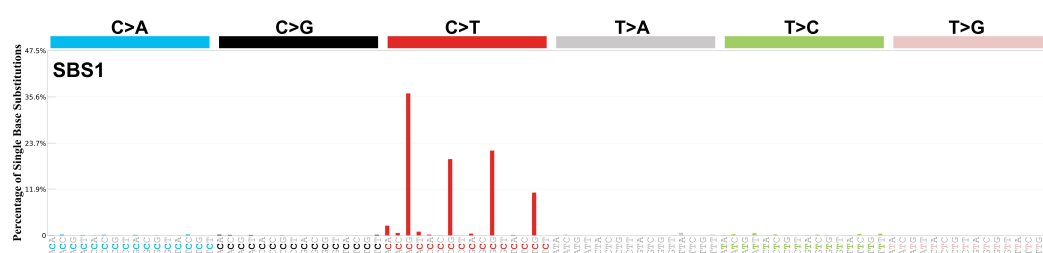


Figura 4 – Assinatura mutacional SBS1 do COSMIC. Sua etiologia é correlacionada com a idade do indivíduo (cosmic.sanger.ac.uk).

Também são identificadas em genomas de câncer, com uma prevalência relativamente baixa, as substituições de bases duplas (DBS), que são modificações simultâneas de dois nucleotídeos consecutivos, e podem ser mapeadas em 16 possíveis trocas. Um exemplo de uma DBS seria um conjunto de pares de base CT:GA mutando para o conjunto de pares de base AA:TT, denotado como CT:GA > AA:TT (Bergstrom et al. 2019). Incluindo as bases vizinhas 5' e 3', adjacentes ao dinucleotídeo alterado, existem 78 tipos de combinações possíveis de mutações DBS nesse contexto de nucleotídeos (Alexandrov et al. 2013a). Combinações únicas desses 78 tipos de mutações DBS num contexto de nucleotídeos 5' e 3' que flanqueiam imediatamente as 16 possíveis trocas ci-

² cosmic.sanger.ac.uk

tadas acima, dão origem a 11 assinaturas mutacionais DBS descritas até o momento no COSMIC. (Alexandrov et al. 2020). Exemplo de assinatura mutacional DBS: (**Figura 5**).

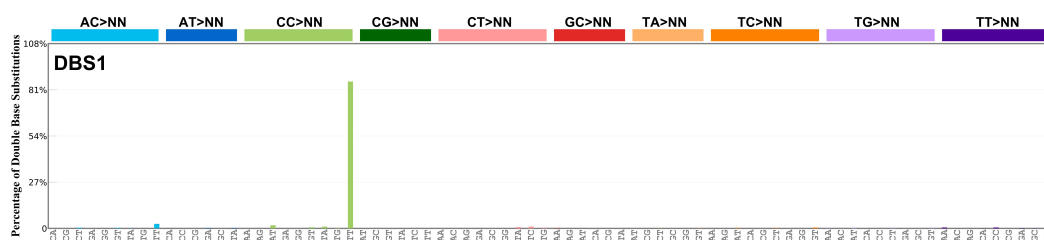


Figura 5 – Assinatura mutacional DBS1 do COSMIC causada por exposição à luz UV (cosmic.sanger.ac.uk).

Além disso, em localizações específicas nos genomas tumorais, são encontradas pequenas inserções e deleções de fragmentos de DNA (geralmente entre 1 e 50 pares de base), conhecidas como indels (ID). Um exemplo de mutação do tipo ID seria a sequência 5'-ATCCG-3' mutada para 5'-ATAAG-3', podendo ser caracterizada como uma deleção de CC:GG e uma inserção de AAA:TTT. A classificação de indels é uma tarefa complexa e não pode ser feita analogamente às classificações SBS e DBS, onde o contexto de nucleotídeos flanqueando imediatamente cada mutação é utilizado para classificar esses eventos mutacionais. Por exemplo, determinar as sequências de nucleotídeos vizinhos à uma deleção (ou inserção) de citosina da sequência 5'-ATCCCCCG-3' não é possível, já que não há como distinguir qual das citosinas foi deletada (Bergstrom et al. 2019). Embora não haja um conjunto bem definido de mutações ID, foi feita uma compilação de 83 tipos de mutações, considerando o tamanho da indel, nucleotídeos afetados e a presença de regiões repetitivas e/ou microhomológicas, e a partir disso, foram extraídas 18 assinaturas mutacionais do tipo ID até o momento, descritas no COSMIC (Alexandrov et al. 2020). Exemplo de assinatura mutacional ID: (**Figura 6**).

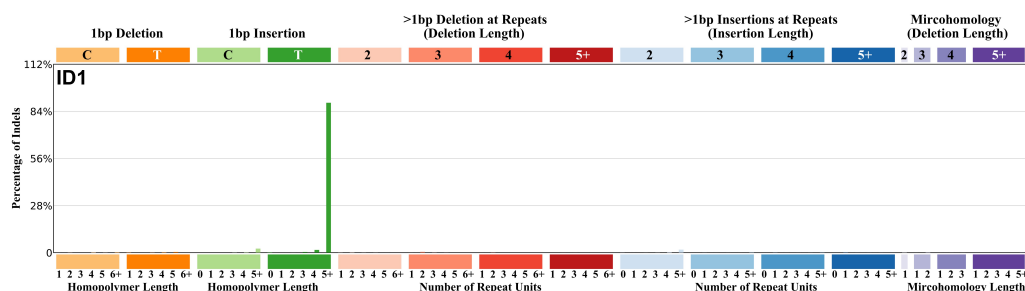


Figura 6 – Assinatura mutacional ID1 do COSMIC causada por falha do mecanismo de reparo de *mismatch* (cosmic.sanger.ac.uk).

Assim, o conhecimento a cerca desses processos pode auxiliar na compreensão da biologia

do câncer em um contexto evolutivo (Helleday et al. 2014). Dentro da linha pesquisa do nosso grupo, contribuímos com esse tema propondo uma abordagem original para a estimação estatística de assinaturas mutacionais com base em um tratamento Bayesiano empírico do modelo NMF (*non-negative matrix factorization*) (Rosales et al. 2017). Essa ferramenta, denominada *signeR*, encontra-se disponível no repositório Bioconductor³ que, em outros estudos do nosso grupo, foi utilizada para investigar danos colaterais no DNA resultante da ação dos genes APOBECs em tumores gástricos associados com infecção viral, especificamente EBV (*Epstein–Barr virus*) (Bobrovitch et al. 2019). E também para investigar assinaturas mutacionais associadas a um melhor prognóstico em tumor gástrico (Buttura et al. 2020).

1.6 IMPACTO DO SOBREPESO NOS MECANISMOS DE REPARO DO DNA

Atualmente encontra-se documentado uma associação inversa entre o IMC e o funcionamento da via NER (Tyson et al. 2009). Sabe-se ainda que defeitos nas vias BER e NER são associados ao aumento das taxas de incidência de câncer. SNVs em genes que codificam proteínas envolvidas nessas vias são responsáveis por diminuir as atividades de enzimas de reparo e aumentar o risco de câncer em humanos (Desaulniers et al. 2015).

O sobrepeso tem impacto em diferentes vias de reparo do DNA. A inflamação aumenta a atividade mitótica da célula, que resulta em erros na replicação do DNA e reduz a duração da fase de reparo. Além disso, vários mecanismos de reparo são afetados pela oxidação do DNA (Setayesh et al. 2018; Jaiswal et al. 2001; Graziewicz et al. 1996).

1.7 REGULAÇÃO EPIGENÉTICA

Foi demonstrado, em ratos, que dietas altamente calóricas podem alterar a disponibilidade de grupos metil, consequentemente levando à uma perturbação epigenética pela alteração dos padrões de metilação do DNA de forma global, impactando a expressão gênica (Zwamborn et al. 2017; Keleher et al. 2018). Dessa forma, há uma hipótese de que mecanismos epigenéticos podem estar envolvidos na regulação de genes que codificam proteínas de reparo do DNA, os quais podem ser impactados pela obesidade e sobrepeso, porém essa associação ainda não é totalmente compreendida (Włodarczyk e Nowicka 2019).

³ bioconductor.org/

Além disso, um estudo demonstrou que, quando os adipócitos são expostos às condições de altos níveis de glicose, fator também comum em organismos obesos e com sobrepeso, um conjunto de genes pró-inflamatórios é ativado devido à alteração dos padrões de metilação de seus promotores, sendo assim esses genes se tornam menos reprimidos, levando a um fenótipo inflamatório (Rønningen et al. 2015).

1.8 ASPECTOS RELEVANTES DESSE PROJETO

A obesidade, sobrepeso e doenças metabólicas associadas estão entre os fatores de risco mais relevantes para a saúde humana em todo o mundo. Entretanto, estudos mostram que a relação do sobrepeso e tumorigênese não está totalmente esclarecida (Setayesh et al. 2018). A literatura citada a respeito da obesidade e câncer são associações epidemiológicas e os aspectos moleculares foram observados, em sua maioria, em modelos animais. Nesse cenário, com a disponibilidade de informações moleculares e clínicas em bancos de dados públicos, abre-se uma oportunidade para explorar dados de pacientes oncológicos em larga escala.

2 OBJETIVOS

2.1 OBJETIVO GERAL

Identificar mecanismos subjacentes da relação entre a obesidade e instabilidade genômica associados à tumorigênese e seu impacto no prognóstico de pacientes oncológicos.

2.2 OBJETIVOS ESPECÍFICOS

- Recuperar informações clínicas e moleculares disponíveis publicamente na base de dados do consórcio O Atlas do Genoma do Câncer (TCGA, *The Cancer Genome Atlas*) e também do A.C. Camargo Cancer Center;
- Definir grupos de pacientes de acordo com a classificação do IMC;
- Detectar genes diferencialmente expressos e metilados entre os grupos;
- Estabelecer um *workflow* de análise de assinaturas mutacionais;
- Estimar a diversidade de células inflamatórias nas amostras;
- Detectar a atuação do mecanismo de reparo de DNA;
- Avaliar possíveis associações desses resultados com o prognóstico dos pacientes.

3 METODOLOGIA

3.1 OBTENÇÃO DOS DADOS

3.1.1 Amostras públicas

Neste trabalho, foram utilizadas amostras tumorais de pacientes com informações clínicas e moleculares, disponíveis na base de dados de domínio público do TCGA. Foram selecionados todos os tumores do TCGA cujos pacientes apresentavam informações de altura e peso. Para obtenção das informações moleculares, as amostras foram submetidas a sequenciamento de RNA (*RNA-Seq*, *RNA-Sequencing*), sequenciamento bissulfito (*Bisulfite-Seq*, *Bisulfite-Sequencing*) e sequenciamento de exoma completo (WES, *Whole Exome Sequencing*), que geraram dados de expressão gênica, *methylation array* e SNVs, respectivamente. A partir desses dados, foram realizadas análises de expressão diferencial, *status* de metilação, assinaturas mutacionais e sobrevida (Figura 7).

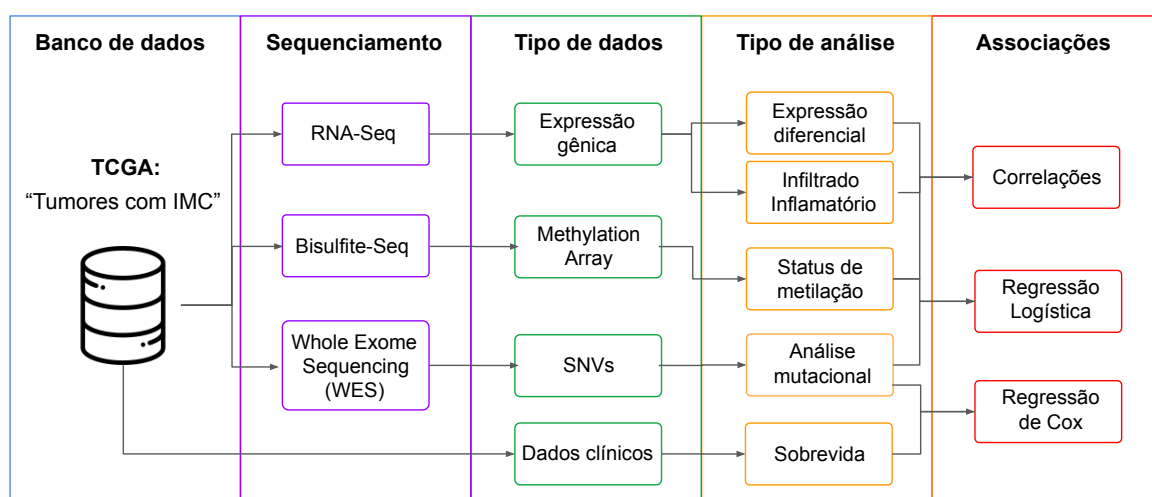


Figura 7 – *Workflow* com os macro temas propostos no presente trabalho.

3.1.1.1 Dados clínicos

A partir de dados clínicos de altura e peso, para cada paciente, foi realizado o cálculo do IMC (1) e, posteriormente, os pacientes foram associados aos respectivos grupos de acordo com a **Tabela 2**, definida pela OMS (World Health Organization 2020).

$$IMC = \frac{Peso(kg)}{Altura(m^2)} \quad (1)$$

Tabela 2 – Classificação oficial de IMC definida pela OMS.

IMC	Classificação
< 18,5	Abaixo do peso
≥ 18,5 e ≤ 24,9	Normal
≥ 25,0 e ≤ 29,9	Sobrepeso
≥ 30	Obesidade

3.1.1.2 Critérios de exclusão

Para alguns tipos de câncer a perda de peso é um dos principais sintomas, uma vez que decorrente do avanço da doença, alguns tumores cursam com disfagia e/ou caquexia tumoral (Baracos et al. 2018). Tais fatores podem levar à perda de peso pela própria dificuldade de ingerir alimentos ou também pela perda de massa muscular. Consequentemente, o uso do IMC ao diagnóstico pode levar a um viés em sua classificação, uma vez que não temos informação de perda de peso do paciente. Entretanto, é improvável que a perda de peso induzida pela doença em estágio inicial leve um indivíduo a uma nova classe de IMC (Quail e Dannenberg 2019). Dessa forma, propomos desconsiderar nas análises subsequentes pacientes com tumor em estágio avançado (III e IV), que justamente são os mais propensos à perda de peso decorrente do avanço do câncer.

Além disso, é conhecido que o tratamento com quimioterápicos pode induzir diversos tipos de mutações no DNA, fator que consequentemente pode levar ao desenvolvimento de assinaturas mutacionais específicas (Christensen et al. 2019), causando um viés na interpretação da possível etiologia das assinaturas mutacionais a serem detectadas nesse estudo. No entanto, os pacientes da coorte do TCGA não foram expostos à terapia neo-adjuvante (Liu et al. 2018).

3.1.1.3 Expressão gênica

Foram recuperados dados de expressão gênica, obtidos por *RNA-Seq*, no formato de contagem de *reads* por gene (TCGA, nível de acesso 2), e também expressão gênica normalizada por FPKM (*Fragments per kilo base Million*) (TCGA, nível de acesso 2).

3.1.1.4 Metilação

Também foram recuperados dados de metilação para todos os genes, obtidos a partir de *Bisulfite-Seq*, seguido pela estratégia de *Methylation Array*, quantificados por β -value, com valores entre 0-1, onde 0 representa "totalmente não metilado" e 1 "completamente metilado" (TCGA, nível de acesso 2).

3.1.1.5 Variantes Somáticas

Foi utilizado o banco de SNVs e INDEL somáticas produzidos pelo projeto TCGA - MC3 (*Multi-Center Mutation Calling in Multiple Cancers*). Essas variantes foram detectadas por meio de uma análise sistemática considerando 7 tipos de algoritmos utilizando dados de sequenciamento de exoma completo (WES, *whole-exome sequencing*) de cerca de 10,000 pares de amostras tumor/-normal, dos 33 tipos de câncer catalogados no TCGA (Ellrott et al. 2018). Ainda, no conjunto de variantes detectadas, os autores aplicaram alguns tipos de filtros para garantir a especificidade da chamada de variantes, os principais dentre eles:

- *Broad Panel of Normals* (Broad PON V2) - um dos filtros mais efetivos para detecção de falsos-positivos e variantes germinativas. Nesse filtro, assume-se que se uma variante foi detectada num conjunto de amostras normais (PoN), então é improvável que a variante seja somática em qualquer amostra tumoral, portanto a variante é considerada como germinativa e excluída;
- OxoG - Nesse filtro, a ferramenta DetOxoG é utilizada para marcar variantes 8-Oxoguanine (OxoG) que possivelmente podem ser artefatos devido à lesão OxoG ser comumente incorporada durante a preparação das bibliotecas de sequenciamento;
- StrandBias - Filtro aplicado para detectar artefatos devido à possíveis vieses de mutação específicas a uma das fitas do DNA;

- *Normal sequencing depth* - Para evitar a chamada de variantes germinativas, são filtradas apenas as variantes com o mínimo *depth coverage* de 8 *reads* na amostra normal em regiões não-dbSNP, e de pelo menos 19 *reads* em regiões dbSNP;
- NonExomic - Nesse filtro são selecionadas somente as variantes encontradas no exoma das amostras.
- Validação de mutação: se uma amostra normal tem uma fração alélica (FA, *allele-fraction*) > 0.2 para SNVs e > 0.1 para INDEL, a variante é classificada como germinativa. Caso não seja classificada como germinativa, e se os *read counts* do alelo alternativo do tumor for pelo menos 2, então a variante é classificada como somática.

3.1.2 Amostras do AC Camargo Cancer Center

Este estudo encontra-se associado ao projeto temático "**Epidemiologia e genômica de adenocarcinomas gástricos no Brasil**" desenvolvido pelo A.C. Camargo Cancer Center e apoiado pela FAPESP. Atualmente, para esse projeto, foram coletados dados clínicos de cerca de 1500 indivíduos de 3 capitais brasileiras, incluindo São Paulo, Fortaleza e Belém. Foram incluídos os pacientes diagnosticados com câncer, pacientes que realizaram exame endoscópico mas não têm diagnóstico de câncer e, finalmente, indivíduos que participaram da campanha de prevenção ao câncer do A.C. Camargo Cancer Center.

No presente estudo, foram selecionados apenas os pacientes diagnosticados com câncer e com variantes somáticas identificadas, obtidas a partir do sequenciamento de exoma (aprovado pelo Comitê de Ética em Pesquisa do A.C. Camargo Cancer Center sob nº 2134/15), totalizando 240 pacientes. Destes, também foram excluídos os pacientes com tumor em estadiamento avançado ($\geq III$), de forma semelhante à coorte do TCGA. Nos pacientes resultantes, o valor de IMC ao diagnóstico foi classificado em grupos conforme sugerido pela OMS (**Tabela 2**).

3.1.2.1 Variantes somáticas

Para os pacientes do A.C. Camargo Cancer Center, foi realizada a chamada de variantes somáticas para o exoma completo por meio de um *pipeline* desenvolvido *in house* baseado nas boas práticas sugeridas pelo Genome Analysis Toolkit (GATK) (McKenna et al. 2010). A partir

de arquivos binários (BAM, *Binary Alignment Map*), com informação do alinhamento das amostras de exoma no genoma de referência (GRCh38), foi realizada a chamada de variantes somáticas por meio do Software Mutect2 (Cibulskis et al. 2013) para as amostras tumorais. Em seguida, as variantes foram filtradas com um painel de 16 amostras de leucócitos não tumorais. Também, foram filtrados os *reads* com baixo *score* de alinhamento no genoma de referência, bem como artefatos induzidos por OxoG e vieses de mutação específicas à uma das fitas do DNA. Além disso, foram removidas potenciais variantes germinativas que estavam descritas no *Genome Aggregation Database* ¹ e no Arquivo Brasileiro Online de Mutações (ABraOM) ².

3.2 ESTRUTURA E WORKFLOW DO PIPELINE DE ASSINATURAS MUTACIONAIS

As assinaturas mutacionais fornecem padrões valiosos sobre processos mutacionais associados ao processo de tumorigênese. Uma análise de assinaturas pode ser realizada utilizando tanto uma abordagem de extração *de novo* ou uma abordagem de *fitting* baseada em assinaturas pré-selecionadas (e. g. as assinaturas descritas no COSMIC). Até o momento, foram desenvolvidas diversas ferramentas de bioinformática nas mais variadas linguagens de programação e plataformas para abordar esse tópico. Uma vez que a resolução desse problema biológico não tem uma solução única, desenvolvemos um *pipeline* para estimação e análise de assinaturas mutacionais que consiste na integração de várias dessas ferramentas, já descritas na literatura, com o objetivo de aumentar a acurácia na escolha de assinaturas e também de ampliar o conjunto de assinaturas mutacionais detectadas. Sua estrutura foi construída por meio do *framework* Snakemake (Köster e Rahmann 2012), um sistema de gerenciamento de *workflow*, permitindo análises de dados reproduzíveis e escalabilidade para ambientes de *cluster* e nuvem.

O *pipeline* é composto por 5 etapas (**Figura 8**) descritas abaixo:

Step 1 (Processamento dos arquivos de entrada): Os arquivos de entrada devem ser mutações descritas em arquivos do tipo *Variant Call Format* (vcf) acompanhados de arquivos do tipo *Browser Extensible Data* (bed). Nessa etapa, é utilizado o *framework* GATK (McKenna et al. 2010) para filtrar apenas as variantes com a tag "PASS", i. e. variantes que passaram por todos os

¹ <https://gnomad.broadinstitute.org/>

² <http://abraom.ib.usp.br/>

filtros especificados no processo anterior de chamada de variantes. Em seguida, todas as mutações do tipo SBS, DBS e ID são mapeadas em sequências incluindo o contexto de bases vizinhas 5' e 3' imediatamente adjacentes à mutação, determinado como *mutation counts*. Em seguida, o espectro com 96 tipos possíveis de mutações SBS, 78 DBS, e 83 ID em um determinado contexto de nucleotídeos, é utilizado nas etapas do *workflow* proposto, que foi dividido em duas abordagens:

(1) assinaturas mutacionais *de novo*

Essa abordagem baseia-se na extração de padrões recorrentes de variantes em um contexto de nucleotídeos obtidos a partir da matriz de *mutation counts*, permitindo a identificação tanto de processos mutacionais novos, quanto de processos mutacionais conhecidos (Maura et al. 2019). Para isso, foi utilizado o *software* EMu (Fischer et al. 2013), e foram desenvolvidos *scripts in house* nas linguagens R e Python, baseados nos pacotes SomaticSignatures (Gehring et al. 2015), signeR (Rosales et al. 2017), MutationalPatterns (Blokzijl et al. 2018) e SigProfiler (Bergstrom et al. 2019), organizados em etapas descritas a seguir:

Step 2.1 (*Opportunities*): A partir do arquivo no formato BED³, o pacote signeR é responsável pelo cálculo das frequências gênicas em que cada tipo de mutação, dentre as combinações possíveis de mutações inseridas em um contexto de tri-nucleotídeos (*i. e., motifs*), pode ocorrer, conhecido como *opportunities*.

Step 3.1 (Decomposição de assinaturas): Baseado na frequência da ocorrência dos *motifs* em múltiplas amostras, estes podem ser decompostos em assinaturas mutacionais (Gehring et al. 2015) por diferentes métodos:

1. SigProfiler: Método clássico proposto por (Alexandrov et al. 2013a) baseado no modelo de *nonnegative matrix factorization* (NMF)
2. MutationalPatterns: Método também baseado no modelo de NMF
3. signeR: método baseado em um tratamento Bayesiano empírico do modelo NMF.
4. EMu: método baseado no modelo *Expectation-Maximization* (EM).

³ <https://m.ensembl.org/info/website/upload/bed.html>

5. SomaticSignatures: método também baseado no modelo NMF, e conta com a opção de abordagem com o método *Principal Component Analysis* (PCA).

Uma vez que foram identificados os processos mutacionais, é necessário estimar o número de assinaturas que melhor descreve os dados. Porém, um parâmetro crítico do modelo NMF é definir o *rank* de fatoração (Blokzijl et al. 2018), ou seja, o número ótimo de assinaturas mutacionais. Em contrapartida, o pacote *signeR* possui uma função implícita, que considera o valor médio do Critério de Informação Bayesiano (BIC, *Bayesian Information Criterion*) para estimar o número ótimo de processos mutacionais, assim como o *software* EMu que infere o número de processos mutacionais baseado no algoritmo EM e o pacote *SigProfiler*, que determina o número ótimo de processos mutacionais baseado no valor que resulta numa decomposição de processos mutacionais com alta reprodutibilidade e baixo erro de reconstrução. Para os outros pacotes utilizados baseados no modelo NMF, foram determinados alguns métodos para definir o *rank* de fatoração:

1. *MutationalPatterns*: Para esse método, a abordagem utilizada foi a escolha do menor *rank* em que o coeficiente de correlação *cophenetic* começa a decrescer (Gaujoux e Seoighe 2010).
2. *SomaticSignatures*: Já para esse método, uma outra abordagem foi a escolha do *rank* com base no primeiro ponto em que o gráfico da soma residual dos quadrados (RSS, *Residual Sum of Squares*) entre a matriz de entrada e sua estimativa mostra um ponto de inflexão (Hutchins et al. 2008).

Step 4.1 Similaridade entre assinaturas: Nesse passo, as assinaturas extraídas são comparadas entre todos os métodos e com as assinaturas disponíveis no COSMIC pela similaridade do cosseno. Nessa métrica o resultado encontra-se no intervalo entre 0 e 1, onde 0 representa nenhuma similaridade e 1, alta similaridade.

(2) *Fitting* de assinaturas

Na segunda abordagem, os dados de entrada são ajustados à assinaturas mutacionais pré-estabelecidas, no caso as assinaturas mutacionais descritas no COSMIC, para que seja estimada a contribuição relativa e absoluta de cada assinatura para cada amostra (Maura et al. 2019).

Step 2.2 (Fitting de assinaturas conhecidas): Nesse passo, cada amostra é representada na forma da fração de mutações em cada um dos contextos de tri-nucleotídeos e então é identificada a combinação ponderada das assinaturas fornecidas que, quando somadas, reconstrói mais precisamente o perfil mutacional. Para essa finalidade, são utilizados os pacotes MutationalPatterns, sigProfiler e deconstructSigs (Rosenthal et al. 2016).

1. MutationalPatterns: Realiza o *fitting* de assinatura com base na resolução do problema de otimização de mínimos quadráticos não negativos (NNLS, *non-negative least-squares*)
2. deconstructSigs: A abordagem deste pacote determina a combinação linear de assinaturas que reconstrói o perfil mutacional mais precisamente utilizando um modelo de regressão linear múltipla.

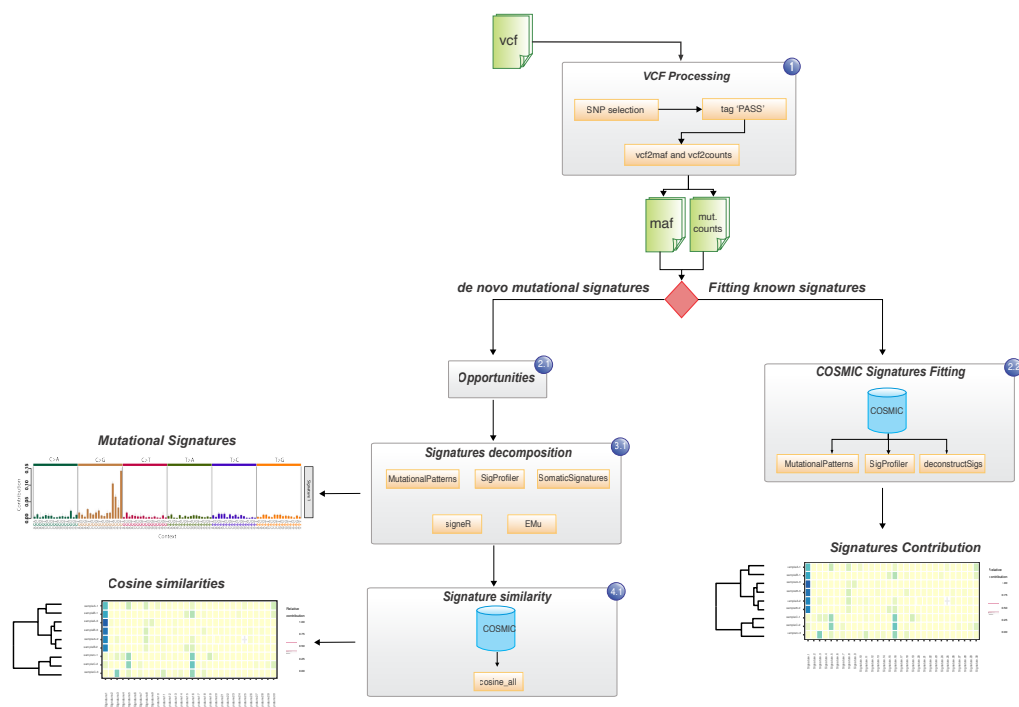


Figura 8 – Workflow do pipeline de assinaturas mutacionais.

3.2.1 Cálculo da assimetria transcricional

Para quantificar a assimetria transcricional entre as fitas de DNA dos pacientes, foi implementado um algoritmo adaptado do *workflow* do AsymTools (Haradhvala et al. 2016), onde a densidade mutacional para uma determinada troca de par de bases $b1:b2 \rightarrow m1:m2$ (ex: A:T $\rightarrow$ C:G) e sua mutação complementar $b2:b1 \rightarrow m2:m1$ (ex: T:A $\rightarrow$ G:C) em uma determinada região é dada por (2). Assim, a assimetria é calculada em uma determinada região por (3).

$$\begin{aligned} r_{b1:b2 \rightarrow m1:m2} &= \frac{n_{b1 \rightarrow m1}}{p * N_{b1}} \\ r_{b2:b1 \rightarrow m2:m1} &= \frac{n_{b2 \rightarrow m2}}{p * N_{b2}} \end{aligned} \quad (2) \quad a_{b \rightarrow m} = \log_2 \left(\frac{r_{b1:b2 \rightarrow m1:m2}}{r_{b2:b1 \rightarrow m2:m1}} \right) \quad (3)$$

onde,

$n_{b \rightarrow m}$: número de observações de mutações $b \rightarrow m$ em relação a fita de DNA de referência

N_b : número de chances dessa mutação acontecer, i.e. o número de ocorrências do *motif* b em uma dada região do genoma de referência

p : número de pacientes analisados.

3.3 ANÁLISE DE EXPRESSÃO DIFERENCIAL

Dados de expressão gênica (RNA-Seq) foram utilizados para avaliar os genes diferencialmente expressos entre as amostras classificadas como obesidade e as amostras normais. Essa etapa foi realizada por meio do pacote DESeq2 (Love et al. 2014). Foram selecionados como diferencialmente expressos os genes com $|\log_2 \text{FoldChange}| \geq 1$ e $p\text{-valor}$ ajustado ≤ 0.05 .

3.4 ANÁLISE DO STATUS DE METILAÇÃO

Foram selecionados apenas os $\beta\text{-values}$ das *probes* na região promotora dos genes, TSS (*Transcription Start Site*) ± 2000 bp. Para identificar os genes com perfil de metilação diferenciada entre os grupos de IMC, foram comparados os $\beta\text{-values}$ dos genes diferencialmente expressos entre os grupos por meio do teste não-paramétrico *Mann-Whitney*. Os genes com $p\text{-value} < 0.05$ e $\beta\text{-value} \geq 0.4$ foram considerados hiper-metilados.

3.5 ENRIQUECIMENTO DE VIAS

A partir do conjunto de genes diferencialmente expressos, foi realizado o enriquecimento de vias pelo *software* GSEA (Subramanian et al. 2005) para determinar diferenças estatisticamente significativas, *false discovery rate* (FDR) ≤ 0.25 , e concordantes entre os dois fenótipos.

3.6 INFILTRADO INFLAMATÓRIO

Os níveis de expressão gênica normalizados por FPKM (*Fragments per kilo base per million mapped reads*) foram utilizados no *software* CIBERSORT (Newman et al. 2015) para estimar a proporção dos 22 tipos de células inflamatórias presentes em cada amostra, incluindo 7 tipos de célula T, células B naïve e de memória, células plasmáticas, células NK (*natural killer*) e subconjuntos mieloides.

3.7 ANÁLISES ESTATÍSTICAS

Para as análises quantitativas, cálculos estatísticos e outros tipos de análise que utilizam grande quantidade de dados, foram desenvolvidos *scripts* nas linguagens de programação R e Python. Estas linguagens possuem repositórios, tais como *Bioconductor* (Huber et al. 2015), e *Bioconda* (Dale et al. 2018), com uma vasta quantidade de pacotes para análises de dados biológicos. As variáveis clínicas e moleculares foram correlacionadas utilizando os métodos *Spearman* ou *Pearson*, conforme apropriado, com o intuito de identificar possíveis associações que são recorrentes nos pacientes oncológicos com sobrepeso.

As médias das variáveis quantitativas de acordo com os diferentes grupos de IMC foram comparadas pelos testes não-paramétricos *Mann-Whitney*, quando comparados apenas dois grupos, e *Kruskal-Wallis* quando comparados três grupos.

O estimador *Kaplan-Meier* foi utilizado para análise de sobrevida, em conjunto com o teste *Log-Rank* para comparação entre as curvas. Também foi realizada regressão multivariada pelo método de Cox considerando variáveis clínicas e moleculares com possível impacto no prognóstico dos pacientes. Além disso, foi utilizado o teste ANOVA para comparar diferentes modelos de Cox com variáveis distintas, a fim de identificar as variáveis mais adequadas ao modelo final.

4 RESULTADOS

4.1 OBTENÇÃO DOS DADOS PÚBLICOS

Dos 33 tipos de tumores disponíveis na base de dados do TCGA, foram selecionados apenas aqueles que tinham informações clínicas de altura e peso, sendo que não são disponibilizadas informações pertinentes à altura, peso ou qualquer outra medida que possa ser utilizada para inferir a classificação dos pacientes pelo IMC antes do diagnóstico. Ao final, foram avaliados 14 tipos tumorais, incluindo 9 tumores associados ao sobrepeso e obesidade: i) carcinoma endometrial de corpo uterino (UCEC, n=500), ii) melanoma cutâneo de pele (SKCM, n=247), iii) carcinoma de células papilares renais (KIRP, n=232), iv) carcinoma de esôfago (ESCA, n=173), v) adenocarcinoma de cólon (COAD, n=222), vi) carcinoma hepatocelular de fígado (LIHC, n=360), vii) linfoma difuso de grandes células B (DLBC, n=37), viii) colangiocarcinoma (CHOL, n=35) e ix) adenocarcinoma de reto (READ, n=76) (Parsi 2013; Colditz e Peterson 2018; Scoppola et al. 2019). E não associados à obesidade, i) carcinoma urotelial de bexiga (BLCA, n=361), ii) carcinoma de células escamosas de colo de útero e adenocarcinoma endocervical (CESC, n=247), iii) timoma (THYM, n=100), iv) melanoma uveal (UVM, n=53), v) carcinosarcoma uterino (UCS, n=52).

A **Figura 9** ilustra o total de pacientes por tumor com informação de IMC, disponíveis na base de dados do TCGA, explorados no presente trabalho.

Foram investigados outros tipos de tumores que não são diretamente associados à obesidade, com informação de IMC, uma vez que desta forma é possível identificar padrões que são recorrentes nos tumores associados à obesidade e não em outros tipos de tumores.

A distribuição das características clínicas de pacientes de toda a coorte do TCGA avaliada nesse estudo, pertinentes a idade, gênero, classificação de IMC, estadiamento do tumor, entre outras, encontra-se no **Apêndice A**.

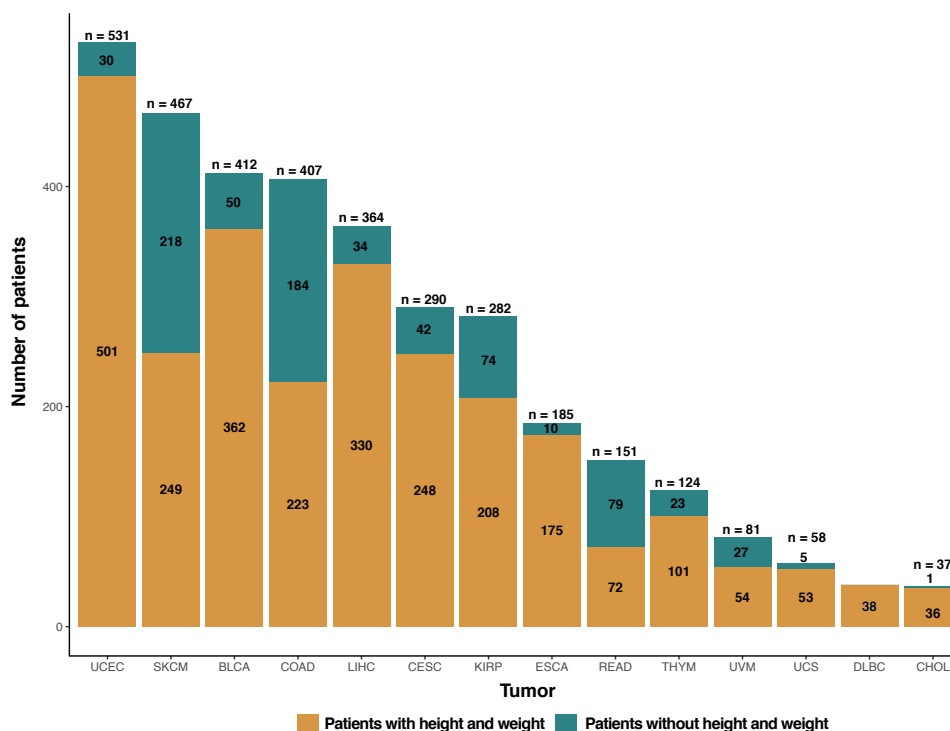


Figura 9 – Total de pacientes no TCGA com informação de peso e altura. No eixo x é representado o número de pacientes total em cada um dos tumores, e no eixo y é representado o número de pacientes.

4.2 ANÁLISE MUTACIONAL

4.2.1 Assinaturas Mutacionais

Foi pré-selecionado um subconjunto com 19 das assinaturas mutacionais descritas no COSMIC v3.1, algumas relacionadas ao estresse oxidativo, incluindo SBS17a, SBS17b, SBS18 e SBS36, outras relacionadas à mecanismos de reparo, incluindo SBS3, SBS6, SBS9, SBS10a, SBS10b, SBS14, SBS15, SBS16 SBS20, SBS21, SBS26, SBS30 e SBS44 e outras como SBS1, que é uma assinatura *clock-like* comum a todos tumores do TCGA, e SBS45 que sinaliza um artefato devido a introdução da base mutagênica 8-oxoG durante o sequenciamento.

A partir desse subconjunto de assinaturas, foi realizada uma avaliação ortogonal com o total de mutações de cada paciente, de cada um dos 14 tumores obtidos a partir do TCGA, utilizando os *softwares* *signeR*, *deconstructSigs* e *MutationalPatterns*. A partir dessa análise, é obtido um nível de presença, ou exposição, de cada uma das assinaturas de interesse nos pacientes. Ao final, foi realizada uma correlação de *Spearman* dos níveis de atividade das assinaturas e o IMC de cada paciente.

Utilizando como base o método signeR, identificamos correlações positivas e significativas entre o IMC dos pacientes com carcinoma de esôfago (ESCA, *esophagus carcinoma*) e as assinaturas SBS1 ($\rho = 0.45$, $p < 0.001$), SBS17a ($\rho = 0.34$, $p = 0.0013$), SBS17b ($\rho = 0.58$, $p < 0.001$) e SBS6 ($\rho = 0.23$, $p = 0.026$), e uma correlação negativa com a assinatura SBS16 ($\rho = -0.48$, $p < 0.001$) (**Figura 10**). Ainda, observa-se uma correlação positiva entre o IMC dos pacientes com linfoma difuso de grandes células B (DLBC, *Lymphoid Neoplasm Diffuse Large B-cell Lymphoma*) e a assinatura SBS17b ($\rho = 0.53$, $p = 0.01$).

Uma vez que o conjunto de amostras do TCGA-DLBC possui apenas 38 pacientes, e destes apenas 2 obesos, decidimos não prosseguir com essa coorte. Ainda, nos tumores restantes, uma vez que as assinaturas relacionadas a danos decorrentes de estresse oxidativo, especificamente SBS17a e SBS17b, mostraram correlação positiva apenas com o IMC dos pacientes do tumor TCGA-ESCA (**Figuras 11-12**), decidimos nos aprofundar nos mecanismos adjacentes a essas associações.

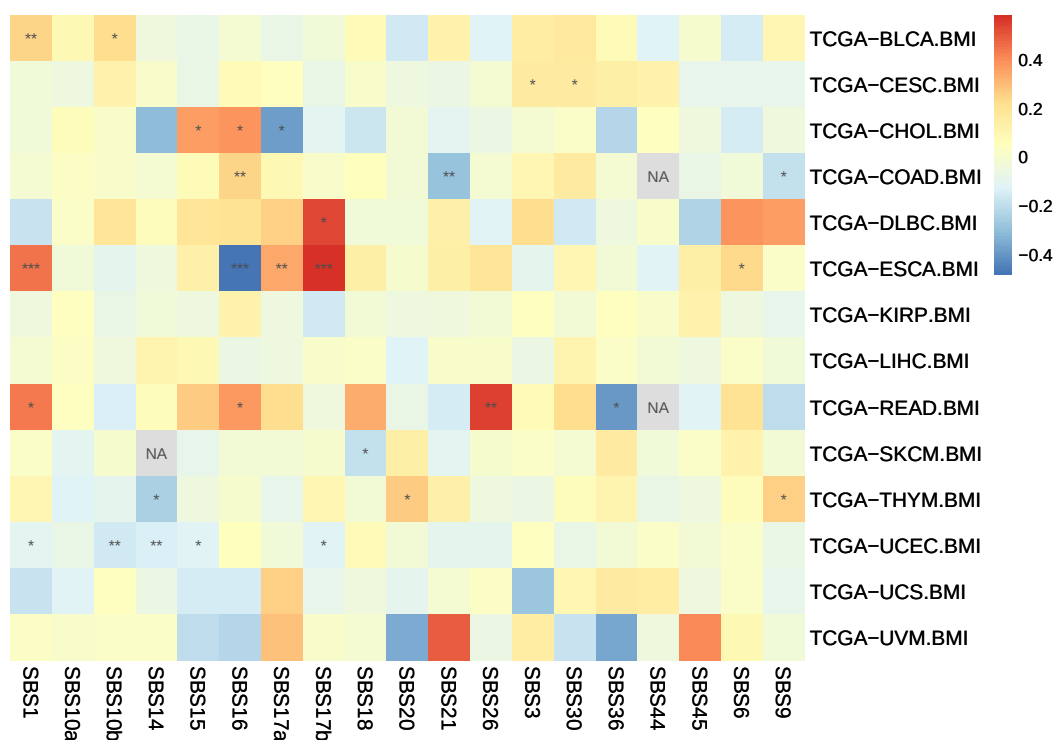


Figura 10 – Valor de correlação entre o subconjunto pré-definido de assinaturas mutacionais do COSMIC v3.1 relacionadas ao estresse oxidativo e a mecanismos de reparo (eixo x), e o IMC dos pacientes de 14 tumores do TCGA (eixo y). A escala de cor representa os valores de correlação. p -value < 0.05 : *; p -value < 0.01 : **; p -value < 0.001 : ***.

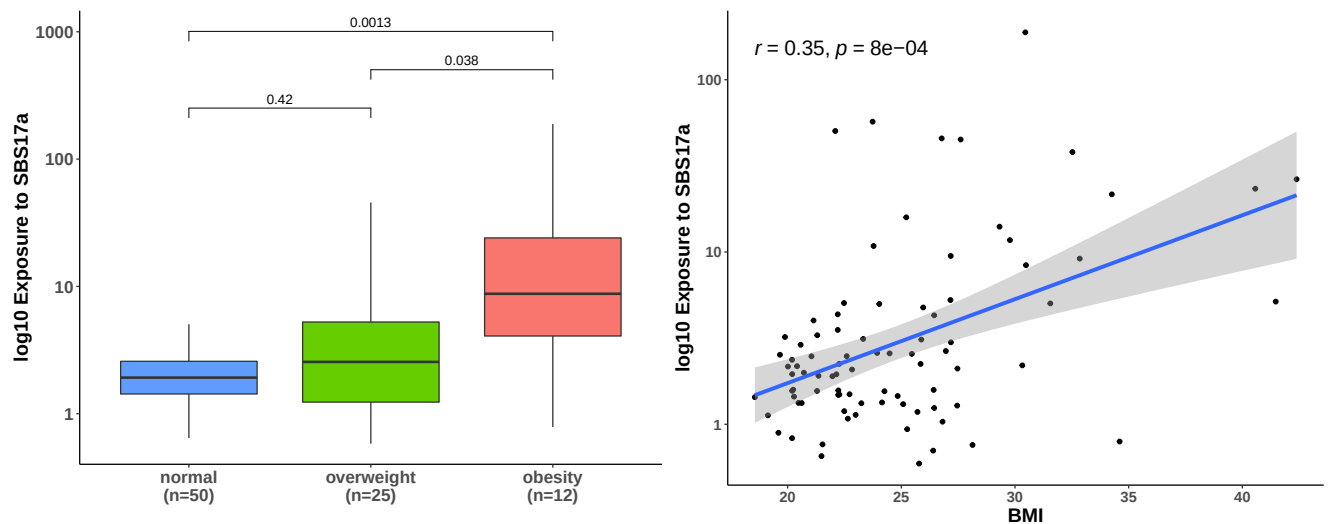


Figura 11 – Níveis de exposição dos pacientes do TCGA-ESCA à assinatura SBS17a do COSMIC v3.2.

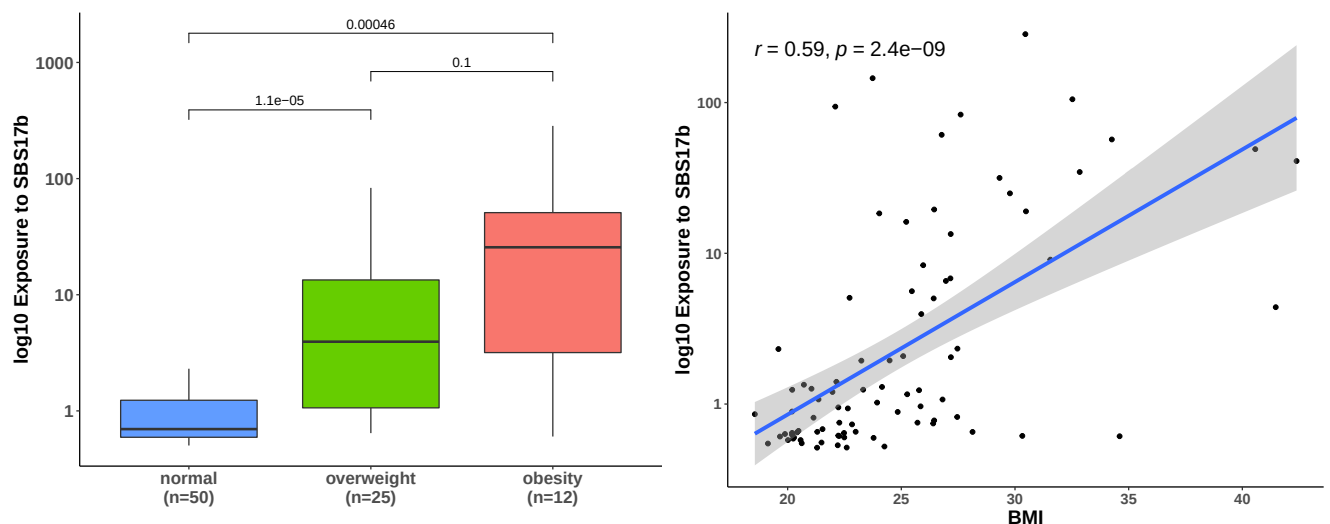


Figura 12 – Níveis de exposição dos pacientes da coorte do TCGA-ESCA à assinatura SBS17b do COSMIC v3.2.

Além disso, ao comparar os níveis de presença, *i.e.* exposições, das assinaturas SBS17a e SBS17b decompostas pelos demais métodos, vemos que os valores de exposição à ambas assinaturas está altamente correlacionado em todos os métodos testados, indicando um consenso e consequentemente reforçando a acurácia na detecção da presença dessas assinaturas nos pacientes testados (**Figura 13**).

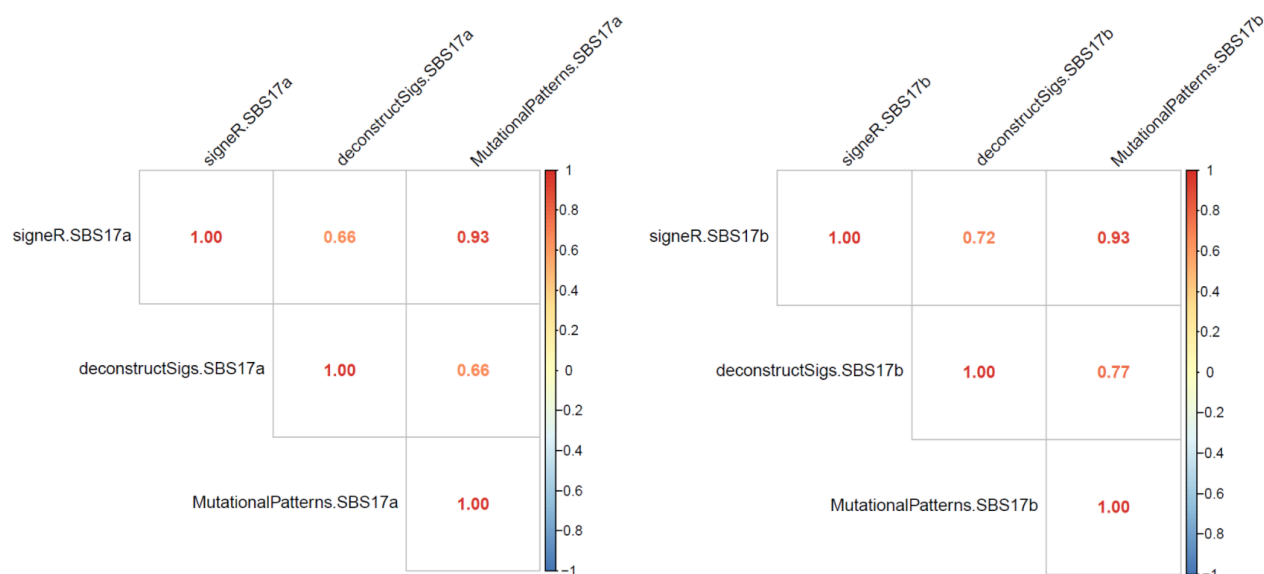


Figura 13 – Correlação entre os valores de exposição às assinaturas SBS17a e SBS17b extraídas pelos métodos signeR, deconstructSigs e MutationalPatterns.

Ainda, na coorte TCGA-ESCA, foram extraídas assinaturas mutacionais SBS *de novo* por 4 diferentes métodos, implementados pelos pacotes SigProfiler, MutationalPatterns, signeR e SomaticSignatures. Consideramos como detectadas na coorte apenas as assinaturas detectadas por 2 ou mais métodos distintos. Seguindo esse critério, foram detectadas assinaturas mutacionais com alta similaridade dos cossenos (≥ 0.7) com diversas assinaturas SBS do COSMIC v3.1, incluindo SBS1, SBS3, SBS5, SBS6, SBS13, SBS15, SBS16, SBS17b, SBS18, SBS29, SBS40 e SBS87. Vale destacar, que todos os métodos implementados em nosso *pipeline* identificaram processos mutacionais *de novo* com alta similaridade com a assinatura SBS17b (**Figura 14**).

Utilizando como base o método signeR, a assinatura mutacional S3 (**Figura 15**), extraída *de novo* apresenta 0.82 de similaridade dos cossenos com a assinatura SBS17b (**Figura 16**) do COSMIC v3.1. Além disso, observamos que pacientes com sobrepeso e obesidade mostram uma tendência a estarem mais expostos à assinatura S3 (equivalente a assinatura SBS17b do COSMIC v3.1) do que os pacientes com IMC normal ($p=0.0027$ e $p=0.00031$, respectivamente). E também, essa assinatura mostra correlação positiva significativa com o IMC dos pacientes do tumor TCGA-ESCA ($\rho = 0.5$, $p < 0.0001$) (**Figura 17**), concordando com os resultados descritos na **Figura 10**.

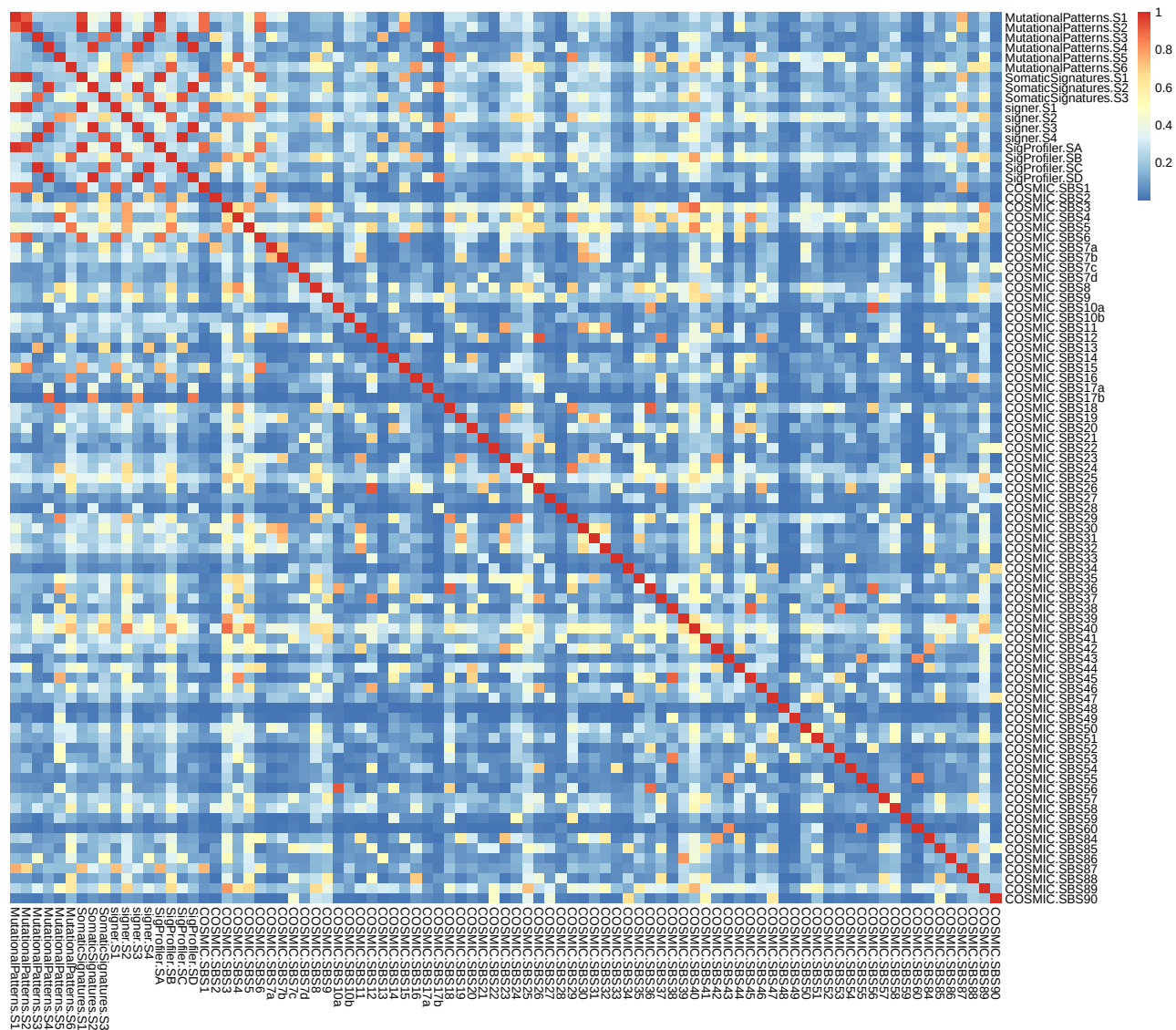


Figura 14 – Similaridade entre as assinaturas extraídas por todos os métodos *de novo* e todas assinaturas SBS descritas no COSMIC v3.2. A escala de cor representa a similaridade dos cossenos

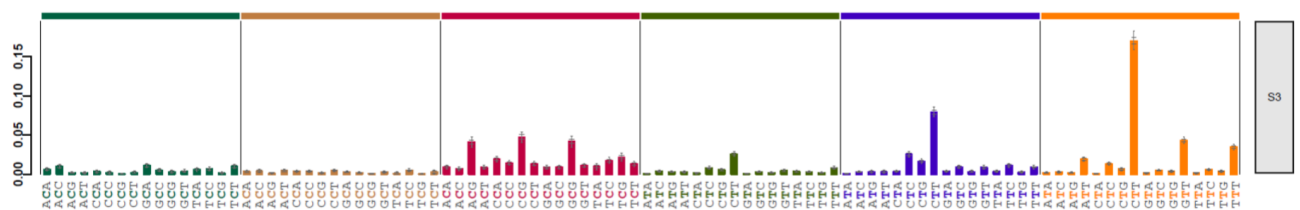


Figura 15 – Assinatura mutacional S3 extraída *de novo* pelo método signerR.

Ainda, a partir do método sigProfiler, foi extraída *de novo* 1 assinatura mutacional DBS da coorte do TCGA-ESCA, DBS78A que mostrou 0.87 de similaridade com as assinaturas DBS2 e DBS4 do COSMIC v3.1 (**Figura 18**).

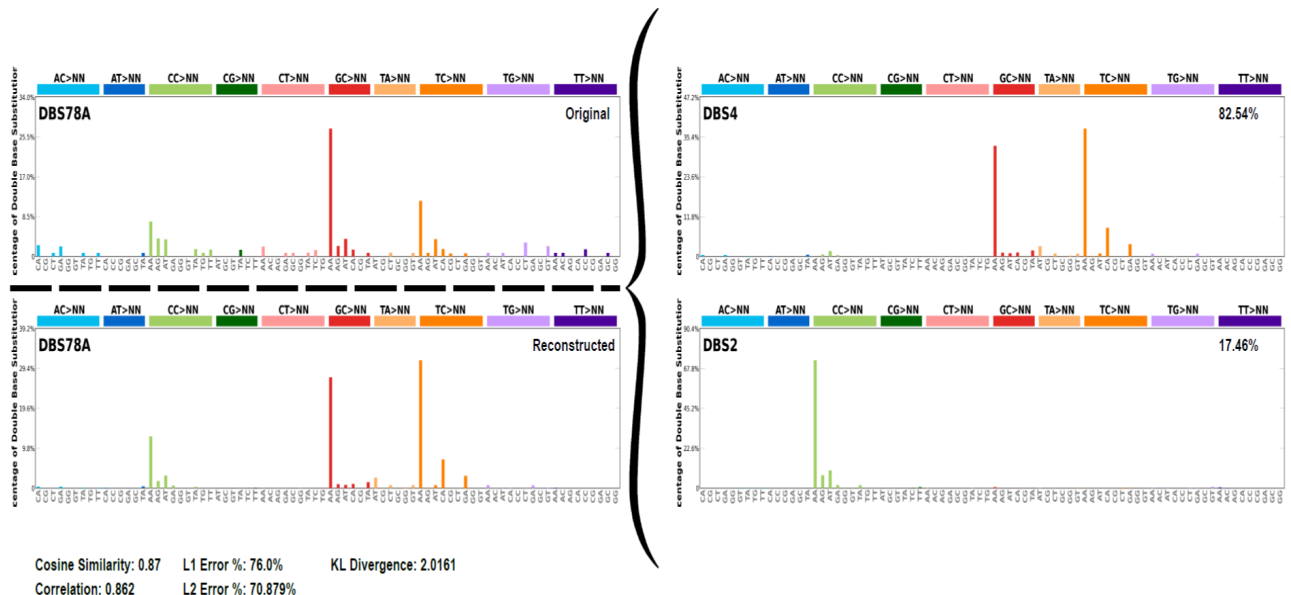


Figura 18 – Assinaturas mutacionais DBS extraídas pelo sigProfiler. No lado superior esquerdo está demonstrada a assinatura mutacional original extraída *de novo* e no lado inferior esquerdo a assinatura mutacional reconstruída. No lado direito são mostradas as assinaturas mutacionais DSB do COSMIC v3.1 que são similares à assinatura extraída *de novo*.

Além disso, também foi extraída *de novo* 1 assinatura mutacional ID pelo sigProfiler, que mostrou 0.97 de similaridade com as assinaturas ID2 e ID7 do COSMIC v3.1 (**Figura 19**).

Foi realizada a correlação dessas assinaturas DBS2, DBS4, ID2 e ID7 detectadas pelo sigProfiler, com o IMC dos pacientes do tumor TCGA-ESCA (**Figura 20**). Nenhuma das assinaturas citadas acima mostrou correlação significativa com o IMC dos pacientes, portanto decidimos não considerar o conjunto de assinaturas mutacionais do tipo DSB e ID, e prosseguir apenas com as assinaturas do tipo SBS, especificamente SBS17a e SBS17b nas análises subsequentes.

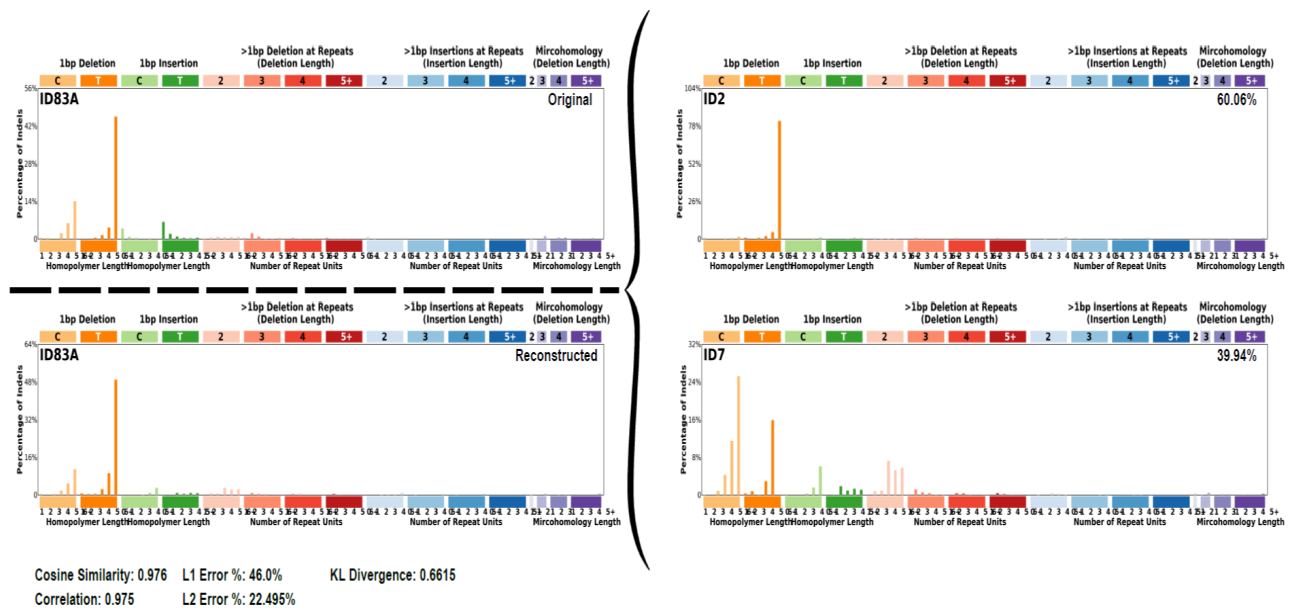


Figura 19 – Assinaturas mutacionais ID extraídas pelo sigProfiler. No lado superior esquerdo está demonstrada a assinatura mutacional original extraída *de novo* e no lado inferior esquerdo a assinatura mutacional reconstruída. No lado direito são mostradas as assinaturas mutacionais ID do COSMIC v3.1 que são equivalentes à assinatura extraída *de novo*.

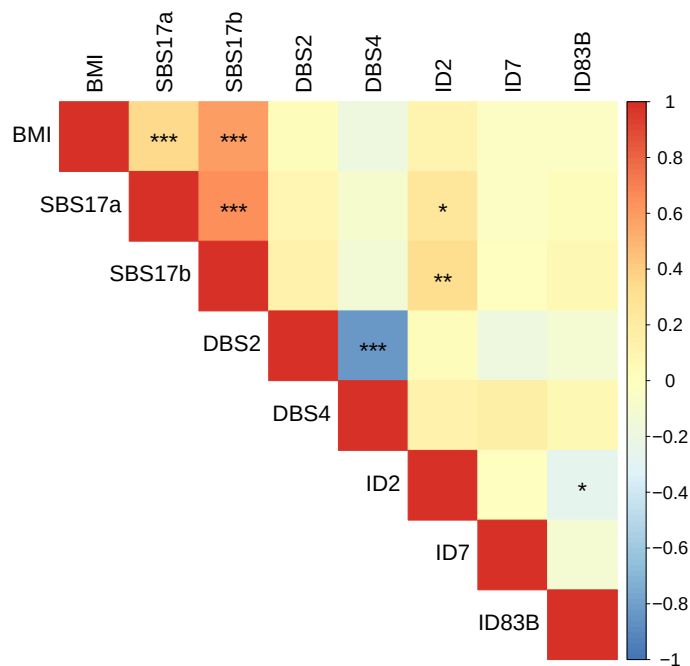


Figura 20 – Correlação entre assinaturas SBS, DBS e ID extraídas pelo sigProfiler, e o IMC dos pacientes da coorte do TCGA-ESCA. A escala de cor representa o valor de correlação. *p-value* <0.05: *; *p-value* <0.01: **; *p-value* <0.001: ***.

Esses resultados indicam que há uma associação entre altos níveis de IMC e a ação de duas assinaturas mutacionais de substituição única, SBS17a e SBS17b, em pacientes do projeto TCGA-

ESCA. Todavia, não encontramos associações entre os níveis de IMC e assinaturas mutacionais de base dupla ou indels. Sendo assim, mostramos que a obesidade está associada à uma classe de mutações específica.

4.2.1.1 Assimetria mutacional

Nas condições biológicas normais, com o mecanismo de reparo acoplado à transcrição (TCR-NER) funcionando normalmente, espera-se que haja uma diferença na densidade mutacional entre as fitas transcrita e não transcrita, uma vez que a fita não transcrita permanece como fita simples durante a transcrição, sendo mais vulnerável a danos, enquanto que a fita transcrita conta com a ação do mecanismo TCR. Essa diferença na densidade mutacional entre as fitas de DNA durante a transcrição pode ser descrita como uma assimetria transcricional. Então com a ação do mecanismo TCR espera-se que a assimetria entre as duas fitas de DNA seja alta, já que apenas uma fita conta com reparo.

A partir da análise de assimetria (*strand bias*) transcricional realizada por meio do pacote MutationalPatterns (Blokzijl et al. 2018), foi possível observar que a densidade mutacional do dano T>G, característico da assinatura SBS17b, ilustrado em rosa, é maior nos pacientes com sobrepeso e obesidade do que nos pacientes normais (**Figura 21 A**). Porém a diferença na densidade mutacional entre as duas fitas, a assimetria mutacional, desse tipo de dano não mostra padrões muito claros ou alguma tendência significativa ($p\text{-value} > 0.05$) de ser menor nos pacientes com sobrepeso e obeso em relação aos normais (**Figura 21 B**). Assim, decidimos investigar em análises subsequentes, os padrões de expressão de genes dessa via de reparo.

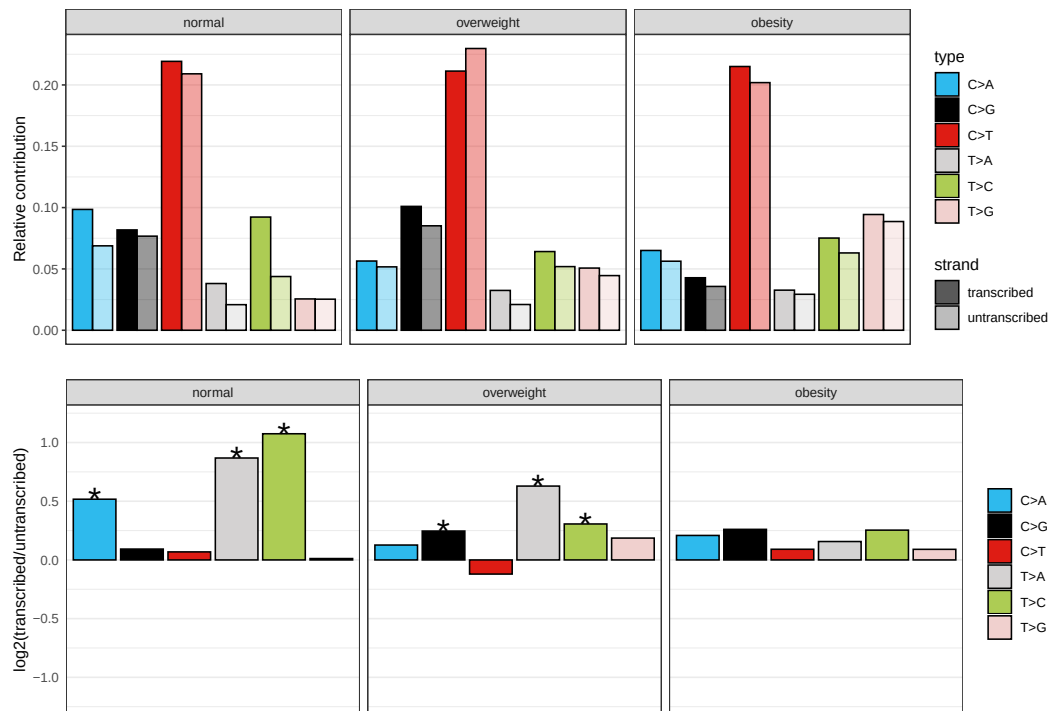


Figura 21 – Diferença na densidade mutacional entre as fitas transcrita e não transcrita dos pacientes com IMC normal, sobrepeso e obesidade da coorte do TCGA-ESCA. As densidades mutacionais entre as duas fitas foram comparadas pelo teste de Poisson. p -value <0.05: *; p -value <0.01: **; p -value <0.001: ***.

4.3 INFILTRADO INFLAMATÓRIO E GENES MARCADORES DE INFLAMAÇÃO

A fim de avaliar o perfil do infiltrado inflamatório nos diferentes grupos de pacientes do TCGA-ESCA, considerando o IMC, foram estimadas as proporções de 22 tipos de células inflamatórias a partir da análise pelo CIBERSORT. A partir do *score* gerado, foi feita a comparação entre os grupos obesidade, sobrepeso e normal pelo teste de Kruskal-Wallis, mostrando que a infiltração das populações de células B *naïve*, células plasmáticas, células T CD8, células T CD4 *memory resting*, células T *follicular helper*, células T regulatórias, células NK ativadas, Monócitos, Macrófagos M0, M1 e M2, células dendríticas em repouso e neutrófilos é significativamente distinta (p -value < 0.05) entre os pacientes obesos, com sobrepeso e normais (**Figura 22**). Esses resultados sugerem que o sobrepeso e obesidade podem alterar a composição de infiltrados inflamatórios do tumor, recrutando uma quantidade maior ou menor de determinadas células imunes do que os pacientes normais.

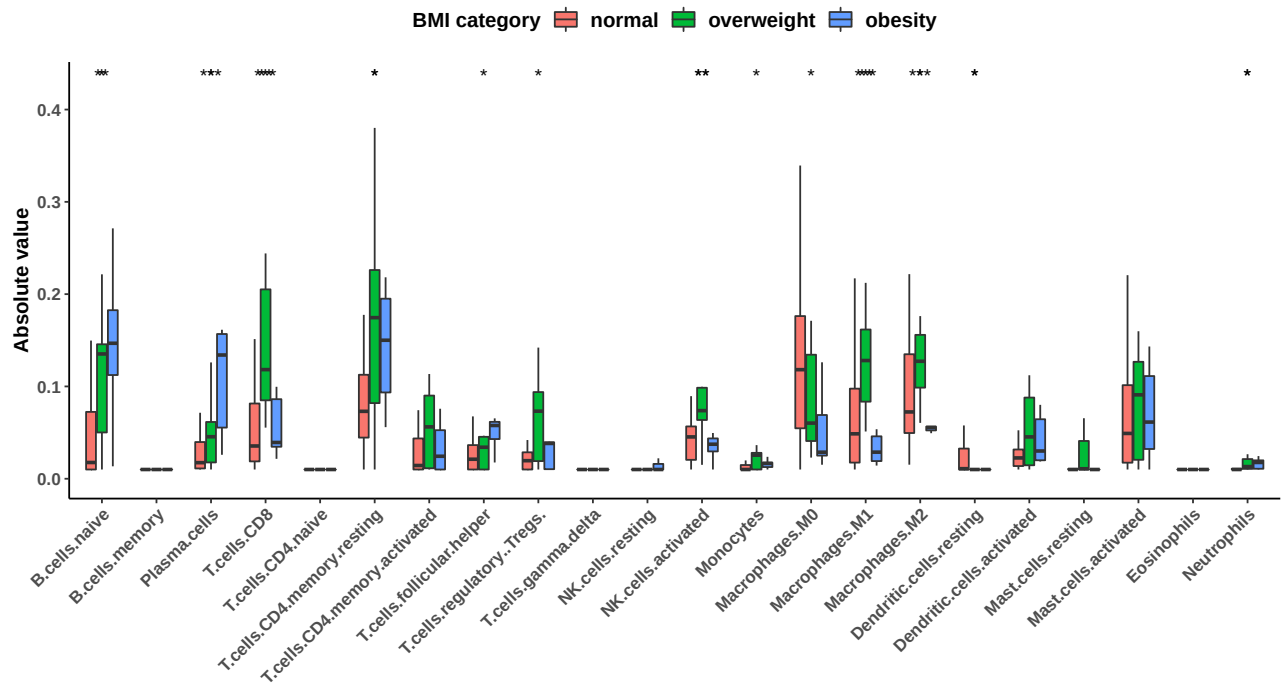


Figura 22 – Infiltrado inflamatório em pacientes obesos, com sobrepeso e normais em TCGA-ESCA. O eixo x representa o tipo celular e o eixo y representa o valor relativo do infiltrado gerado pelo *software* CIBERSORT em cada grupo de IMC. Os grupos foram comparados pelo teste não paramétrico Kruskal-Wallis. *p-value* <0.05: *; *p-value* <0.01: **; *p-value* <0.001: ***.

Além disso, observa-se que a infiltração de células B *naïve*, células plasmáticas, células T CD4 de memória em repouso e neutrófilos mostram correlações positivas significativas com o IMC e com os níveis de exposição às assinaturas SBS17a e SBS17b e os macrófagos M0 mostram correlações negativas (**Figura 23**).

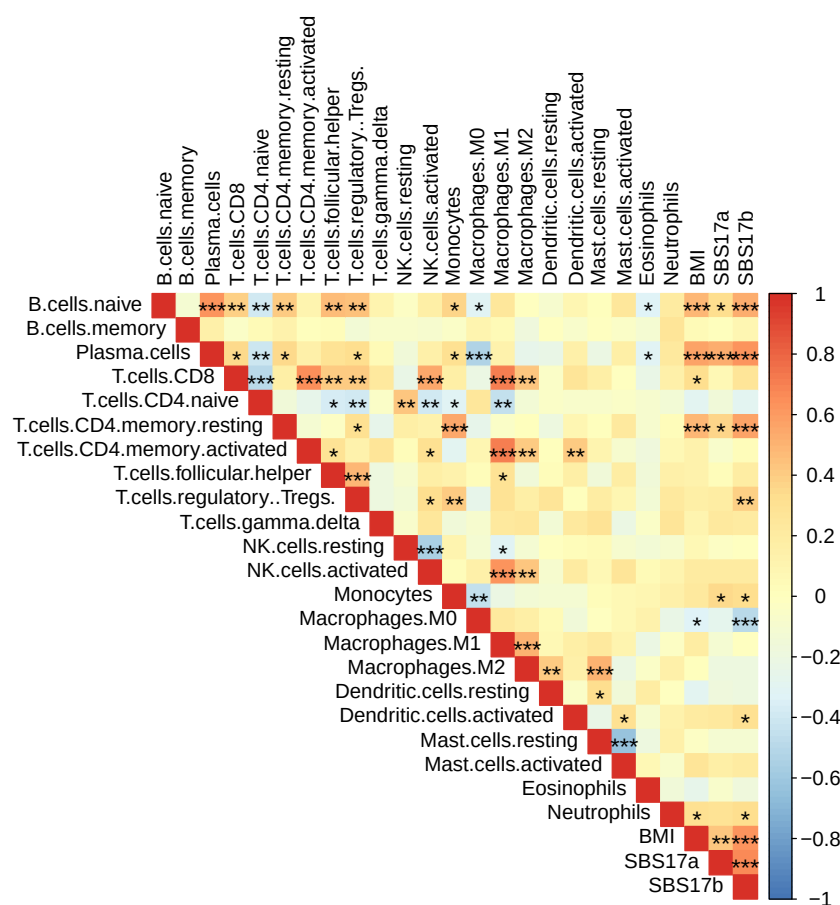


Figura 23 – Correlação entre os infiltrados inflamatórios e as variáveis BMI, exposição às assinaturas SBS17a e SBS17b. A escala de cor representa o valor de correlação. p -value <0.05: *; p -value <0.01: **; p -value <0.001: ***.

Ao avaliar os níveis de expressão normalizados de diversos genes marcadores imunes comparados por meio do teste de *Kruskal-Wallis*, foi observado que dentre os marcadores pró-inflamatórios, os marcadores IL17F, NOS2, IL12A, IL12B, STAT6, IL2 e IL17A apresentaram níveis mais elevados de expressão nos pacientes com sobrepeso e obesidade (**Figura 24**) do que nos pacientes normais. Já os marcadores HIF1A, REL, IL1A e IL6, apresentam níveis mais baixos de expressão nos pacientes com sobrepeso e obesidade do que nos pacientes normais.

Além disso, é possível observar que os níveis de expressão normalizados de alguns desses genes pró-inflamatórios, incluindo IL17F, NOS2, IL12B e STAT6, mostram correlações positivas significativas com o IMC e os níveis de exposição às assinaturas SBS17a e SBS17b. Já o gene HIF1A mostra correlações negativas (**Figura 25**).

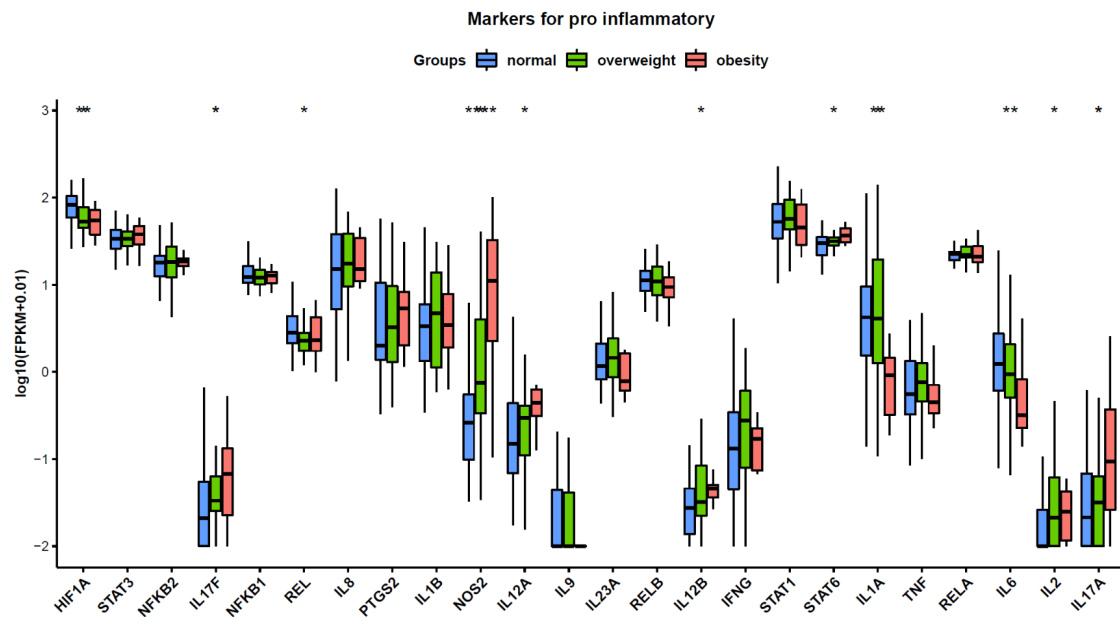


Figura 24 – Expressão normalizada de genes marcadores pró-inflamatórios nos grupos normal, sobrepeso e obesidade da coorte do TCGA-ESCA. O eixo x representa o gene marcador e o eixo y representa seu nível de expressão em log₁₀ FPKM para cada grupo. *p-value* <0.05: *; *p-value* <0.01: **; *p-value* <0.001: ***.

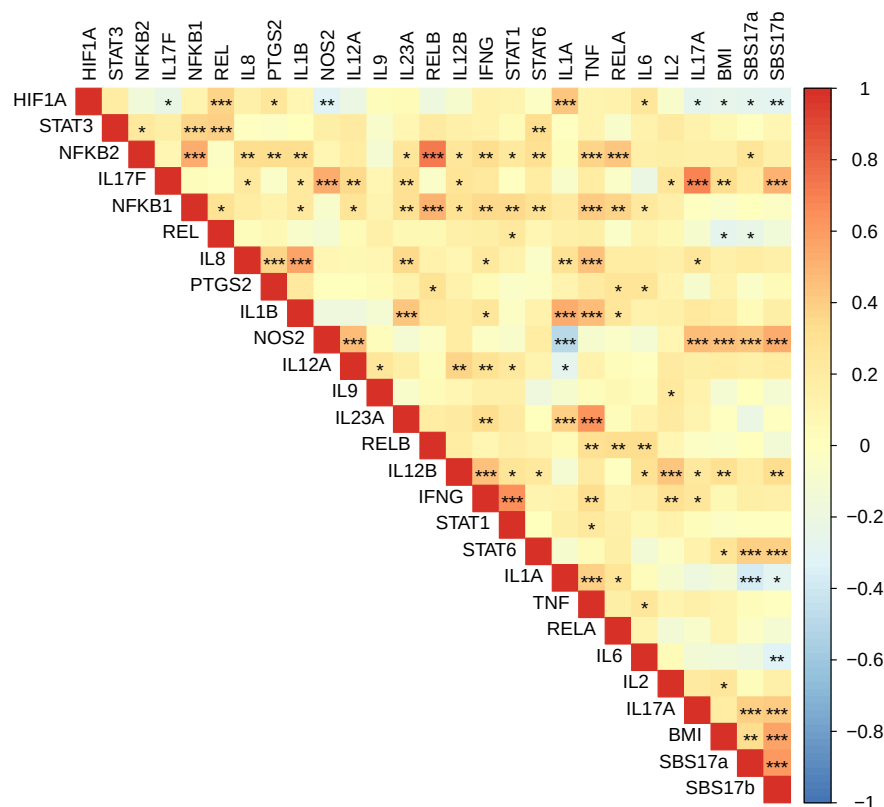


Figura 25 – Correlação entre genes pró-inflamatórios e as variáveis BMI, exposição às assinaturas SBS17a e SBS17b. *p-value* <0.05: *; *p-value* <0.01: **; *p-value* <0.001: ***.

Dos marcadores anti-inflamatórios, destacam-se diferenças nos marcadores TGFB1 e IL10 que apresentam níveis mais baixos de expressão nos pacientes com sobrepeso e obesos do que nos pacientes normais (**Figura 26**). Ainda, observa-se uma correlação significativa negativa dos níveis de expressão normalizados do gene TGFB1 com o IMC e os níveis de exposição às assinaturas SBS17a e SBS17b (**Figura 27**).

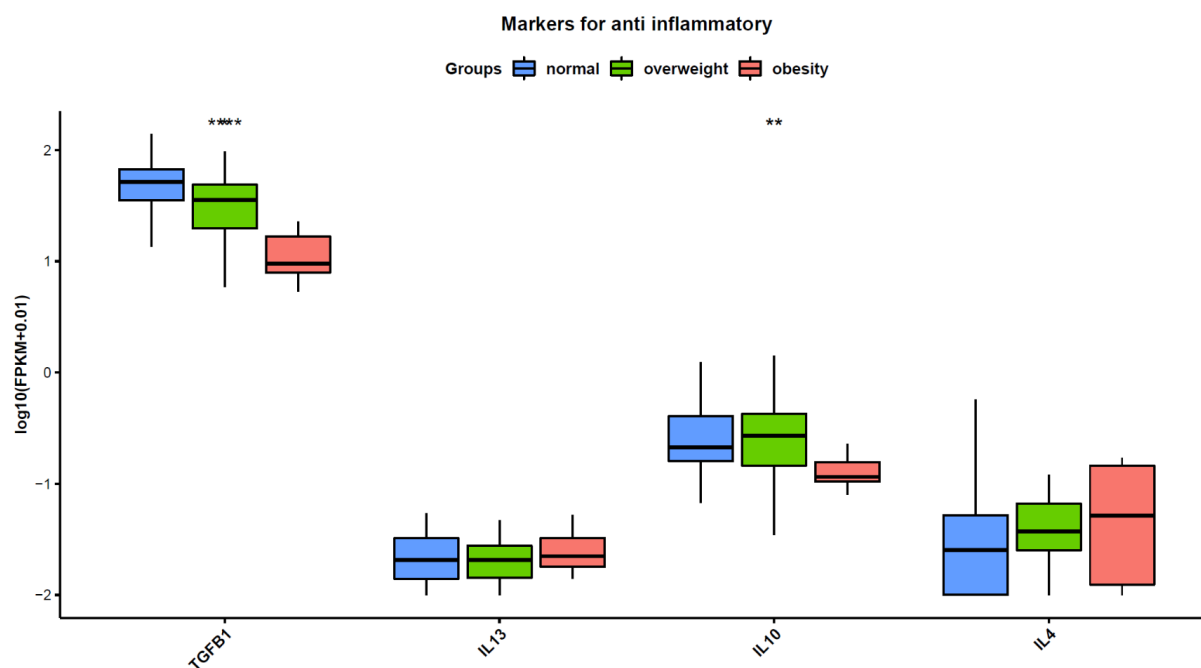


Figura 26 – Expressão normalizada de genes marcadores anti-inflamatórios nos grupos normal, sobrepeso e obesidade. O eixo x representa o gene marcador e o eixo y representa seu nível de expressão em FPKM normalizadas em $\log_{10}$ para cada grupo. Os grupos foram comparados pelo teste não paramétrico de Kruskal-Wallis. *p-value* <0.05: *; *p-value* <0.01: **; *p-value* <0.001: ***.

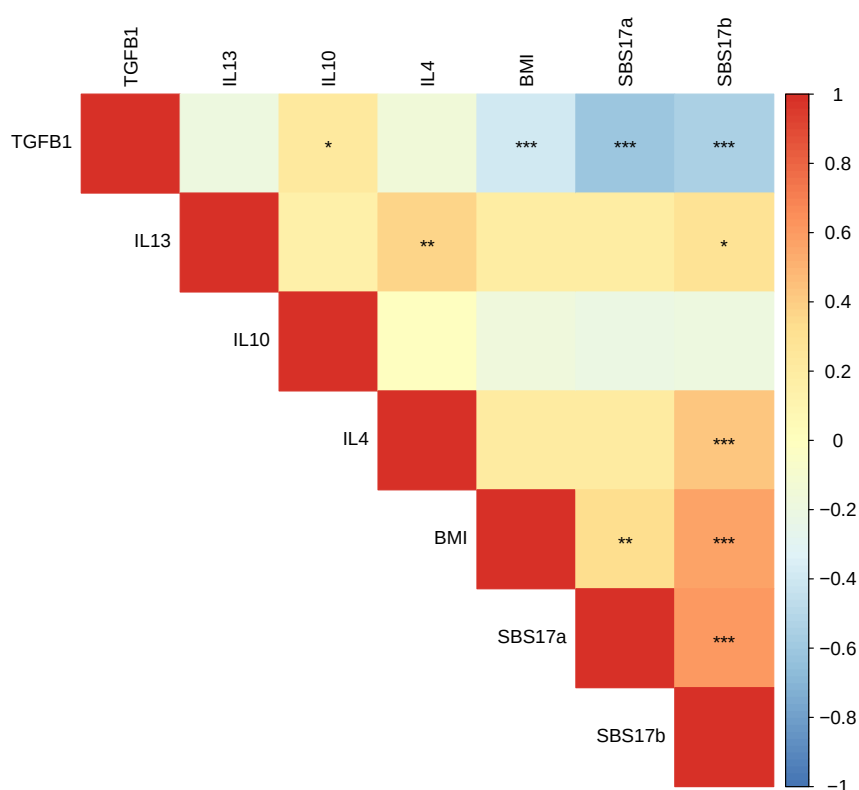


Figura 27 – Correlação entre genes anti-inflamatórios e as variáveis BMI, exposição às assinaturas SBS17a e SBS17b. A escala de cor representa o valor de correlação. *p-value* <0.05: *; *p-value* <0.01: **; *p-value* <0.001: ***.

Assim, mostramos uma desregulação no perfil inflamatório de pacientes com altos níveis de IMC a nível celular, pelo recrutamento de diferentes tipos de células inflamatórias, e também à nível genético, pela desregulação da expressão de genes de vias pró- e anti-inflamatórias.

4.4 GENES MARCADORES DE REPARO

Foram selecionados genes marcadores das principais vias de reparo responsáveis por reparar danos oxidativos, incluindo BER, NER e MMR. De acordo com a **Figura 28**, observa-se que na via BER os níveis de expressão normalizada dos genes PARP2, APEX1, TDG, NEIL3 e POLB são menores nos pacientes com sobrepeso e obesidade do que nos pacientes normais. Já alguns genes como OGG1, MUTYH e NEIL1 tem maiores níveis de expressão normalizadas em pacientes com sobrepeso e obesos. Inclusive, alguns desses genes, como PARP2, APEX1 e NEIL3 mostram correlações negativas significativas com o IMC dos pacientes e os níveis de exposição às assinaturas

SBS17b, enquanto que os genes OGG1 e NEIL1 mostram correlações positivas (**Figura 29**).

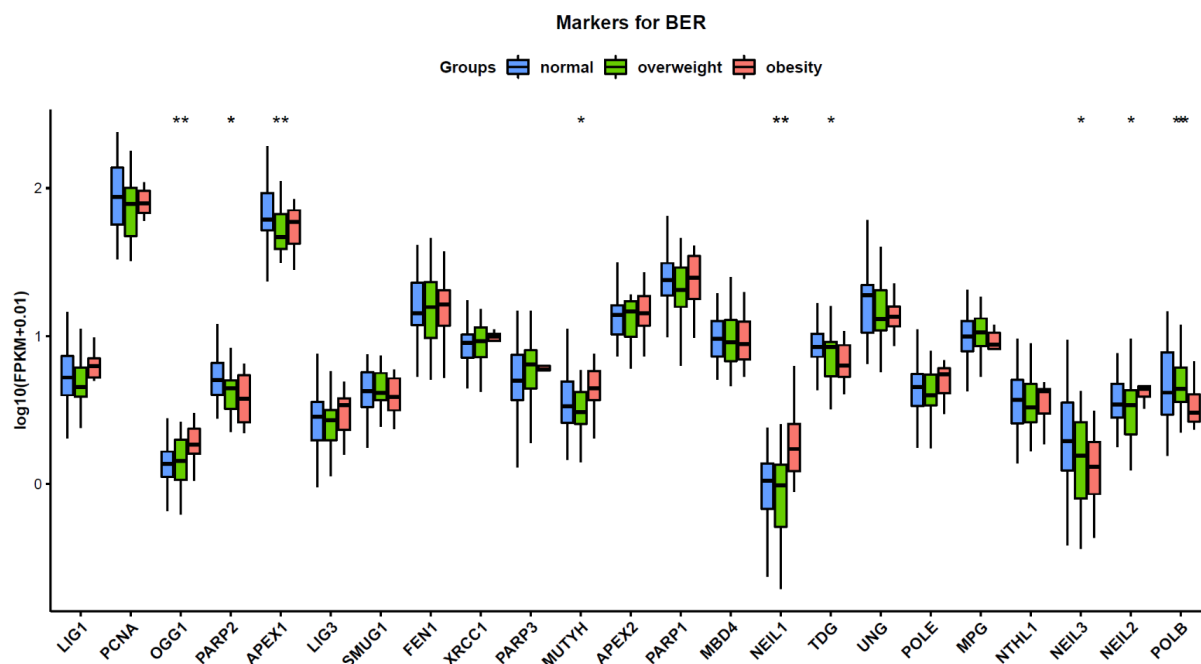


Figura 28 – Expressão normalizada de genes marcadores da via BER nos grupos normal, sobrepeso e obesidade. O eixo x representa o gene marcador e o eixo y representa seu nível de expressão em log₁₀ FPKM para cada grupo. p -value <0.05: *; p -value <0.01: **; p -value <0.001: ***.

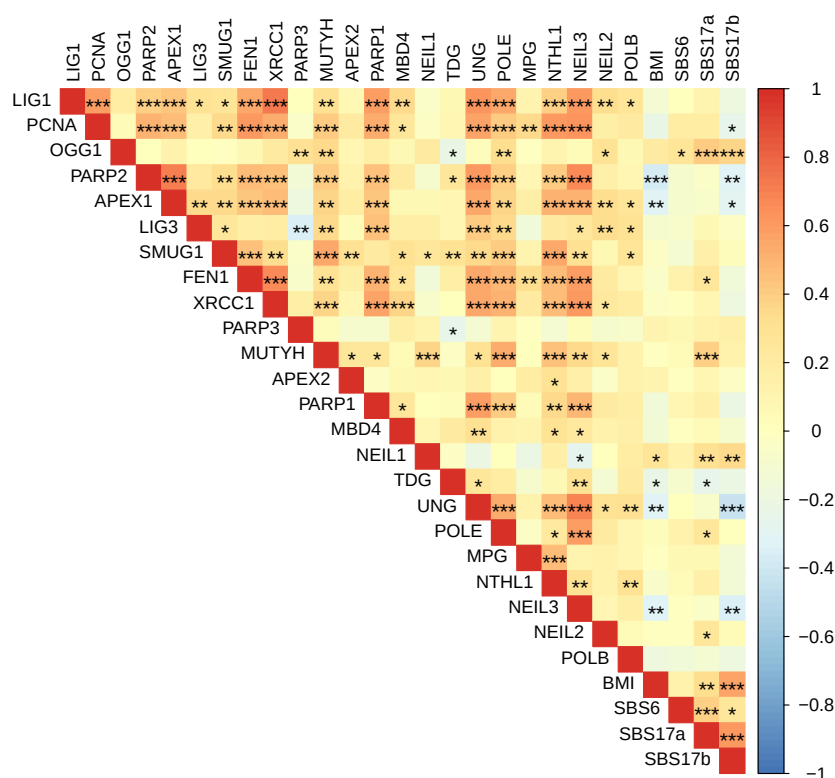


Figura 29 – Correlação entre genes da via BER e as variáveis BMI, exposição às assinaturas SBS17a e SBS17b. p -value <0.05: *; p -value <0.01: **; p -value <0.001: ***.

Na via NER nota-se que os níveis de expressão normalizados dos genes RPA1, MNAT1, ERCC3, GTF2H3, DDB2 e XAB2 estão diminuídos nos pacientes com sobrepeso e obesidade em relação aos normais, enquanto que dos genes UVSSA, RPA4, e ERCC5, estão aumentados (**Figura 30**). Além disso, todos os genes mencionados com expressão diminuída, exceto ERCC3, também mostram correlação negativa significativa com o IMC e os níveis de expressão às assinaturas SBS17a e SBS17b, e todos os genes com expressão aumentada mostram correlação positiva significativa com o IMC e os níveis de expressão às assinaturas SBS17a e SBS17b (**Figura 31**).

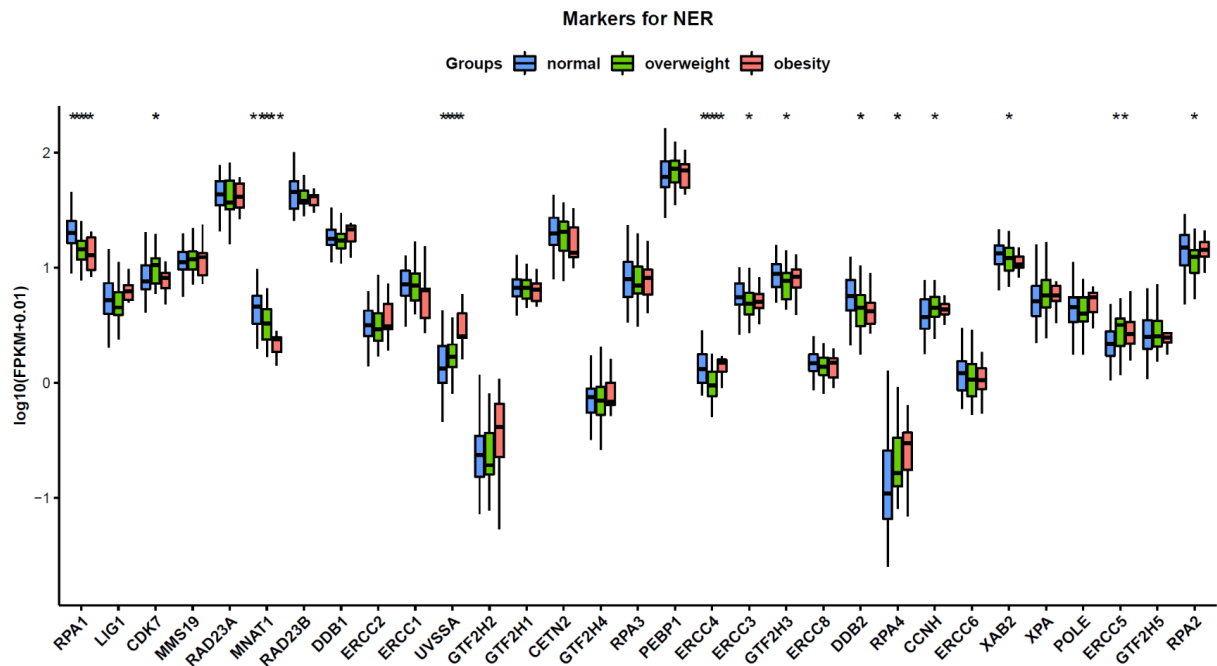


Figura 30 – Expressão normalizada de genes marcadores da via NER nos grupos normal, sobrepeso e obesidade. O eixo x representa o gene marcador e o eixo y representa seu nível de expressão em $\log_{10}$ FPKM para cada grupo. p -value <0.05: *; p -value <0.01: **; p -value <0.001: ***.

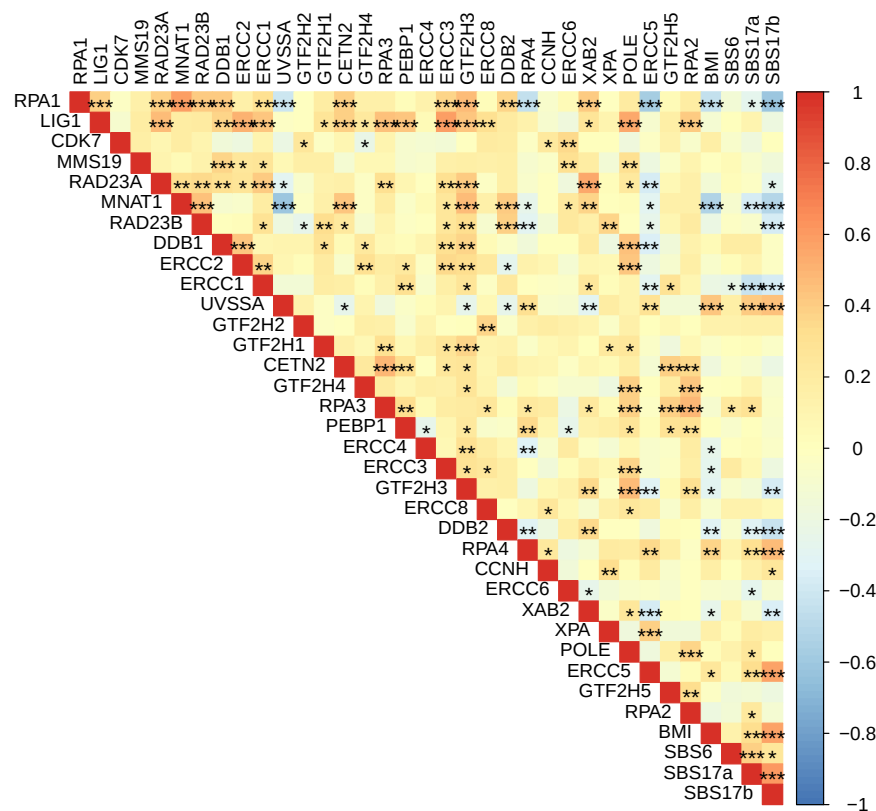


Figura 31 – Correlação entre genes da via NER e as variáveis BMI, exposição às assinaturas SBS17a e SBS17b. p -value <0.05: *; p -value <0.01: **; p -value <0.001: ***.

Na via MMR, a **Figura 32** mostra que os genes MSH6 e PMS2 tem os níveis de expressão normalizados diminuídos quando comparados pacientes com sobrepeso e obesidade com os pacientes normais, enquanto que os genes MLH3 e MSH5 mostram aumento nesses níveis. Além disso, gene MSH6 mostra correlação negativa significativa com o IMC e com os níveis de exposição à assinatura SBS17b (**Figura 33**).

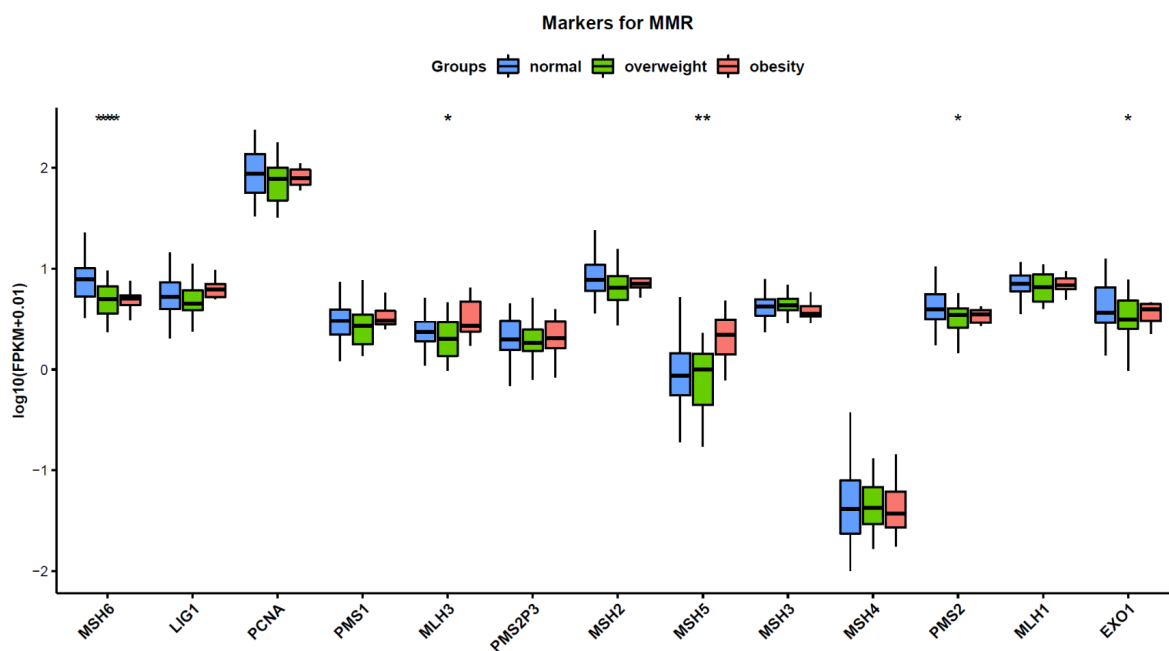


Figura 32 – Expressão normalizada de genes marcadores da via MMR nos grupos normal, sobrepeso e obesidade. O eixo x representa o gene marcador e o eixo y representa seu nível de expressão em $\log_{10}$ FPKM para cada grupo. p -value <0.05: *; p -value <0.01: **; p -value <0.001: ***.

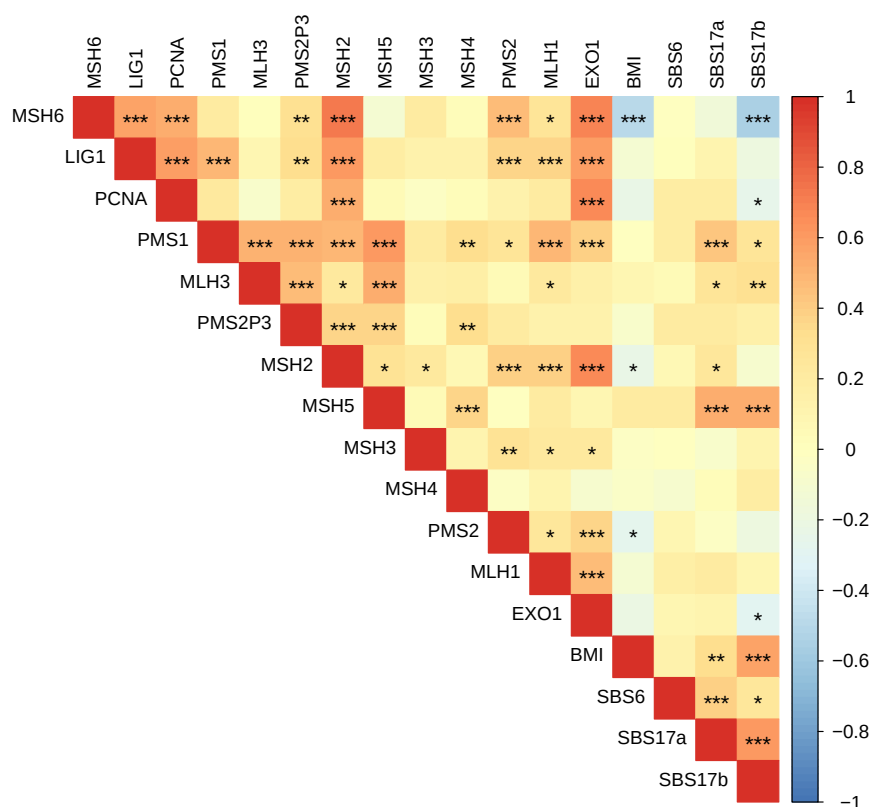


Figura 33 – Correlação entre genes da via MMR e as variáveis BMI, exposição às assinaturas SBS17a e SBS17b. *p-value* <0.05: *; *p-value* <0.01: **; *p-value* <0.001: ***.

Desta forma, considerando as principais vias responsáveis por reparar danos oxidativos, mostramos uma associação entre os altos níveis de IMC, a desregulação de genes envolvidos nessas vias e a atividade mutacional das assinaturas SBS17a e SBS17b.

4.5 IDENTIFICAÇÃO DOS GENES DIFERENCIALMENTE EXPRESSOS

Foram recuperadas informações de expressão gênica a partir de RNA-Seq da base de dados do TCGA, no formato de *read counts* por gene, para um total de 53 pacientes (10 obesos e 43 normais) da coorte do TCGA-ESCA. De acordo com os critérios de $|\log_2\text{FoldChange}| \geq 1$ e $p\text{-valor}$ ajustado ≤ 0.01 pelo método de *Benjamini-Hochberg*, foram identificados 2516 genes diferencialmente expressos (DEGs, *Differentially Expressed Genes*) entre os grupos obesidade e normal, sendo 447 *up*-regulados e 2069 *down*-regulados (**Figura 34**).

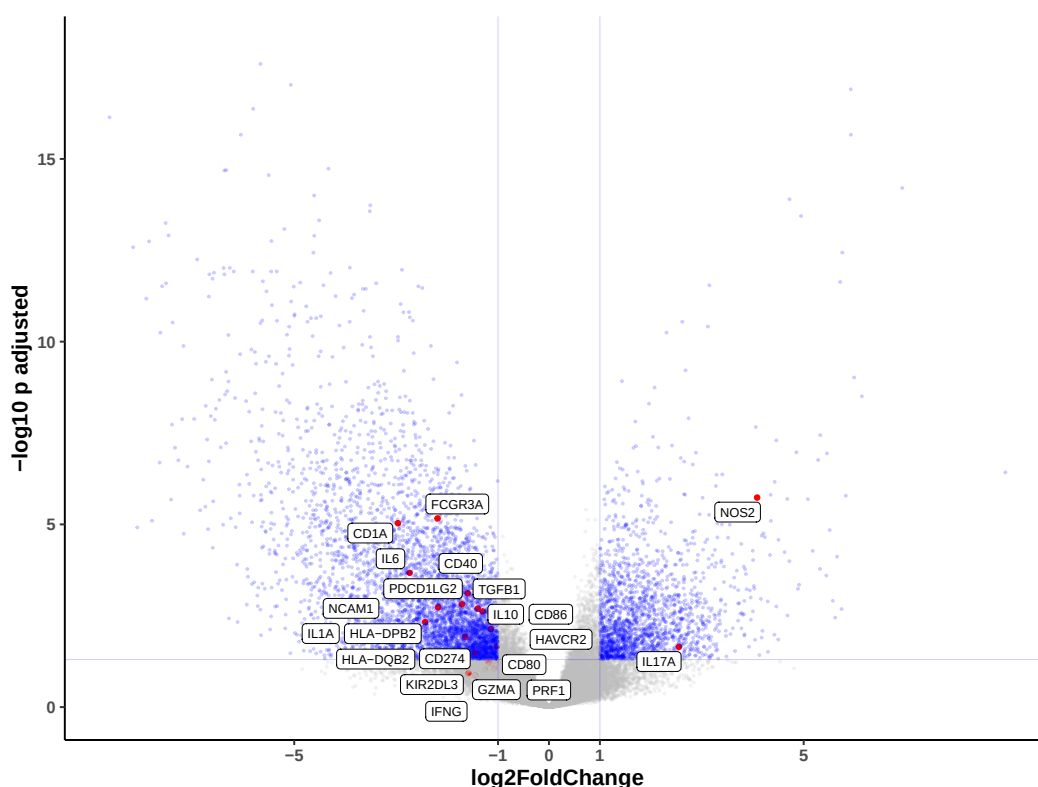


Figura 34 – Perfil de expressão diferencial entre os grupos normal e obesidade. Os pontos azuis indicam os genes diferencialmente expressos de acordo com os critérios definidos acima. Os pontos vermelhos são genes envolvidos em processos oxidativos, inflamação e supressão/exaustão. No eixo x está representado o $p\text{-value}$ ajustado, normalizado em $-\log_{10}$, no eixo y está representado o $\log_2$ Fold Change entre os grupos obesidade e normal.

A partir da Análise de Componentes Principais (PCA, *Principal Component Analysis*) foi observado que apesar do baixo número de amostras no grupo obesidade, são formados dois *clusters*, um com uma maior concentração de amostras normais e outro com uma maior concentração de amostras do grupo obesidade. As componentes PC1 e PC2 explicam 60% da variância dos dados (**Figura 35**).

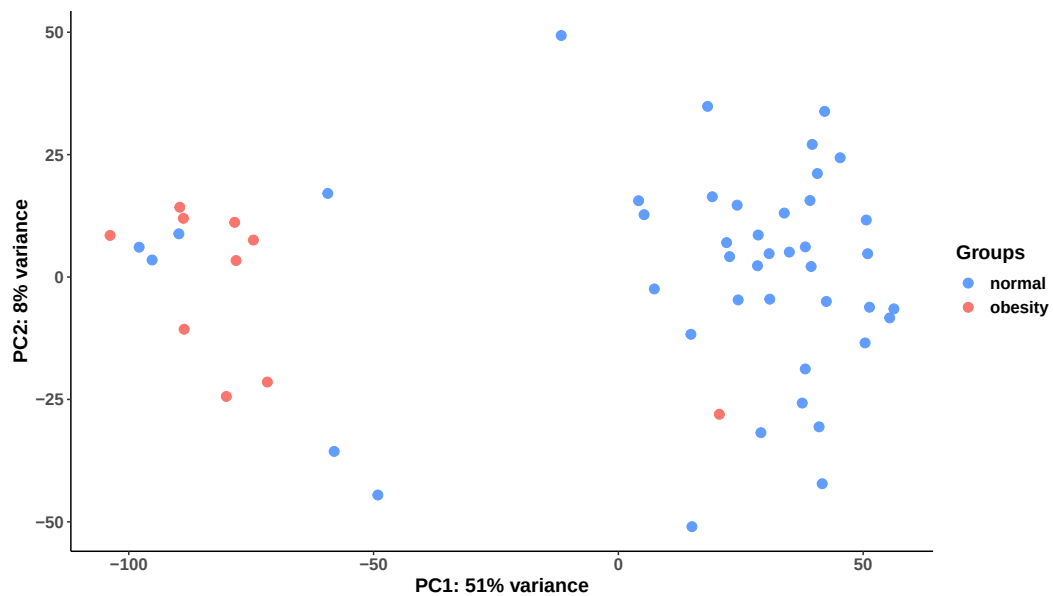


Figura 35 – Clusterização das amostras de acordo com os grupos normal e obesidade. Os pontos vermelhos representam pacientes obesos e os pontos azuis representam pacientes normais.

O *heatmap* dos DEGs foi ilustrado na **Figura 36**. A escala de cor representa os *read counts* transformados pela estabilização de variância (*vst*, *Variance Stabilizing Transformation*) e clusterizados de acordo com os grupos obesidade e normal. Os *clusters* foram definidos de forma hierárquica, não supervisionada, a partir de distância euclidiana e pelo método de clusterização de Ward.

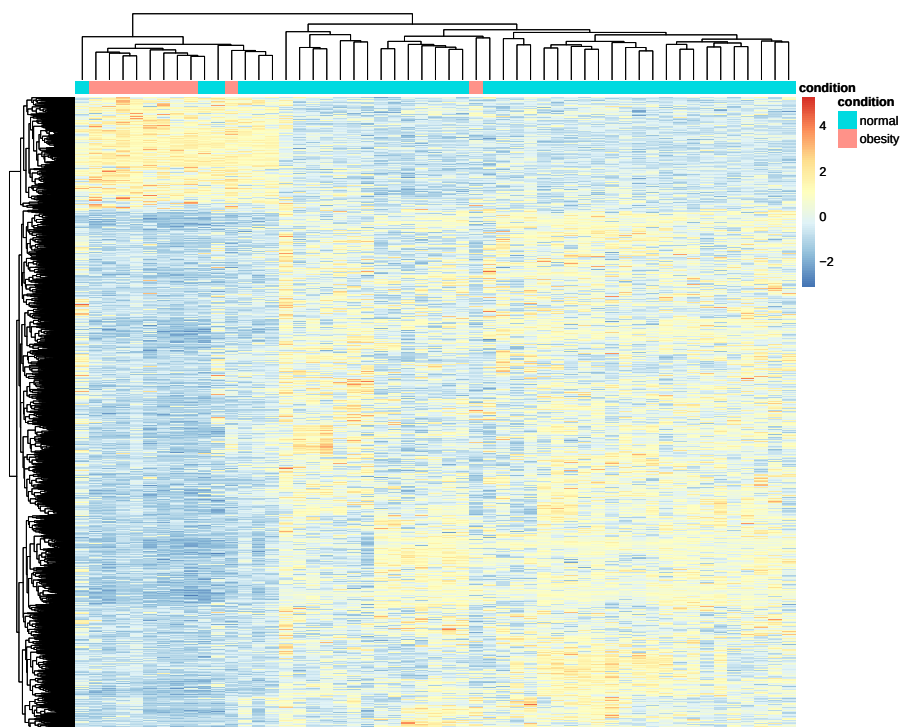


Figura 36 – *Heatmap* dos DEGs.

4.6 ANÁLISE DO STATUS DE METILAÇÃO PARA DEGS

Na **Figura 36** descrita acima, foi observado que grande parte dos DEGs estão *down-regulados* em sua maioria em pacientes obesos. Para investigar possíveis mecanismos biológicos induzindo alteração na expressão dos DEGs identificados, *e.g.* uma possível alteração epigenética, foi realizada uma análise de metilação para esse conjunto de genes.

O *heatmap* dos valores de metilação para os DEGs foi ilustrado na **Figura 37**. A escala de cor representa o β -value para cada gene, agrupados seguindo a mesma ordem de clusterização do *heatmap* mostrado acima.

Como resultado, é possível observar que o *cluster* de genes *down-regulados* em pacientes obesos, ilustrados na **Figura 36**, possuem altos β -values (≥ 0.5), indicando que a *down*-regulação desses é decorrente de altos valores de metilação, ou seja, esses genes contam com hiper-metilação de seus promotores em pacientes obesos. Por outro lado, o *cluster* de genes *up-regulados* em obesos, mostrado no *heatmap* acima, possuem baixos β -values (≤ 0.4), indicando que a *up*-regulação desses genes se dá pela hipo-metilação de seus promotores (**Figura 37**).

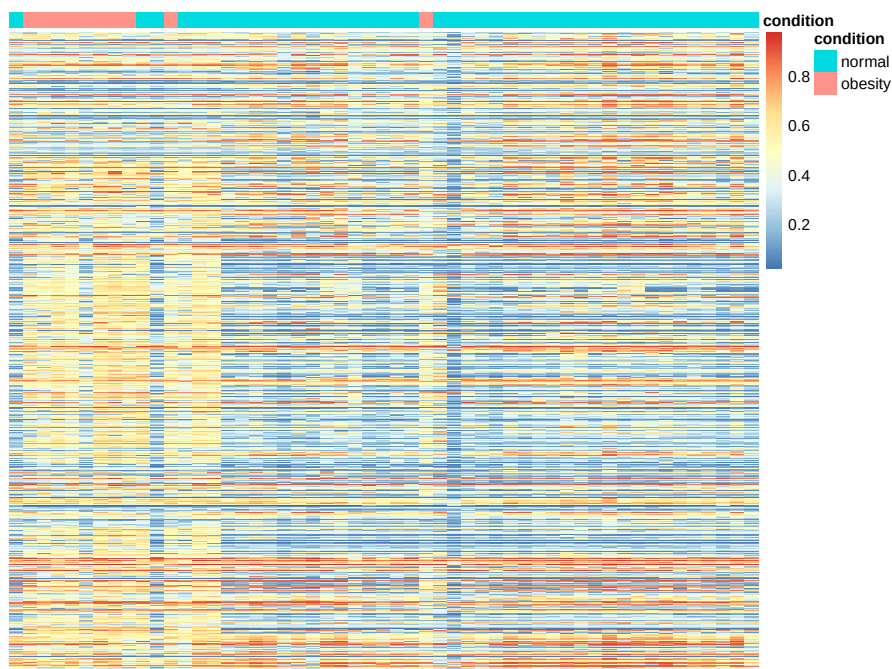


Figura 37 – Heatmap do β -value dos DEGs.

Ainda, é possível observar de uma forma global, que pacientes obesos possuem valores muito mais altos de β -value do que os pacientes normais (**Figura 38**), e essa diferença é estatisticamente significativa **Figura 39** (p -value < 0.0001).

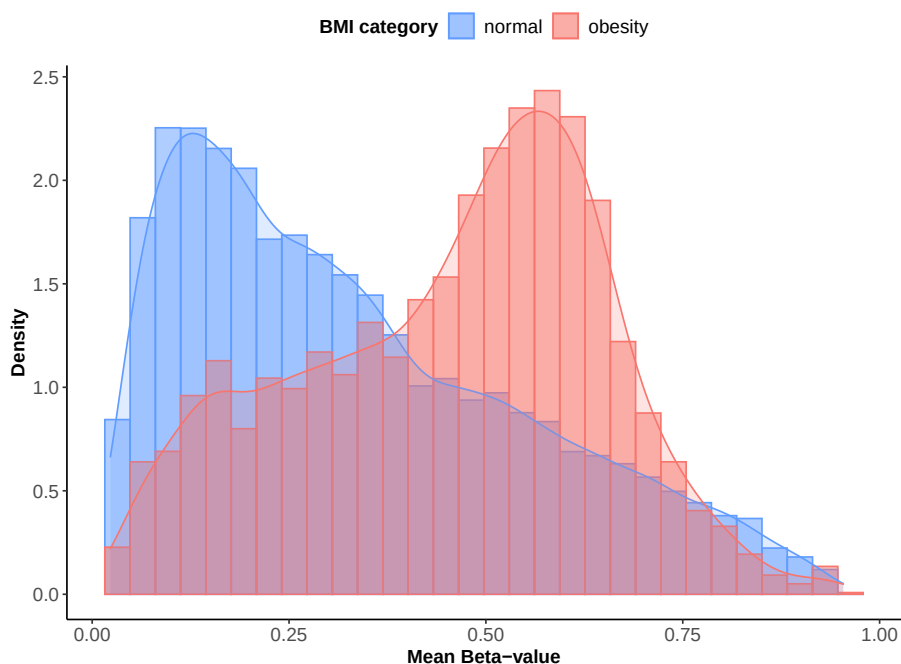


Figura 38 – Distribuição dos β -values de pacientes obesos e normais.

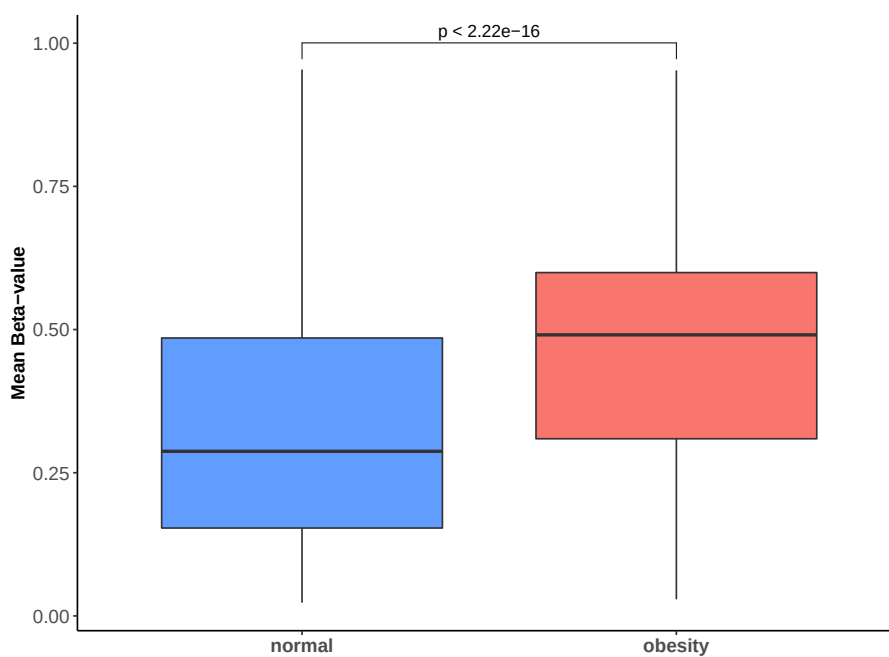


Figura 39 – Comparação dos β -values entre pacientes obesos e normais.

Juntos, esses resultados indicam que a obesidade induz alterações epigenéticas no genoma tumoral de forma global, particularmente por meio da desregulação dos padrões de metilação.

4.7 ENRIQUECIMENTO DE VIAS PARA DEGS

Para explorar a função biológica desses DEGs, foi realizada uma análise de enriquecimento de vias do *Gene Ontology* (GO) (Carbon et al. 2019) por meio do pacote clusterProfiler (Yu et al. 2012).

De acordo com as vias GO selecionadas com *p-value* ajustado pelo método de *Benjamini-Hochberg* < 0.05 para funções moleculares (MF, *Molecular Function*), foi possível observar que, os genes *up*-regulados em pacientes obesos, estão associados à diversas vias relacionadas à condição de obesidade, como atividade de transporte de lipídeos, atividade de transferência de esterol e colesterol, atividade de oxidoreductase, entre outras (**Figura 40 esquerda**). Já os genes *down*-regulados em obesidade, estão envolvidos em vias responsáveis pelas atividades de citocinas, ligação de receptores de IL-1, ligação de fatores de crescimento, entre outras (**Figura 40 direita**).

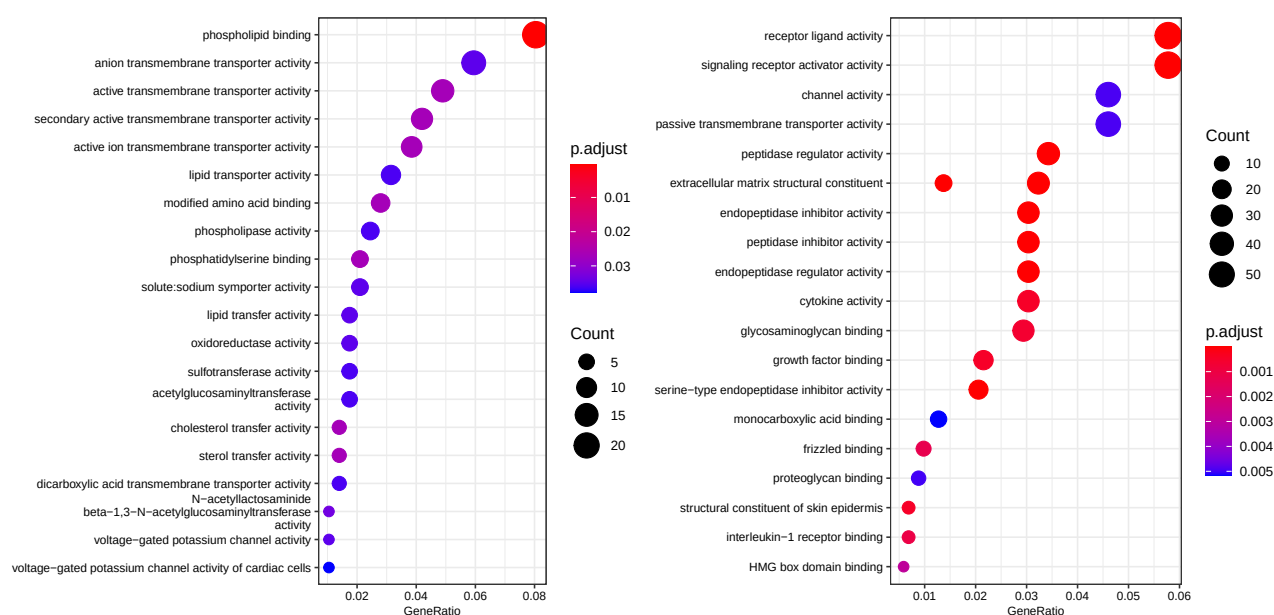


Figura 40 – Vias alteradas em pacientes obesos a partir do enriquecimento de vias pelo GO de acordo com MF.

4.8 ANÁLISE DO STATUS DE METILAÇÃO PARA MARCADORES INFLAMATÓRIOS

Os processos inflamatórios são de particular interesse para esse estudo. Na **Figura 24**, foi mostrado que diversos genes pró-inflamatórios têm expressão aumentada em pacientes obesos quando comparados aos pacientes normais. A fim de investigar possíveis mecanismos epigenéticos envolvidos na regulação da expressão desses genes, foram avaliados os valores de metilação de seus promotores.

De importância, observamos (**Figura 41**) que o nível de metilação (β -value) do gene NOS2 é significativamente menor em pacientes obesos comparados aos normais (p -value<0.001), ao passo que podemos observar uma relação inversa nos níveis de expressão gênica nesses pacientes (**Figura 24**) indicando, assim, uma hipo-metilação de seu promotor.

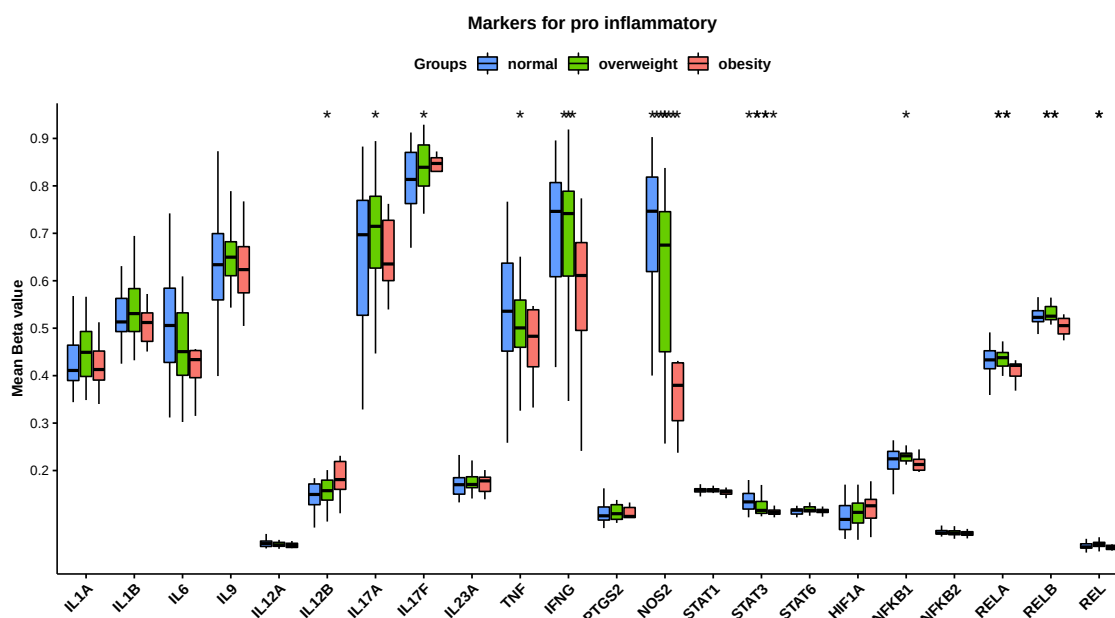


Figura 41 – Comparação dos β -values de genes inflamatórios entre pacientes obesos e normais. p -value <0.05: *; p -value <0.01: **; p -value <0.001: ***.

De forma similar, observa-se uma correlação negativa entre os valores de expressão gênica e metilação do gene NOS2 ($r=-0.46$, p -value<0.001), indicando que conforme a expressão gênica aumenta, menor o β -value. Ainda, é mostrado que essa associação inversa é mais forte em pacientes obesos ($r=-0.85$, p -value=0.02), que também possuem menor β -value e maior expressão de NOS2 que os demais (**Figura 42**).

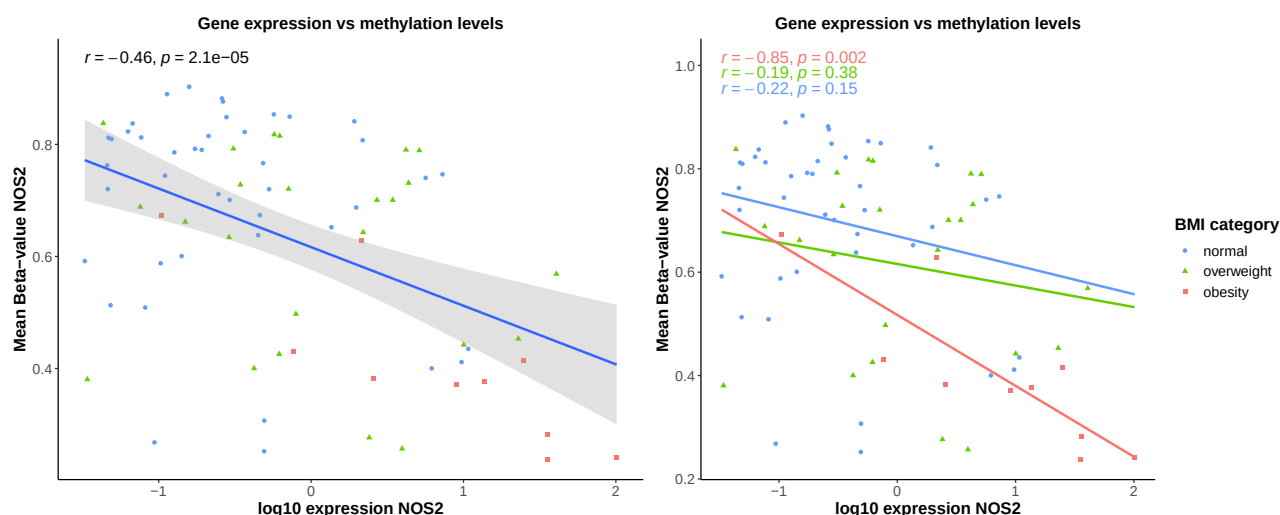


Figura 42 – Correlação de *Pearson* entre o β -value e a expressão normalizada do gene NOS2 na escala $\log_{10}$. A escala de cores representa a classificação de IMC.

4.9 ASSOCIAÇÃO ENTRE EXPRESSÃO GÊNICA E METILAÇÃO DO GENE NOS2, E A ASSINATURA MUTACIONAL SBS17B

Conforme descrito em resultados anteriores, observa-se uma associação positiva entre a expressão gênica de NOS2 e a assinatura SBS17b, ambos fatores também associados à altos níveis de IMC (**Figura 25**). Também, foi mostrado que o aumento de expressão gênica de NOS2 se dá pela hipo-metilação de seu promotor em pacientes obesos (**Figuras 41-42**).

Tendo isso em vista, foram investigadas possíveis associações diretas entre a assinatura SBS17b e os níveis de metilação de NOS2. Como resultado, foi possível observar que, de fato, os níveis de exposição à assinatura SBS17b e os níveis de metilação de NOS2 possuem uma associação inversa ($r = -0.54$, $p\text{-value} < 0.001$), mostrando que, quanto maior os níveis de exposição à assinatura SBS17b, menor o β -value do gene NOS2 (**Figura 43**).

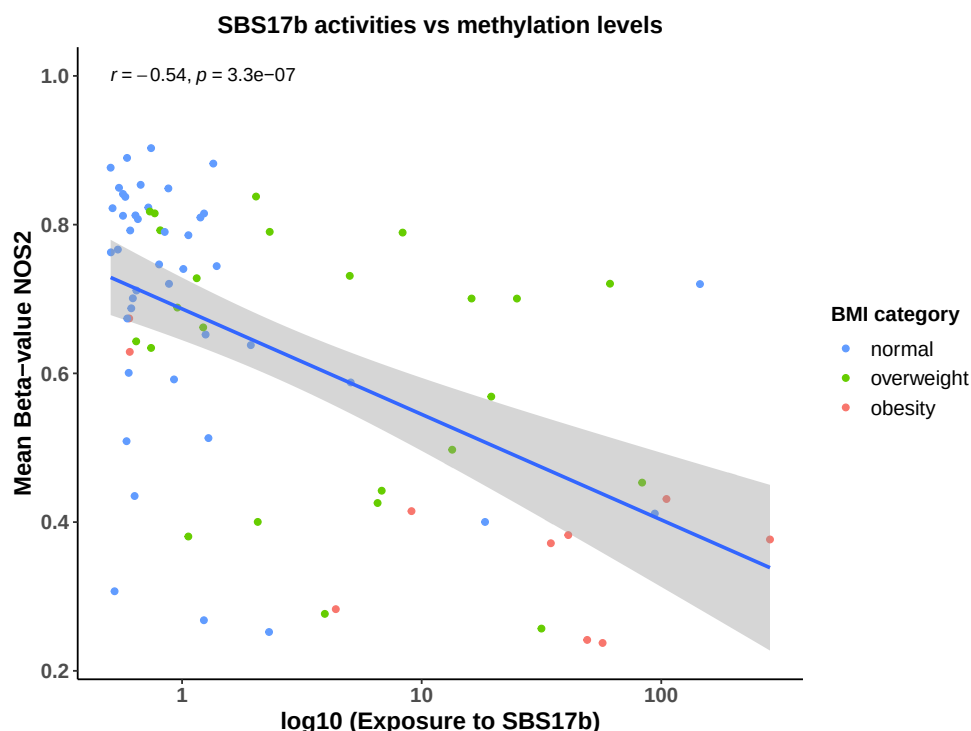


Figura 43 – Correlação de *Pearson* entre o β -value do gene NOS2 e a exposição à assinatura SBS17b na escala $\log_{10}$. A escala de cores representa a classificação de IMC.

Com o intuito de avaliar o real impacto da assinatura SBS17b na metilação desse gene, foi realizada uma regressão logística simples considerando como variável dependente o *status* de metilação, onde pacientes com β -value NOS2 ≥ 0.4 , foram classificados como "NOS2 Metilado", e β -value < 0.4 "NOS2 Não metilado". Como resultado, observa-se que a assinatura SBS17b aumenta em 1.14 vezes a chance do promotor do gene NOS2 estar "Não metilado" (**Tabela 3**).

Tabela 3 – Regressão logística simples para a metilação do NOS2 e a assinatura SBS17b. OR, *Odds Ratio* (Razão de chances); CI, *Confidence Interval* (Intervalo de confiança).

Variável	Coeficiente	Erro Padrão	CI (95%)		p-value
			Lower	Upper	
Intercept	2.3670	0.4623	1.460893	3.2730235	3.05e-07
$\log_{10}(\text{SBS17b})$	-1.1417	0.3970	-1.919858	-0.3635269	0.00403

A performance e o poder de discriminação do modelo de regressão logística para a metilação do gene NOS2 pela assinatura mutacional SBS17b é mostrada a seguir (**Figura 44**).

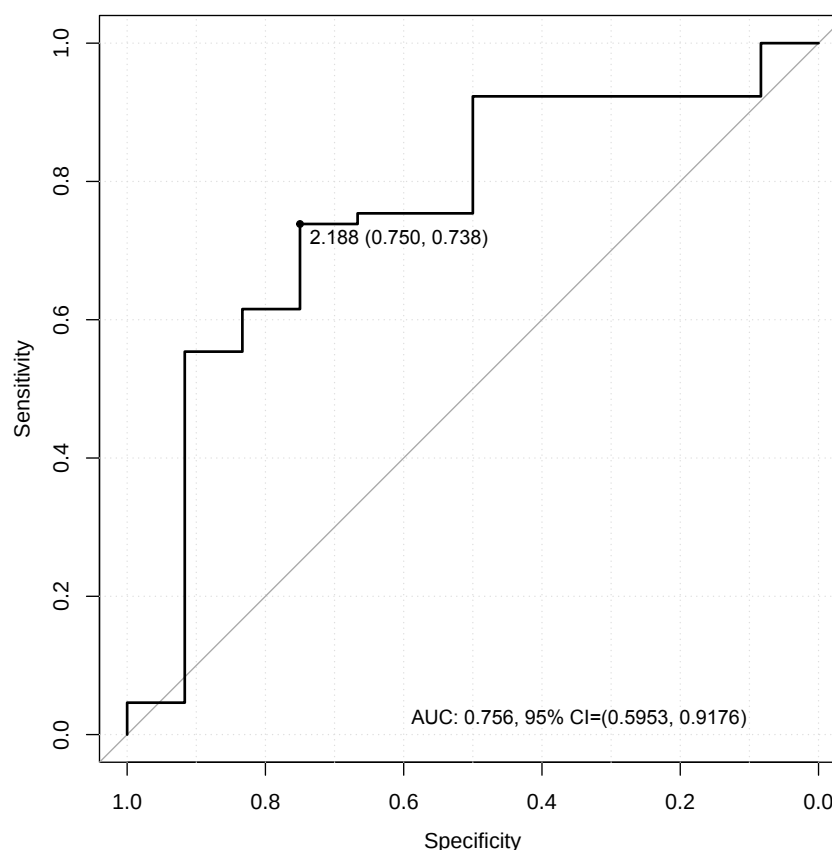


Figura 44 – Curva ROC (*Receiver Operating Characteristic*) mostrando o poder de discriminação do modelo de regressão logística. AUC, *Area under the curve* (Área sob a curva); CI, *Confidence Interval* (Intervalo de confiança).

Juntos, esses resultados nos indicam que a presença da assinatura SBS17b está diretamente associada com a hipo-metilação do promotor do gene NOS2, em pacientes obesos. Sendo assim, concluímos que a hipo-metilação do gene NOS2, em pacientes obesos com carcinoma de esôfago, em estágio inicial pode ser considerado um substituto para a assinatura mutacional SBS17b.

4.10 EXPRESSÃO E METILAÇÃO DO GENE NOS2 EM OUTROS TUMORES DO TCGA

A fim de investigar se a hiper-expressão do NOS2, decorrente da hipo-metilação do seu promotor, em pacientes obesos ocorre em diferentes tumores ou se é tumor-específica, foram realizadas análises de expressão diferencial e do *status* de metilação entre pacientes obesos e normais para todos os 14 tumores do TCGA explorados nesse projeto. Como resultado, a **Figura 45** abaixo ilustra que o gene NOS2 é somente diferencialmente expresso no tumor de esôfago (TCGA-ESCA).

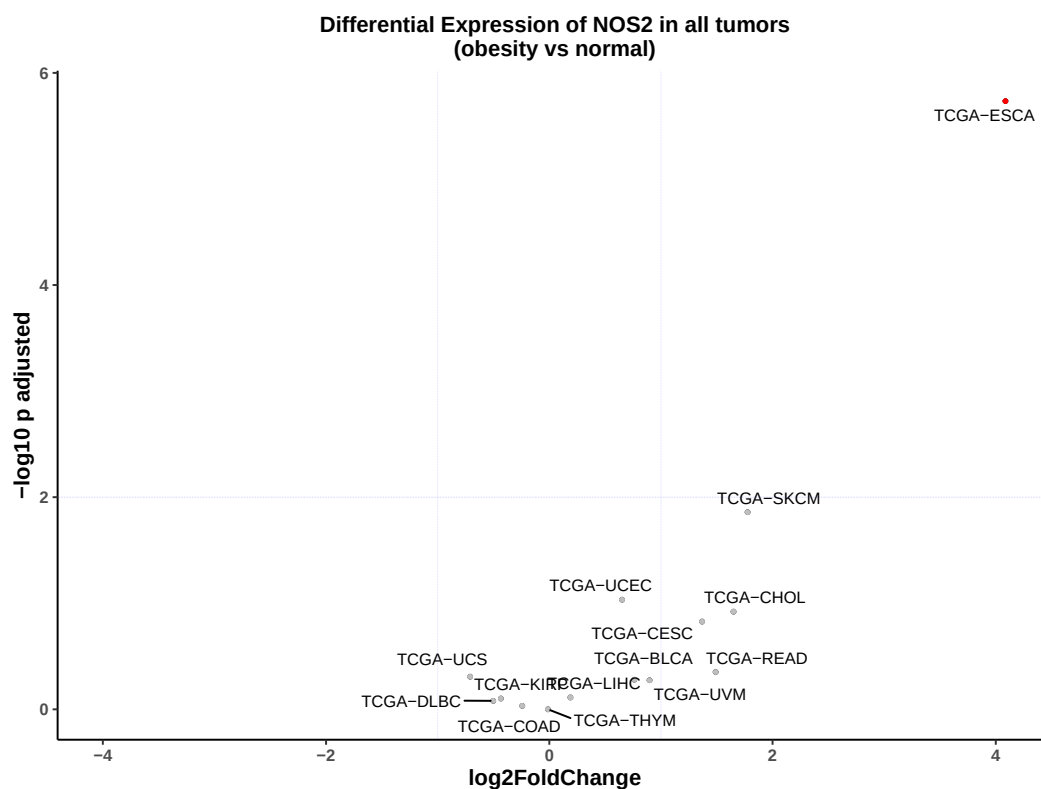


Figura 45 – *Fold change* do gene NOS2 em outros tumores do TCGA.

Além disso, quando comparados os valores de metilação, *i.* e *β-values*, do gene NOS2 nos outros tumores do TCGA, vemos claramente que essa desregulação do padrão de metilação é exclusiva do TCGA-ESCA (**Figura 46**). Sendo assim, concluímos que a desregulação desse gene à nível de expressão e metilação, dentre os tumores comparados, é exclusiva de esôfago.

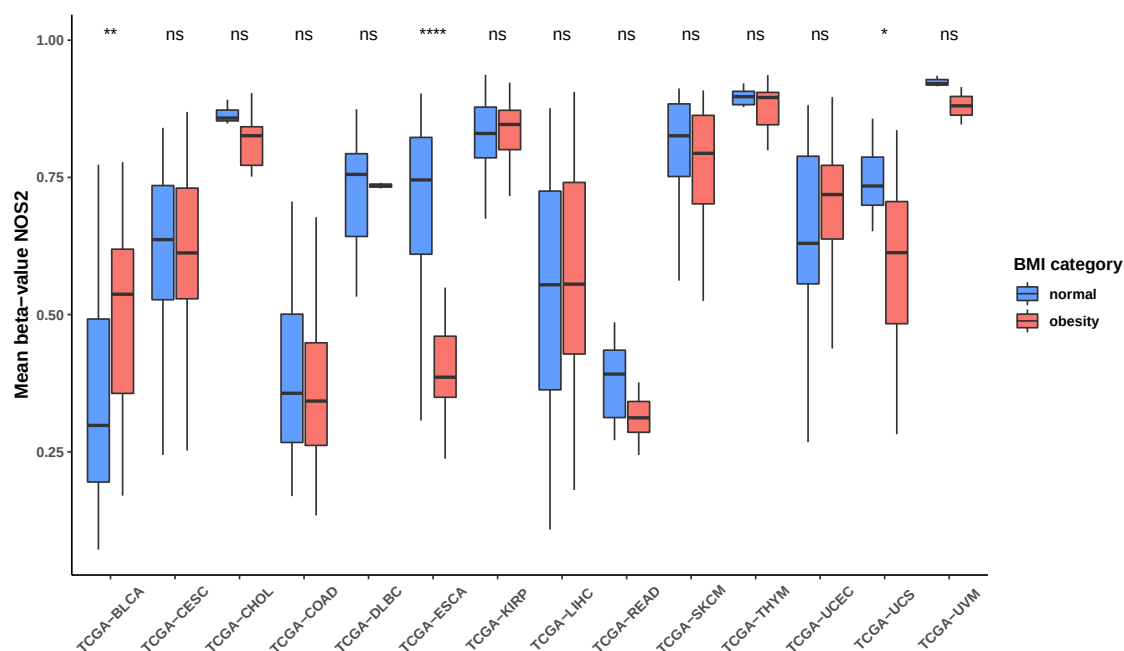


Figura 46 – Média dos β -values do gene NOS2 em outros tumores do TCGA.

4.11 SOBREVIDA GLOBAL

Foi realizada uma análise de sobrevida global, utilizando o estimador *Kaplan Meier* (Goel et al. 2010), para os pacientes do TCGA-ESCA de acordo com a categoria de IMC e também com os níveis de exposição às assinaturas mutacionais SBS17a e SBS17b. Foi utilizado o pacote *maxstat* para definir os grupos *high* e *low* de acordo com a exposição às assinaturas, onde o ponto de corte ótimo foi próximo à mediana, então pacientes com exposição às assinaturas SBS17a e SBS17b $\geq$ mediana foram definidos como *high* e os outros pacientes foram definidos como *low*, para a respectiva assinatura. A comparação entre as curvas foi realizada pelo teste Log-Rank.

Foi possível observar que as condições de sobrepeso e obesidade não apresentam diferenças significativas ($p=0.35$) na probabilidade de sobrevida dos pacientes em até 120 meses (10 anos) (**Figura 47**).

Além disso, observa-se que altos níveis de exposição à assinatura SBS17a (*high*) não tem impacto significativo ($p\text{-value}=0.066$) na probabilidade de sobrevida dos pacientes em até 10 anos, porém mostra uma tendência de piora na probabilidade de sobrevida comparados com baixos níveis de exposição (*low*) (**Figura 48**).

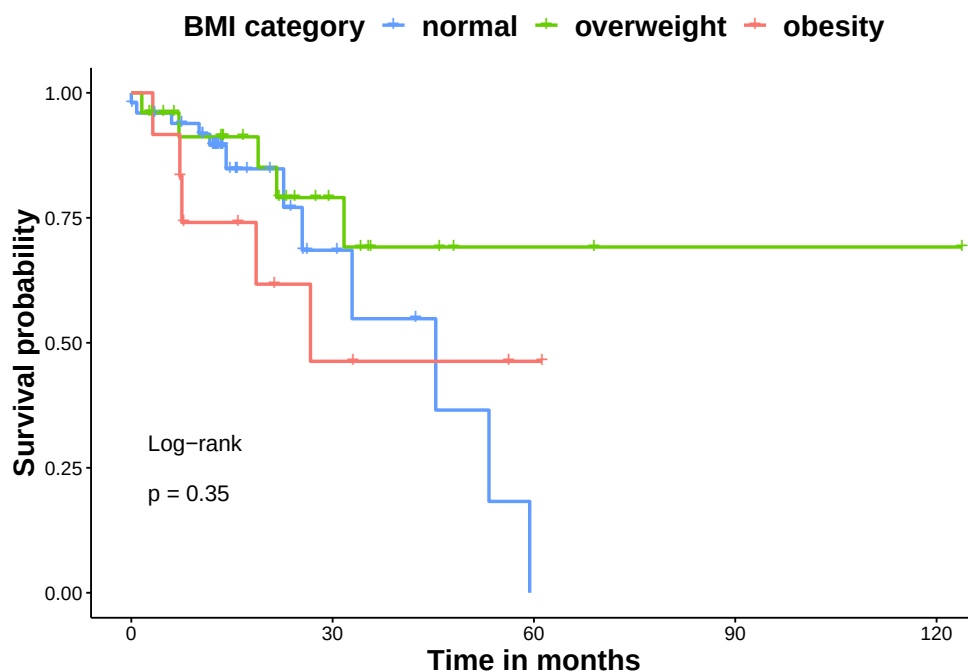


Figura 47 – Sobrevida global em até 10 anos para pacientes da coorte do TCGA-ESCA e IMC classificado como normal ou obesidade.

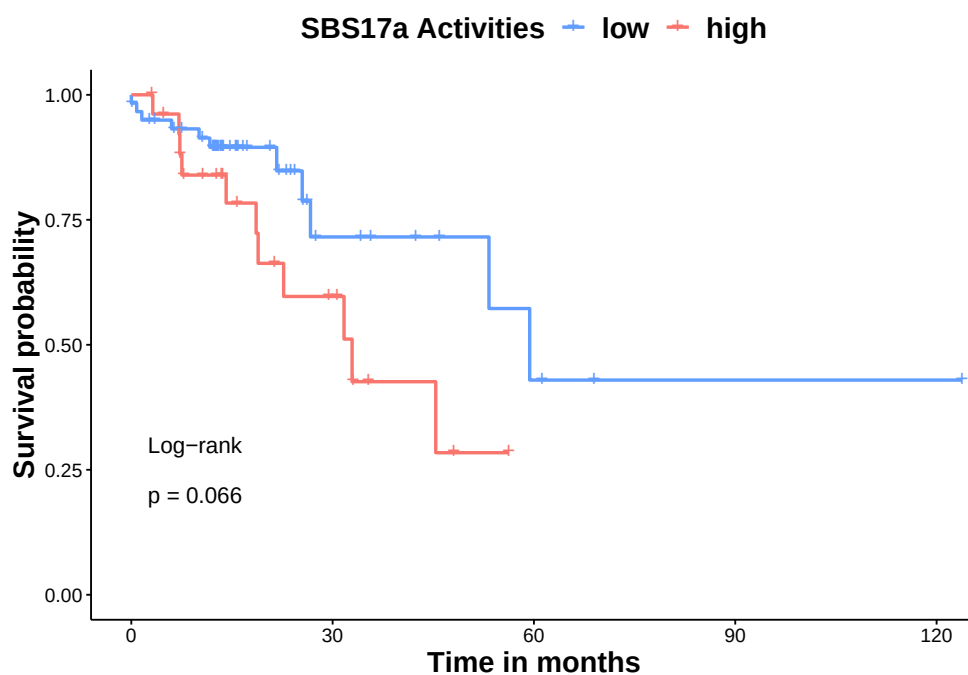


Figura 48 – Sobrevida global em até 10 anos para pacientes da coorte do TCGA-ESCA e níveis *high* e *low* de exposição à assinatura SBS17a.

Por outro lado, pacientes com altos níveis de exposição à assinatura SBS17b (*high*) mostram claramente uma piora significativa na probabilidade de sobrevida ($p\text{-value}=0.018$), comparados aos pacientes com baixo níveis de exposição à assinatura (*low*) (**Figura 49**).

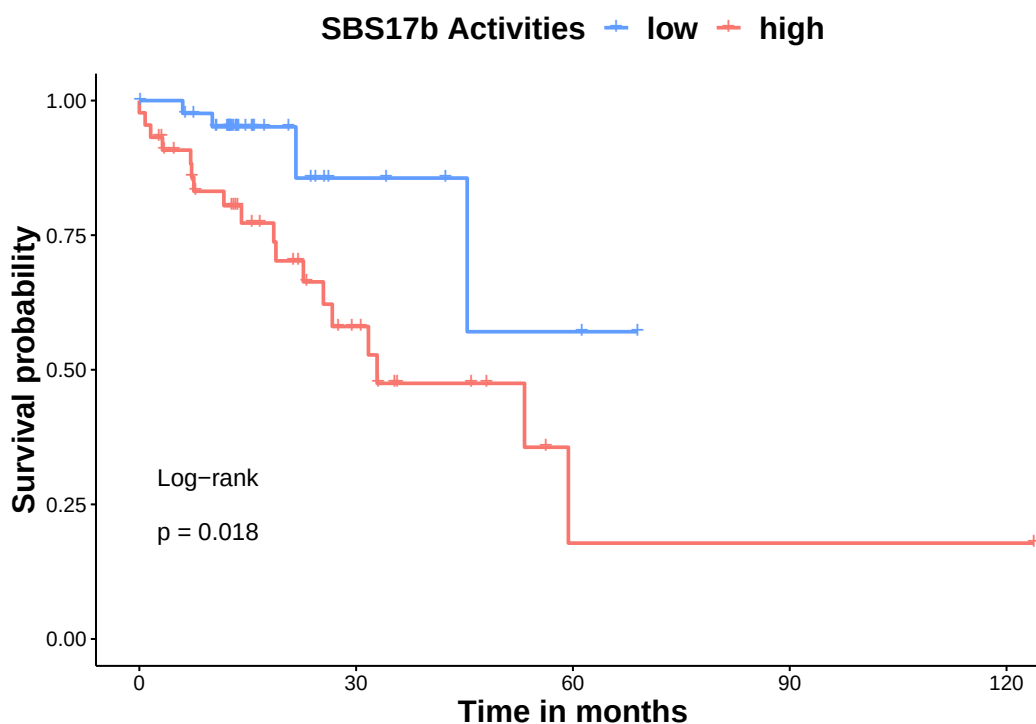


Figura 49 – Sobrevida global em até 10 anos para pacientes da coorte do TCGA-ESCA e níveis *high* e *low* de exposição à assinatura SBS17b.

A seguir, foi realizada uma regressão de Cox multivariada, ajustada por idade, gênero, categoria de IMC, níveis de exposição à assinatura SBS17b, tipo histológico do tumor, e estadiamento patológico (**Figura 50**). Como resultado, observa-se que no modelo múltiplo, quando ajustada às outras variáveis clínicas mencionadas, altos níveis de exposição à assinatura SBS17b (*high*) aumentam em quase 8 vezes a razão de risco (HR, *Hazard Ratio*) de óbito do paciente, HR = 7.70, 95% IC (2.05 - 25.05), $p = 0.002$, quando comparados com pacientes com baixo nível de exposição à assinatura (*low*). Além disso, foi construído um modelo multivariado com as mesmas variáveis clínicas já mencionadas, porém sem os níveis de exposição à assinatura SBS17b e outro modelo multivariado incluindo esta variável, mencionado acima (**Figura 50**). Esses dois modelos foram comparados pelo teste ANOVA, que mostrou significância estatística ($p = 0.014$), ou seja, quando adicionamos a variável de níveis de exposição à assinatura SBS17b ao modelo múltiplo, de fato esta variável se mostra como um fator prognóstico negativo melhor do que as variáveis clínicas inseridas no modelo, incluindo subtipo histológico e o próprio estadiamento do tumor.

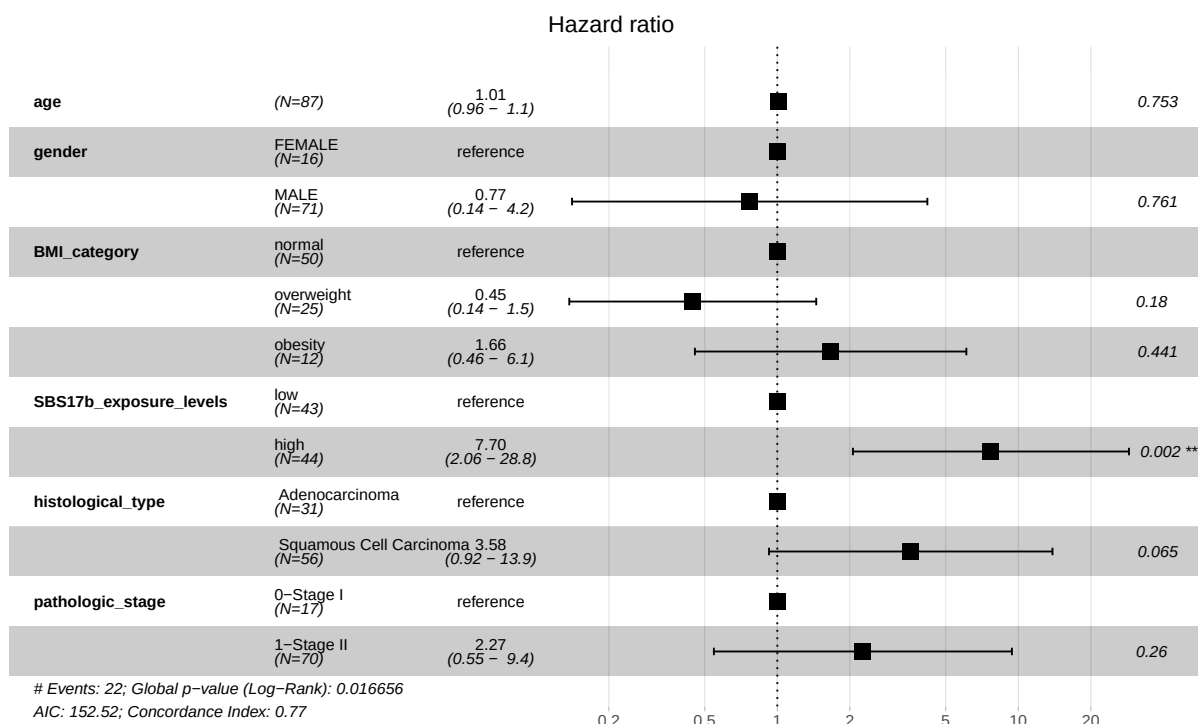


Figura 50 – Regressão de Cox múltipla para as variáveis selecionadas. *Hazard Ratio*, (HR) razão de riscos; AIC, *Akaike Information Criterion* (Critério de Informação de Akaike). A última coluna representa o *p-value*.

4.12 AMOSTRAS DO AC CAMARGO CANCER CENTER

A fim de validar os resultados de assinatura mutacional descritos anteriormente em uma coorte distinta, foi realizada uma análise mutacional com os pacientes participantes do projeto temático de adenocarcinoma gástrico do A.C. Camargo Cancer Center. Foram selecionados apenas aqueles com tumor em localização anatômica similar ao TCGA-ESCA, que compreende à localização "Esôfago distal".

Na **Tabela 4** abaixo, podemos observar como estão distribuídos os pacientes da coorte do A.C Camargo Cancer Center, em termos de idade, gênero, classificação de IMC e estadiamento do tumor.

Tabela 4 – Informações clínicas de pacientes com adenocarcinoma gástrico do A.C. Camargo Cancer Center.

Variável	Subtipo Esôfago distal
Idade (anos)	
Média	62.29
Mediana	62.0
Amplitude	28-81
Gênero	
Feminino	5 (17.85%)
Masculino	23 (82.14%)
IMC (Kg/m²)	
Média	26.46
Mediana	27.11
Amplitude	19.92-35.50
Categoria de IMC	
Normal	11 (39.28%)
Sobrepeso	14 (50%)
Obesidade	3 (10.71%)
Estadiamento patológico	
I	3 (10.71%)
II	8 (28.57%)
–	17 (60.71%)

Como resultado, também foi detectada nessa coorte a presença da assinatura S3, com similaridade dos cossenos de 0.91 com a assinatura SBS17b do COSMIC v3.2, a partir do método signeR (**Figura 51**). No entanto, ao comparar a exposição a essa assinatura entre os diferentes grupos de IMC, é possível observar que a mediana dos valores de assinatura dos pacientes obesos é maior que a mediana dos demais pacientes, porém não foi detectada significância estatística ($p\text{-value} > 0.05$) (**Figura 52**). Tal fator pode ser facilmente explicado devido ao baixo número amostral (n=28).

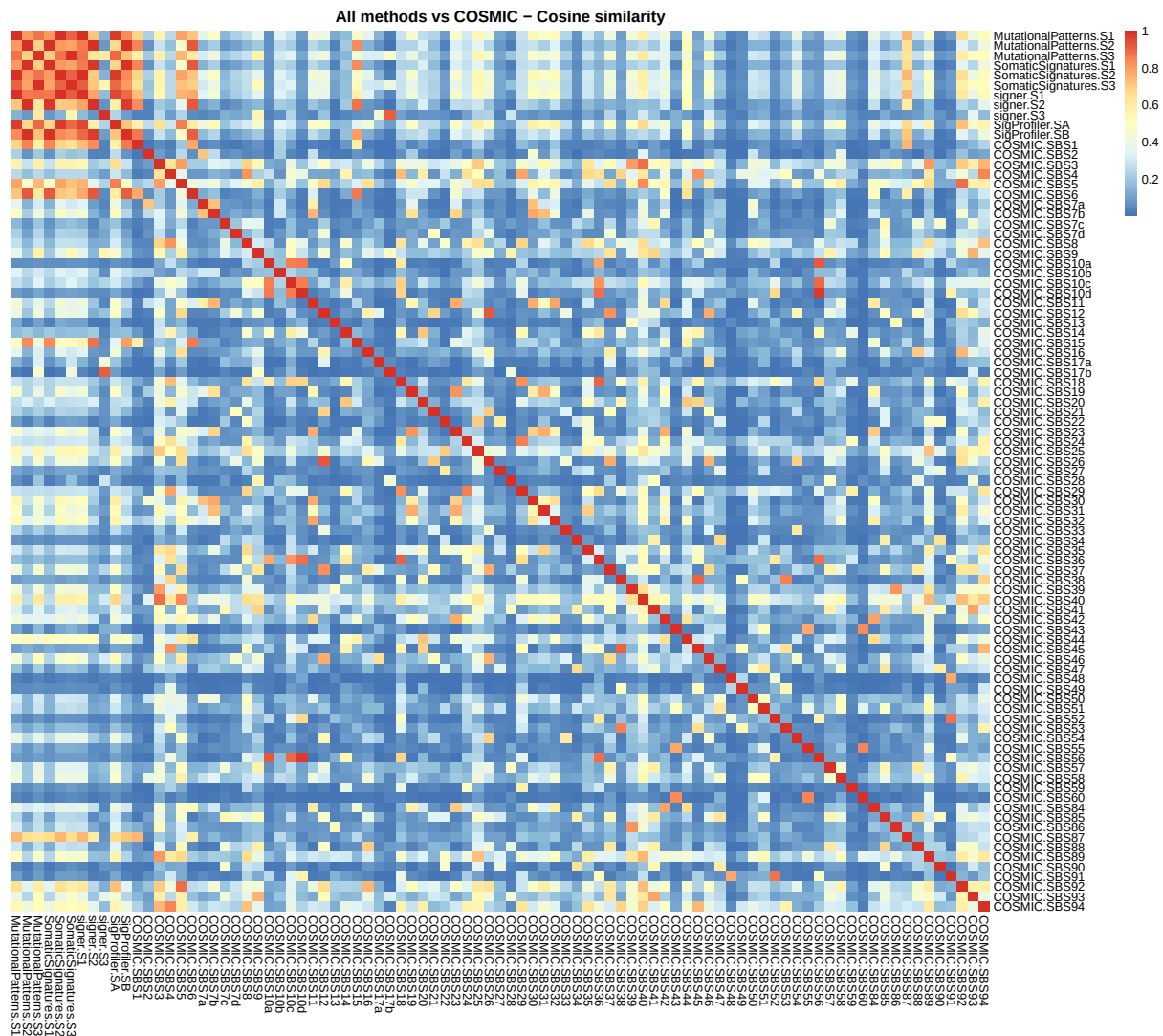


Figura 51 – Similaridade dos cossenos entre as assinaturas mutacionais extraídas por todos os métodos *de novo* e as assinaturas SBS descritas no COSMIC v3.2.

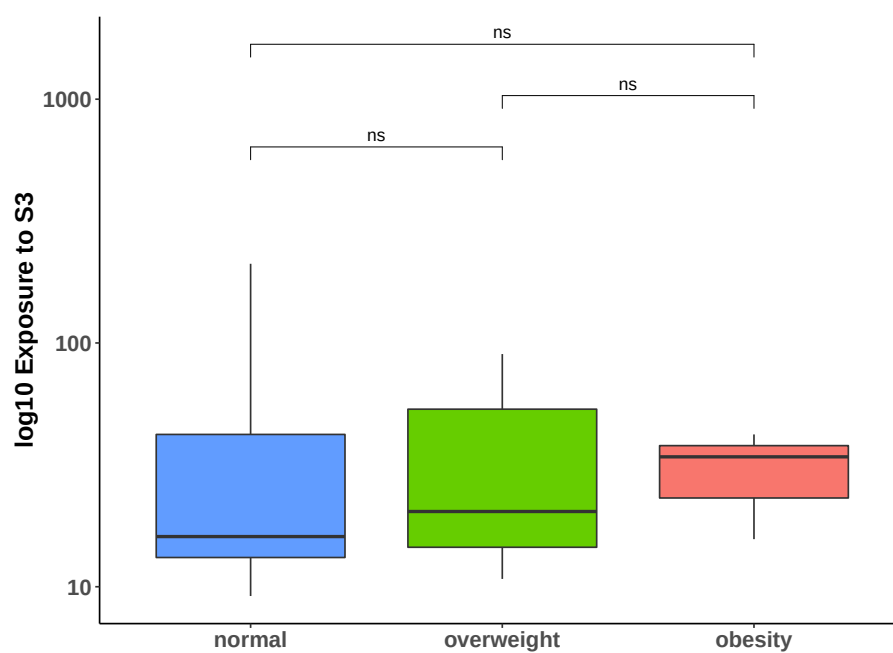


Figura 52 – Níveis de exposição dos pacientes do A.C. Camargo à assinatura S3 na escala $\log_{10}$. *ns*, não-significativo.

5 DISCUSSÃO

Neste estudo, conduzimos análises integrativas numa abordagem *multi-omics*, onde avaliamos mutações obtidas a partir de chamada de variantes de exoma, também avaliamos expressão gênica obtidas a partir de RNA-Seq e dados de metilação obtidos a partir de Bisulfite-Seq seguido por *Methylation Array*. Essas análises foram realizadas com o intuito de identificar o impacto do sobrepeso e obesidade nos mecanismos moleculares adjacentes à inflamação, estresse oxidativo e instabilidade do genoma associados à tumorigênese, e como esses fatores podem afetar o prognóstico de pacientes oncológicos.

Incluímos em nosso estudo os pacientes com tumor em estágio inicial (*I* e *II*), uma vez que os pacientes em estágio avançado são mais propensos à perda de peso causada por disfagia e/ou caquexia tumoral decorrente do avanço da doença. Destes, apenas incluímos os pacientes sem tratamento neoadjuvante, ou seja, sem tratamento *a priori*. Destacamos que de acordo com nossos critérios, temos pouco espaço para erro na classificação do IMC decorrente de perda de peso relacionada ao câncer, e também eliminamos algum possível viés relacionado a interpretação da etiologia de assinaturas mutacionais decorrente de tratamento com quimioterápicos.

Nós realizamos uma análise ortogonal de um conjunto de assinaturas relacionadas à mecanismos de reparo e também a danos oxidativos, e dentre esse conjunto identificamos duas assinaturas mutacionais, SBS17a e SBS17b, associadas ao IMC dos pacientes da coorte do TCGA-ESCA (**Figuras 11-12**). Conforme descrito na literatura, essas assinaturas mutacionais foram detectadas nos tumores de adenocarcinoma de esôfago (Secrier et al. 2016) e estômago (Lin et al. 2020). A **Figura 53**, disponível no COSMIC v3.2, ilustra o número de mutações (por megabase) em amostras humanas de câncer, onde as assinaturas SBS17a e SBS17b estão presentes, respectivamente.

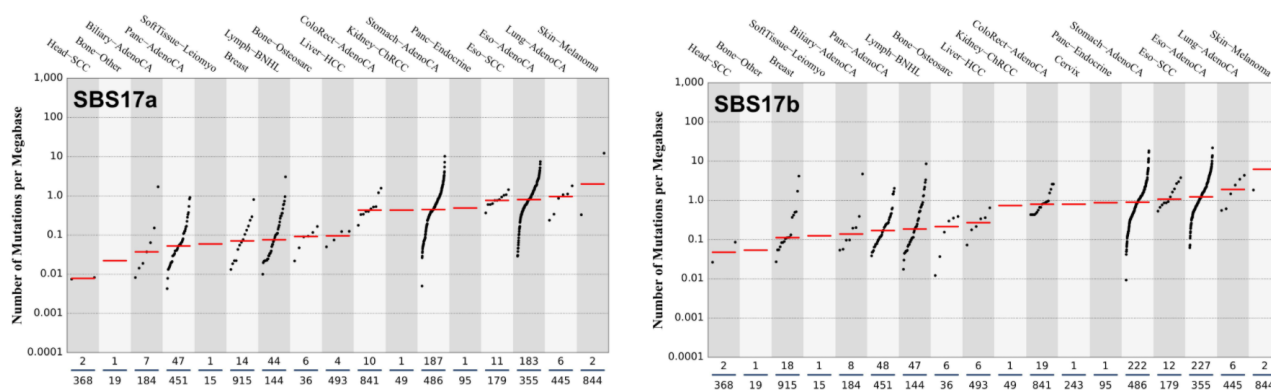


Figura 53 – Número de mutações por megabase atribuídas a cada assinatura mutacional onde a assinatura está presente.

A partir dessas figuras, é possível observar que, de fato o número de mutações associadas a essas assinaturas mutacionais são maiores em adenocarcinoma de esôfago e estômago, e também podem ocorrer em menor quantidade em linfoma de células B e adenocarcinoma de pâncreas, e numa frequência muito baixa em tumores de cabeça e pescoço, mama, adenocarcinoma biliar, osteosarcoma, fígado, colorretal, carcinoma de células escamosas de esôfago, pulmão e pele. Alguns desses tumores faziam parte da nossa coorte inicial, incluindo esôfago, linfoma de células B, fígado, colorretal e pele, e em parte, podemos justificar a detecção das assinaturas SBS17a e SBS17b apenas no carcinoma de esôfago e no linfoma de células B e não nos outros tumores devido a esse fato.

Um outro ponto interessante é que, a partir de associações estatísticas, alguns estudos recentes têm sugerido que, nos tumores onde essas assinaturas foram detectadas (esôfago e estômago), o dano oxidativo causado por ROS pode ser uma possível etiologia para essas assinaturas (Nones et al. 2014; Tomkova et al. 2018; Poetsch 2020). Já é conhecido que a obesidade pode levar à produção de ROS (Huizinga et al. 2011; Usman e Volpi 2018; Korac et al. 2021), e no nosso estudo mostramos que o IMC elevado, comum a indivíduos com sobrepeso e obesos, está associado ao aumento de mutações características das assinaturas SBS17a e SBS17b, ligando a produção de ROS decorrente da condição de obesidade ao dano no DNA.

Em particular, uma outra possível etiologia da assinatura SBS17b é o dano induzido pelo tratamento com o quimioterápico 5-Fluorouracil (Christensen et al. 2019), mas ressaltamos que com o intuito de evitar esse viés de interpretação, apenas utilizamos pacientes sem tratamento neoadjuvante, conforma descrito na seção **5. Materiais e Métodos**, subtópico **Crítérios de Exclusão**.

Complementando nossos resultados mencionados acima, realizamos extração de assinaturas mutacionais *de novo* para o tumor TCGA-ESCA, nesse caso sem pré-selecionar um conjunto de assinaturas específicas, e foi detectada, dentre outras, a assinatura mutacional S3 pelo signeR (**Figura 15**), que possui alta similaridade com a assinatura SBS17b do COSMIC v3.2. A assinatura mutacional S3, concordantemente também mostra uma associação com o IMC dos pacientes (**Figura 17**).

No COSMIC v3.2, além das assinaturas SBS17a e SBS17b, existem outras duas assinaturas mutacionais relacionadas ao estresse oxidativo, especificamente SBS18 e SBS36. Essas assinaturas surgem devido ao dano por ROS associado a falha no reparo pela via BER, devido a mutações germinativas bialélicas ou somáticas no gene MUTYH (Viel et al. 2017; Pilati et al. 2017). Observamos que as assinaturas SBS18 e SBS36 não mostram correlação significativa com o IMC dos pacientes do TCGA-ESCA (**Figura 10**) e, analisando mutações somáticas na nossa coorte, observamos que apenas 4 pacientes possuem mutações no gene MUTYH. Esses resultados nos indicam que, apesar de estarem descritas outras assinaturas mutacionais causadas por danos oxidativos, como SBS18 e SBS36, elas não estão associadas ao sobrepeso e obesidade, sendo geradas por mecanismos distintos.

Descrevemos o perfil imune de infiltrados inflamatórios em pacientes do TCGA-ESCA, onde mostramos um aumento significativo de células B *naïve*, e plasmáticas, células T CD4 de memória ativadas, células T *helper*, e neutrófilos, e uma diminuição significativa de macrófagos M0 e células dendríticas em repouso em pacientes com sobrepeso e obesos em relação aos normais (**Figura 22**). Inclusive, a infiltração dessas células inflamatórias mencionadas está também associada às assinaturas mutacionais SBS17a e SBS17b (**Figura 23**), mostrando que há uma relação entre o excesso de IMC, a infiltração dessas células imune e danos ao DNA descritos nos padrões das assinaturas SBS17a e SBS17b.

Avaliamos a expressão de genes marcadores de inflamação dessa mesma coorte do TCGA-ESCA, onde mostramos que os genes IL17F, NOS2, IL12B e STAT6 tendem a ter expressão aumentada (**Figura 24**), ao passo que o gene anti-inflamatório TGFB1 tem expressão diminuída em pacientes obesos e com sobrepeso em relação aos normais (**Figura 26**). Destes, os genes NOS2 e TGFB1 também estão diferencialmente expressos em obesos, com $\log_2 \text{Fold Change} = 4.87$ ($p\text{-value}$ ajustado < 0.0001) e $\log_2 \text{Fold Change} = -1.40$ ($p\text{-value}$ ajustado $= 0.002$), respectivamente.

Além disso, esses genes pró- e anti-inflamatórios também estão associados às assinaturas SBS17a e SBS17b (**Figuras 25-27**). Assim, esses resultados nos indicam que a obesidade e sobrepeso estão associadas à desregulação da expressão de genes pró- e anti-inflamatórios, que consequentemente estão associados à desregulação da infiltração de células imune e à presença das assinaturas mutacionais SBS17a e SBS17b.

Adicionalmente, investigamos as vias de reparo BER, NER e MMR, principais vias responsáveis por reparar danos oxidativos. Comparando a expressão de genes descritos para cada via, observamos que na via BER, os genes PARP2, APEX1 e NEIL3 têm expressão diminuída, enquanto que os genes OGG1 e NEIL1 têm aumento de expressão em pacientes com sobrepeso e obesos em relação aos normais (**Figura 28**). Destes genes, também estão diferencialmente expressos PARP2 ($\log_2 \text{Fold Change} = -0.5$, $p\text{-value}$ ajustado = 0.03), NEIL3 ($\log_2 \text{Fold Change} = -0.89$, $p\text{-value}$ ajustado = 0.03) e NEIL1 ($\log_2 \text{Fold Change} = 1.17$, $p\text{-value}$ ajustado < 0.001). Os genes mencionados também mostram associação com a assinatura SBS17b (**Figura 29**).

Na via NER, mostramos que os níveis de expressão dos genes RPA1, MNAT1, GTF2H3, DDB2 e XAB2 estão diminuídos, enquanto que nos genes UVSSA, RPA4 e ERCC5 estão aumentados nos pacientes obesos e com sobrepeso em relação aos normais (**Figura 30**). Destes, estão diferencialmente expressos os genes RPA1 ($\log_2 \text{Fold Change} = -0.67$, $p\text{-value}$ ajustado = 0.01), MNAT1 ($\log_2 \text{Fold Change} = -1.10$, $p\text{-value}$ ajustado < 0.001) e UVSSA ($\log_2 \text{Fold Change} = 0.93$, $p\text{-value}$ ajustado = 0.001). Esses genes também estão correlacionados com a assinatura SBS17b (**Figura 31**).

Na via MMR, o gene MSH6 mostra expressão diminuída em pacientes com sobrepeso e obesos em relação aos normais (**Figura 32**), também está diferencialmente expresso, com $\log_2 \text{Fold Change} = -0.71$, $p\text{-value}$ ajustado = 0.004, e mostra correlação negativa com a assinatura SBS17b (**Figura 33**). Esses resultados nos indicam que a obesidade e sobrepeso estão associadas à desregulação de diversos genes em vias de reparo, como BER, NER e MMR, que consequentemente estão associados aos danos resultantes da assinatura mutacional SBS17b.

Avaliamos também o perfil de metilação dos promotores de genes diferencialmente expressos, onde identificamos alterações nos padrões de metilação de forma global em pacientes obesos, uma vez que estes apresentam valores de metilação, *i. e.* $\beta\text{-value}$ significativamente maiores do que pacientes normais para a maioria dos genes (**Figuras 38-39**).

Além disso, investigamos especificamente genes envolvidos com processos inflamatórios, onde identificamos uma hipo-metilação do promotor do gene NOS2 (*Nitric Oxide Synthase 2*) (**Figura 41**), responsável pela produção do radical óxido nítrico, ao passo que observamos um aumento de aproximadamente 4 *fold* em sua expressão em pacientes obesos (**Figura 24**).

Como mostramos anteriormente que a expressão do gene NOS2 está associada à assinatura SBS17b (**Figura 25**), também investigamos possíveis associações entre seus níveis de metilação e a assinatura SBS17b, onde observamos que os valores de metilação do NOS2 e os níveis de exposição à assinatura SBS17b estão inversamente correlacionados (**Figura 43**). Isso nos indica que conforme maiores os níveis de exposição à assinatura SBS17b, menores os β -values do NOS2. De fato, por meio de uma regressão logística mostramos que a assinatura SBS17b aumenta em 1.141 vezes a chance de hipo-metilação do promotor do NOS2 (**Tabela 3**). Sendo assim, propomos que a hipo-metilação do promotor do NOS2 em pacientes obesos com carcinoma de esôfago pode ser considerada um substituto para a detecção da assinatura mutacional SBS17b.

Ao comparar os padrões de expressão e metilação do gene NOS2 nos outros tumores do TCGA descritos neste estudo, observamos que a *up*-regulação e hipo-metilação do promotor desse gene em pacientes obesos, dentre os tumores comparados, só ocorre em carcinoma de esôfago (TCGA-ESCA), sendo assim tumor-específica (**Figuras 45-46**).

Sabe-se que a obesidade é fator de risco para diversos tipos de câncer, conforme mostra a **Tabela 1**, e podemos observar que o risco relativo é diferente entre homens e mulheres. Além disso, são bem descritas as associações entre a idade e o risco de câncer, sendo os indivíduos mais velhos mais propensos ao desenvolvimento de alguns tipos de câncer (Calcinotto et al. 2019).

Mostramos num modelo de regressão de Cox multivariado ajustado por variáveis clínicas como idade, gênero e tipo histológico, que altos níveis de exposição à assinatura SBS17b aumentam em quase 5 vezes o risco de óbito dos pacientes do TCGA-ESCA, sendo um fator prognóstico negativo melhor do que as variáveis clínicas mencionadas e o próprio estadiamento do tumor (**Figura 50**).

Por fim, com o objetivo de identificar mecanismos recorrentes e validar as associações propostas nesse trabalho entre o excesso de peso corporal e danos ao DNA, detectados pela presença das assinaturas SBS17a e SBS17b na coorte do TCGA-ESCA, realizamos a análise de assinaturas mutacionais por meio do nosso *pipeline* para a coorte de adenocarcinoma gástrico do A.C. Ca-

margo Cancer Center, utilizando as amostras obtidas de tumor de esôfago. De forma similar ao TCGA-ESCA, também foi detectada a presença da assinatura mutacional SBS17b (**Figura 51**), que mostra uma tendência a ser mais alta em pacientes obesos (**Figura 52**). No entanto, não obtivemos significância estatística ($p\text{-value} > 0.05$) devido ao baixo número amostral ($n=33$).

Assim como boa parte dos estudos oncológicos, destacamos a limitação do número amostral, tanto dos pacientes do TCGA quanto do A.C. Camargo Cancer Center, que permitiria realizar testes estatísticos mais robustos para consolidar e replicar nossos achados.

Além disso, apontamos que há uma limitação em utilizar o IMC como única medida de obesidade e sobrepeso que, apesar de ser uma estimativa robusta da gordura corporal, pode superestimar a gordura corporal em atletas e outras pessoas com ganho de massa muscular, por exemplo. Outro ponto é que o IMC também pode subestimar a gordura corporal em pessoas mais velhas e pessoas com perda de massa muscular. Sendo assim, o ideal seria aliar a medida de IMC com medidas antropométricas, como a circunferência da cintura, que ajudaria a rastrear a localização da gordura e possíveis riscos à saúde de acordo com essa informação (National Institutes of Health 2020). No entanto, não contamos com essas medidas nos bancos de dados obtidos, sendo assim somente possível investigar o IMC.

Inclusive, destaca-se outra limitação pertinente ao uso do IMC, que seria seu cálculo somente ao diagnóstico. Dessa forma, não é possível inferir a classificação de IMC antes do surgimento do tumor. O ideal seria aliar informações de IMC ao diagnóstico e também em períodos anteriores à formação do câncer. Porém, não contamos com esse tipo de informação para a coorte do TCGA.

Para futuras direções deste estudo, ressaltamos a importância de decifrar o perfil imunológico encontrado em pacientes obesos com carcinoma de esôfago, e validar a abundância dessas células experimentalmente.

Um outro ponto interessante seria replicar as análises conduzidas neste trabalho em outras coortes robustas de carcinoma de esôfago, onde os pacientes possuam informação de IMC, e possivelmente, amostras pareadas de tecido adjacente não tumoral para que seja possível comparar pacientes obesos com e sem tumor.

Além disso, como mostramos uma desregulação global do perfil de metilação e associações dos níveis de metilação dos promotores de alguns genes com a assinatura SBS17b, também é inte-

ressante avaliar mutações similares aos padrões mutacionais dessa assinatura na região promotora desses genes. Inclusive está descrita na literatura que mutações somáticas podem alterar a atividade dos promotores de genes em um contexto tumoral, e com isso foi descrito o pacote *proActiv*, disponível no Bioconductor, para avaliar a atividade de promotores a partir de dados de RNA-Seq (Demircioğlu et al. 2019). Nesse contexto, além da metilação abre-se a oportunidade de investigar diferentes marcas epigenéticas como acetilações e modificações pós-traducionais relacionadas à obesidade.

Também, faz-se necessário investigar genes diferencialmente expressos envolvidos em vias de sinalização de leptina e insulina, conhecidas por estarem desreguladas em pacientes obesos, bem como possíveis mutações germinativas nesses genes.

Para as amostras do A.C. Camargo, uma vez que foram coletadas informações referente a dieta dos pacientes para esse estudo, uma alternativa também seria associar os componentes da dieta e obesidade com o prognóstico dos pacientes e também com a atividade mutacional.

6 CONCLUSÕES

Este estudo evidencia, por diferentes níveis moleculares, como a obesidade e sobrepeso impactam a regulação de genes com importantes funções biológicas em um ambiente tumoral, incluindo genes de reparo, genes anti- e pró-inflamatórios, que também estão associados com a desregulação da infiltração de células do sistema imunológico.

Desenvolvemos um *workflow* para análise e detecção de assinaturas mutacionais somáticas, pelo qual identificamos associações positivas entre obesidade e sobrepeso e danos ao DNA por meio da detecção da presença das assinaturas mutacionais SBS17a e SBS17b no exoma tumoral, cuja etiologia remete ao dano causado por ROS.

Analisando padrões de metilação de genes pró-inflamatórios observamos uma associação direta entre os níveis de presença da assinatura SBS17b e a hipo-metilação do promotor do gene NOS2, envolvido na produção do radical óxido nítrico. Sendo assim, concluímos que a assinatura SBS17b está diretamente associada à hipo-metilação desse gene, e portanto, sua hipo-metilação pode ser considerada um substituto para a detecção da assinatura SBS17b.

Por fim, mostramos em pacientes com tumor em estadiamento inicial, que altos níveis de exposição à assinatura SBS17b diminui a probabilidade de sobrevida, sendo um fator prognóstico pior que o próprio estadiamento do tumor.

7 REFERÊNCIAS

- 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; 578:94–101.
- Alexandrov LB, Nik-Zainal S, Wedge DC, Campbell PJ, Stratton MR. Deciphering Signatures of Mutational Processes Operative in Human Cancer. *Cell Reports* 2013a; 3:246–259.
- Alexandrov LB, Nik-Zainal S, Wedge DC, Aparicio SA, Behjati S, Biankin AV, et al. Signatures of mutational processes in human cancer. *Nature* 2013b; 500:415–421.
- AMERICAN ASSOCIATION FOR CANCER RESEARCH. *Cancer Progress Report Graphics*. 2019. Disponível em: <https://www.cancerprogressreport.org/>, Acesso em: 31 jul. 2019 .
- Avgerinos KI, Spyrou N, Mantzoros CS, Dalamaga M. Obesity and cancer risk: Emerging biological mechanisms and perspectives. *Metabolism: clinical and experimental* 2019; 92:121–135.
- Azzi A. Oxidative stress: A dead end or a laboratory hypothesis? *Biochemical and Biophysical Research Communications* 2007; 362:230–232.
- Baracos V, Martin L, Korc M, Guttridge D, Fearon K. Cancer-associated cachexia. *Nat Rev Dis Primers* 2018; 4:17105.
- Barregard L, Møller P, Henriksen T, Mistry V, Koppen G, Rossner P, et al. Human and Methodological Sources of Variability in the Measurement of Urinary 8-Oxo-7,8-dihydro-2-deoxyguanosine. *Antioxidants & Redox Signaling* 2013; 18:2377–2391.
- Bergstrom EN, Huang MN, Mahto U, Barnes M, Stratton MR, Rozen SG, et al. SigProfilerMatrix-Generator: A tool for visualizing and exploring patterns of small mutational events. *BMC Genomics* 2019; 20:1–12.
- Blokzijl F, Janssen R, van Boxtel R, Cuppen E. MutationalPatterns: Comprehensive genome-wide analysis of mutational processes. *Genome Medicine* 2018; 10:1–11.
- 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. *Int J Cancer* 2019;:.
- Brennan P, Wild CP. Genomics of Cancer and a New Era for Cancer Prevention. *PLoS Genet* 2015; 11:.
- Buttura JR, Provisor MN, Valieris R, Calsavara VF, Drummond RD, Defelicibus A, et al. Mutational signatures driven by epigenetic determinants stratify patients for therapeutic interventions in gastric cancer. *Cancers* 2020;:.
- Calcinotto A, Kohli J, Zagato E, Pellegrini L, Demaria M, Alimonti A. Cellular Senescence: Aging, Cancer, and Injury. *Physiol Rev* 2019; 2:1047–1078.
- Carbon S, Douglass E, Dunn N, Good B, Harris NL, Lewis SE, et al. The Gene Ontology Resource: 20 years and still GOing strong. *Nucleic Acids Research* 2019; 47:D330–D338.

- Castillo JJ, Ingham RR, Reagan JL, Furman M, Dalia S, Mitri J. Obesity is associated with increased relative risk of diffuse large B-cell lymphoma: a meta-analysis of observational studies. *Clinical lymphoma, myeloma & leukemia* 2014; 14:122–30.
- Cerdá C, Sánchez C, Climent B, Vázquez A, Iradi A, El Amrani F, et al. Oxidative Stress and DNA Damage in Obesity-Related Tumorigenesis. In: . [S.l.: s.n.], 2014. 824, 5–17.
- Chapman MA, Lawrence MS, Keats JJ, Cibulskis K, Sougnez C, Schinzel AC, et al. Initial genome sequencing and analysis of multiple myeloma. *Nature* 2011; 471:467–472.
- Chen Y, Wang X, Wang J, Yan Z, Luo J. Excess body weight and the risk of primary liver cancer: an updated meta-analysis of prospective studies. *European journal of cancer (Oxford, England : 1990)* 2012; 48:2137–45.
- Christensen S, Van der Roest B, Besselink N, Janssen R, Boymans S, Martens JW, et al. 5-Fluorouracil treatment induces characteristic T>G mutations in human cancer. *Nature Communications* 2019; 10:1–11.
- Cibulskis K, Lawrence MS, Carter SL, Sivachenko A, Jaffe D, Sougnez C, et al. Sensitive detection of somatic point mutations in impure and heterogeneous cancer samples. *Nature Biotechnology* 2013; 31:213–219.
- Codoñer-Franch P, Valls-Bellés V, Arilla-Codoñer A, Alonso-Iglesias E. Oxidant mechanisms in childhood obesity: The link between inflammation and oxidative stress. *Translational Research* 2011; 158:369–384.
- Colditz GA, Peterson LL. Obesity and Cancer: Evidence, Impact, and Future Directions. *Clinical chemistry* 2018; 64:154–162.
- Collaborative Group on Epidemiological Studies of Ovarian Cancer. Ovarian cancer and body size: individual participant meta-analysis including 25,157 women with ovarian cancer from 47 epidemiological studies. *PLoS medicine* 2012; 9:e1001200.
- Dale R, Grüning B, Sjödin A, Rowe J, Chapman BA, Tomkins-Tinch CH, et al. Bioconda: Sustainable and comprehensive software distribution for the life sciences. *Nature Methods* 2018; 15: 475–476.
- David SS, O’Shea VL, Kundu S. Base-excision repair of oxidative DNA damage. *Nature* 2007; 447:941–950.
- Demircioğlu D, Cukuroglu E, Kindermans M, Nandi T, Calabrese C, Fonseca NA, et al. A Pan-cancer Transcriptome Analysis Reveals Pervasive Regulation through Alternative Promoters. *Cell* 2019; 178:1465–1477.e17.
- Desaulniers D, Al-Mulla F, Al-Temaimi R, Amedei A, Azqueta A, Bisson WH, et al. Causes of genome instability: The effect of low dose chemical exposures in modern society. *Carcinogenesis* 2015; 36:S61–S88.
- Ellrott K, Bailey MH, Saksena G, Covington KR, Kandoth C, Stewart C, et al. Scalable Open Science Approach for Mutation Calling of Tumor Exomes Using Multiple Genomic Pipelines. *Cell Systems* 2018; 6:271–281.e7.
- Fernández-Sánchez A, Madrigal-Santillán E, Bautista M, Esquivel-Soto J, Morales-González Á, Esquivel-Chirino C, et al. Inflammation, oxidative stress, and obesity. *International Journal of Molecular Sciences* 2011; 12:3117–3132.

- Finucane MM, Stevens GA, Cowan MJ, Danaei G, Lin JK, Paciorek CJ, et al. National, regional, and global trends in body-mass index since 1980: Systematic analysis of health examination surveys and epidemiological studies with 960 country-years and 9.1 million participants. *The Lancet* 2011; 377:557–567.
- Fischer A, Illingworth CJ, Campbell PJ, Mustonen V. EMu: Probabilistic inference of mutational processes and their localization in the cancer genome. *Genome Biology* 2013; 14:.
- Font-Burgada J, Sun B, Karin M. Obesity and Cancer: The Oil that Feeds the Flame. *Cell Metabolism* 2016; 23:48–62.
- Fousteri M, Mullenders LH. Transcription-coupled nucleotide excision repair in mammalian cells: Molecular mechanisms and biological effects. *Cell Research* 2008; 18:73–84.
- Gao CL, Zhu C, Zhao YP, Chen XH, Ji CB, Zhang CM, et al. Mitochondrial dysfunction is induced by high levels of glucose and free fatty acids in 3T3-L1 adipocytes. *Molecular and Cellular Endocrinology* 2010; 320:25–33.
- Gaujoux R, Seoighe C. A flexible R package for nonnegative matrix factorization. *BMC Bioinformatics* 2010; 11:.
- Gehring JS, Fischer B, Lawrence M, Huber W. SomaticSignatures: Inferring mutational signatures from single-nucleotide variants. *Bioinformatics* 2015; 31:3673–3675.
- Goel MK, Kishore J, Khanna P, Cunningham. Understanding survival analysis: Kaplan-Meier estimate. *International Journal of Ayurveda Research* 2010; 1:274.
- Gomes VA, Casella-Filho A, Chagas AC, Tanus-Santos JE. Enhanced concentrations of relevant markers of nitric oxide formation after exercise training in patients with metabolic syndrome. *Nitric Oxide - Biology and Chemistry* 2008; 19:345–350.
- Goodwin PJ, Stambolic V. Impact of the obesity epidemic on cancer. *Annual review of medicine* 2015; 66:281–96.
- Graziewicz M, Wink DA, Laval F. Nitric oxide inhibits DNA ligase activity: Potential mechanisms for NO-mediated DNA damage. *Carcinogenesis* 1996; 17:2501–2505.
- Hainaut P, Pfeifer GP. Patterns of p53 G→T transversions in lung cancers reflect the primary mutagenic signature of DNA-damage by tobacco smoke. *Carcinogenesis* 2001; 22:367–374.
- Hanawalt PC, Spivak G. Transcription-coupled DNA repair: Two decades of progress and surprises. *Nature Reviews Molecular Cell Biology* 2008; 9:958–970.
- Haradhvala NJ, Polak P, Stojanov P, Covington KR, Shinbrot E, Hess JM, et al. Mutational Strand Asymmetries in Cancer Genomes Reveal Mechanisms of DNA Damage and Repair. *Cell* 2016; 164:538–549.
- Helleday T, Eshtad S, Nik-Zainal S. Mechanisms underlying mutational signatures in human cancers. *Nat Rev Genet* 2014; 15:585–98.
- Huber W, Carey VJ, Gentleman R, Anders S, Carlson M, Carvalho BS, et al. Orchestrating high-throughput genomic analysis with Bioconductor. *Nature Methods* 2015; 12:115–121.

- Huizinga T, Nigrovic P, Ruderman E, Schulze-Koops H. Mitochondrial reactive oxygen species promote production of proinflammatory cytokines and are elevated in TNFR1-associated periodic syndrome (TRAPS): Commentary. *International Journal of Advances in Rheumatology* 2011; 9: 108.
- Hutchins LN, Murphy SM, Singh P, Graber JH. Position-dependent motif characterization using non-negative matrix factorization. *Bioinformatics* 2008; 24:2684–2690.
- IBGE. *Pesquisa Nacional de Saúde 2019: Percepção do estado de saúde, estilos de vida, doenças crônicas e saúde bucal*. 2020. Disponível em: <https://www.ibge.gov.br/estatisticas/sociais/saude/9160-pesquisa-nacional-de-saude.html>, Acesso em: 15 out. 2020 .
- INTERNATIONAL AGENCY FOR RESEARCH ON CANCER. *Cancer attributable to obesity*. 2012. Disponível em: <https://gco.iarc.fr/causes/obesity>, Acesso em: 26 jul. 2019 .
- Jaiswal M, LaRusso NF, Nishioka K, Nakabeppu Y, Gores GJ. Erratum: Human Ogg1, a protein involved in the repair of 8-oxoguanine, is inhibited by nitric oxide. *Cancer Research* 2001; 61: 7705.
- Jiang G, Sancar A. Recruitment of DNA Damage Checkpoint Proteins to Damage in Transcribed and Nontranscribed Sequences. *Molecular and Cellular Biology* 2006; 26:39–49.
- Jinks-Robertson S, Bhagwat AS. Transcription-Associated Mutagenesis. *Annual Review of Genetics* 2014; 48:341–359.
- Karin M, Lin A. NF- κ B at the crossroads of life and death. *Nat Immunol* 2002; 3,3:221–227.
- Keleher MR, Zaidi R, Hicks L, Shah S, Xing X, Li D, et al. A high-fat diet alters genome-wide DNA methylation and gene expression in SM/J mice 06 Biological Sciences 0604 Genetics. *BMC Genomics* 2018; 19:1–13.
- Klungland A, Rosewell I, Hollenbach S, Larsen E, Daly G, Epe B, et al. Accumulation of premutagenic DNA lesions in mice defective in removal of oxidative base damage. *Proceedings of the National Academy of Sciences* 1999; 96:13300–13305.
- Korac B, Kalezić A, Pekovic-vaughan V, Korac A. Redox Biology Redox changes in obesity , metabolic syndrome , and diabetes. *Redox Biology* 2021;:101887.
- Köster J, Rahmann S. Snakemake-a scalable bioinformatics workflow engine. *Bioinformatics* 2012; 28:2520–2522.
- Lawrence MS, Stojanov P, Polak P, Kryukov GV, Cibulskis K, Sivachenko A, et al. Mutational heterogeneity in cancer and the search for new cancer-associated genes. *Nature* 2013; 499:214–218.
- Lichtman MA. Obesity and the Risk for a Hematological Malignancy: Leukemia, Lymphoma, or Myeloma. *The Oncologist* 2010; 15:1083–1101.
- Lin Y, Luo Y, Sun Y, Guo W, Zhao X, Xi Y, et al. Genomic and transcriptomic alterations associated with drug vulnerabilities and prognosis in adenocarcinoma at the gastroesophageal junction. *Nature Communications* 2020; 11:1–14.
- Liu J, Lichtenberg T, Hoadley KA, Poisson LM, Lazar AJ, Cherniack AD, et al. An Integrated TCGA Pan-Cancer Clinical Data Resource to Drive High-Quality Survival Outcome Analytics. *Cell* 2018; 173:400–416.e11.

- Love MI, Huber W, Anders S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biology* 2014; 15:1–21.
- Maura F, Degasperi A, Nadeu F, Leongamornlert D, Davies H, Moore L, et al. A practical guide for mutational signature analysis in hematological malignancies. *Nature Communications* 2019; 10:.
- McKenna A, Hanna M, Banks E, Sivachenko A, Cibulskis K, Kernytsky A, et al. The Genome Analysis Toolkit: A MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Research* 2010; 20:254–260.
- MINISTÉRIO DA SAÚDE. *Vigitel*. 2018. Disponível em: <http://www.saude.gov.br/noticias/agencia-saude/45612-brasileiros-atingem-maior-indice-de-obesidade-nos-ultimos-treze-anos>, Acesso em: 31 jul. 2019 .
- Mittal M, Siddiqui MR, Tran K, Reddy SP, Malik AB. Reactive oxygen species in inflammation and tissue injury. *Antioxidants and Redox Signaling* 2014; 20:1126–1167.
- Mostoslavsky R, Chua KF, Lombard DB, Pang WW, Fischer MR, Gellon L, et al. Genomic instability and aging-like phenotype in the absence of mammalian SIRT6. *Cell* 2006; 124:315–329.
- NATIONAL INSTITUTES OF HEALTH. *Assessing Your Weight and Health Risk*. 2020. Disponível em: https://www.nhlbi.nih.gov/health/educational/lose_wt/risk.htm, Acesso em: 05 jul. 2021 .
- 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; 12:453–457.
- Nishimura S, Manabe I, Nagasaki M, Eto K, Yamashita H, Ohsugi M, et al. CD8 + effector T cells contribute to macrophage recruitment and adipose tissue inflammation in obesity. *Nature Publishing Group* 2009;:.
- Nones K, Waddell N, Wayte N, Patch AM, Bailey P, Newell F, et al. Genomic catastrophes frequently arise in esophageal adenocarcinoma and drive tumorigenesis. *Nature Communications* 2014; 5: .
- Oltra AM, Carbonell F, Tormos C, Iradi A, Sáez GT. Antioxidant enzyme activities and the production of MDA and 8-oxo-dG in chronic lymphocytic leukemia. *Free Radical Biology and Medicine* 2001; 30:1286–1292.
- Ottobelli Chielle E, de Souza WM, da Silva TP, Moresco RN, Moretto MB. Adipocytokines, inflammatory and oxidative stress markers of clinical relevance altered in young overweight/obese subjects. *Clinical Biochemistry* 2016; 49:548–553.
- Parsi MA. Obesity and cholangiocarcinoma. *World journal of gastroenterology* 2013; 19:457–62.
- Pilati C, Shinde J, Alexandrov LB, Assié G, André T, Hélias-Rodzewicz Z, et al. Mutational signature analysis identifies MUTYH deficiency in colorectal cancers and adrenocortical carcinomas. *Journal of Pathology J Pathol* 2017; 242:10–15.
- Pleasance ED, Cheetham RK, Stephens PJ, McBride DJ, Humphray SJ, Greenman CD, et al. A comprehensive catalogue of somatic mutations from a human cancer genome. *Nature* 2010; 463: 191–196.

- Poetsch AR. The genomics of oxidative DNA damage, repair, and resulting mutagenesis. In: *Computational and Structural Biotechnology Journal*. [S.l.]: Elsevier B.V., 2020. 18, 207–219.
- Quail DF, Dannenberg AJ. The obese adipose tissue microenvironment in cancer development and progression. *Nature Reviews Endocrinology* 2019; 15:139–154.
- Regueira T, Andresen M, Bruhn A, Perez D, Strobel P, Dougnac A, et al. Lipoperoxidation and protein oxidative damage exhibit different kinetics during septic shock. *Mediators of Inflammation* 2008; 2008:.
- Renahan AG, Tyson M, Egger M, Heller RF, Zwahlen M. Body-mass index and incidence of cancer: a systematic review and meta-analysis of prospective observational studies. *Lancet (London, England)* 2008; 371:569–78.
- Rezende LFM, Arnold M, Rabacow FM, Levy RB, Claro RM, Giovannucci E, et al. The increasing burden of cancer attributable to high body mass index in Brazil. *Cancer Epidemiology* 2018; 54: 63–70.
- Ribeiro ML, Priolli DG, Miranda DD, Arçari DP, Pedrazzoli J, Martinez CA. Analysis of Oxidative DNA Damage in Patients with Colorectal Cancer. *Clinical Colorectal Cancer* 2008; 7:267–272.
- Roberts DL, Dive C, Renahan AG. Biological Mechanisms Linking Obesity and Cancer Risk: New Perspectives. *Annual Review of Medicine* 2010a; 61:301–316.
- Roberts RA, Smith RA, Safe S, Szabo C, Tjalkens RB, Robertson FM. Toxicological and pathophysiological roles of reactive oxygen and nitrogen species. *Toxicology* 2010b; 276:85–94.
- Rosales RA, Drummond RD, Valieris R, Dias-Neto E, Da Silva IT. signeR: An empirical Bayesian approach to mutational signature discovery. *Bioinformatics* 2017; 33:8–16.
- Rosenthal R, McGranahan N, Herrero J, Taylor BS, Swanton C. deconstructSigs: Delineating mutational processes in single tumors distinguishes DNA repair deficiencies and patterns of carcinoma evolution. *Genome Biology* 2016; 17:1–11.
- Russo MT, De Luca G, Degan P, Bignami M. Different DNA repair strategies to combat the threat from 8-oxoguanine. *Mutation Research - Fundamental and Molecular Mechanisms of Mutagenesis* 2007; 614:69–76.
- Rønningen T, Shah A, Reiner AH, Collas P, Øivind Moskaug J. Epigenetic priming of inflammatory response genes by high glucose in adipose progenitor cells. *Biochemical and Biophysical Research Communications* 2015; 467:979–986.
- Scoppola A, Strigari L, Barnabei A, Petasecca P, De Galitiis F, Fulgenzi CAM, et al. Insulin resistance as a risk factor for cutaneous melanoma. a case control study and risk-assessment nomograms. *Frontiers in Endocrinology* 2019; 10:757.
- Secrier M, Li X, De Silva N, Eldridge MD, Contino G, Bornschein J, et al. Mutational signatures in esophageal adenocarcinoma define etiologically distinct subgroups with therapeutic relevance. *Nature Genetics* 2016; 48:1131–1141.
- Setayesh T, Nersesyan A, Mišík M, Ferk F, Langie S, Andrade VM, et al. Impact of obesity and overweight on DNA stability: Few facts and many hypotheses. *Mutation Research - Reviews in Mutation Research* 2018; 777:64–91.

- Shang XL, Fan FH, Zhi JG, Nagai R, Wei RZ, Zhi QL, et al. Advanced oxidation protein products accelerate atherosclerosis through promoting oxidative stress and inflammation. *Arteriosclerosis, Thrombosis, and Vascular Biology* 2006; 26:1156–1162.
- Shibutani S, Takeshita M, Grollman AP. Insertion of specific bases during DNA synthesis past the oxidation-damaged base 8-oxodG. *Nature* 1991; 349:431–434.
- Slupphaug G, Kavli B, Krokan HE. The interacting pathways for prevention and repair of oxidative DNA damage. *Mutation Research - Fundamental and Molecular Mechanisms of Mutagenesis* 2003; 531:231–251.
- Spivak G, Ganesan AK. The complex choreography of transcription-coupled repair. *DNA Repair* 2014; 19:64–70.
- Stratton MR, Campbell PJ, Futreal PA. The cancer genome. *Nature* 2009; 458:719–724.
- Subramanian A, Tamayo P, Mootha VK, Mukherjee S, Ebert BL, Gillette MA, et al. Gene set enrichment analysis: A knowledge-based approach for interpreting genome-wide expression profiles. *Proceedings of the National Academy of Sciences of the United States of America* 2005; 102:15545–15550.
- Surmi B, Hasty A. The role of chemokines in recruitment of immune cells to the artery wall and adipose tissue. *Vascular Pharmacology* 2010; 52:27–36.
- Tamir S, Burney S, Tannenbaum SR. DNA damage by nitric oxide. *Chemical Research in Toxicology* 1996; 9:821–827.
- Tomkova M, Tomek J, Kriaucionis S, Schuster-Böckler B. Mutational signature distribution varies with DNA replication timing and strand asymmetry. *Genome Biology* 2018; 19:.
- Townsend K, Tseng YH. Brown adipose tissue. *Adipocyte* 2012; 1:13–24.
- Tyson J, Caple F, Spiers A, Burtle B, Daly AK, Williams EA, et al. Inter-individual variation in nucleotide excision repair in young adults: Effects of age, adiposity, micronutrient supplementation and genotype. *British Journal of Nutrition* 2009; 101:1316–1323.
- Usman M, Volpi EV. DNA damage in obesity: Initiator, promoter and predictor of cancer. *Mutation Research - Reviews in Mutation Research* 2018; 778:23–37.
- Viel A, Bruselles A, Meccia E, Fornasarig M, Quaia M, Canzonieri V, et al. A Specific Mutational Signature Associated with DNA 8-Oxoguanine Persistence in MUTYH-defective Colorectal Cancer. *EBioMedicine* 2017; 20:39–49.
- Wang Z, Rhee DB, Lu J, Bohr CT, Zhou F, Vallabhaneni H, et al. Characterization of oxidative guanine damage and repair in mammalian telomeres. *PLoS Genetics* 2010; 6:28.
- Włodarczyk M, Nowicka G. Obesity , DNA Damage , and Development of Obesity-Related Diseases. *Int J Mol Sci* 2019; 20:1146.
- Wolin K, Carson K, Colditz G. Obesity and Cancer. *The Oncologist* 2010; 15:556–565.
- WORLD HEALTH ORGANIZATION. *Fact Sheet No. 311. Obesity and Overweight*. 2020. Disponível em: <https://www.who.int/news-room/fact-sheets/detail/obesity-and-overweight>, Acesso em: 26 jul. 2019 .

Yan B, Wang H, Rabbani ZN, Zhao Y, Li W, Yuan Y, et al. Tumor Necrosis Factor- α Is a Potent Endogenous Mutagen that Promotes Cellular Transformation. *Cancer Res* 2006; 66:11565–70.

Yu G, Wang LG, Han Y, He QY. ClusterProfiler: An R package for comparing biological themes among gene clusters. *OMICS A Journal of Integrative Biology* 2012; 16:284–287.

Zwamborn RAJ, Slieker RC, Mulder PCA, Zoetemelk I, Verschuren L, Eka H, et al. Prolonged high-fat diet induces gradual and fat depot-specific DNA methylation changes in adult mice. *Sci Rep* 2017; 7:43261.

APÊNDICES

Apêndice A – Análise descritiva dos pacientes da coorte do TCGA

Variables	BLCA (n=361)	CESC (n=247)	CHOL (n=36)	COAD (n=222)	DLBC (n=37)	ESCA (n=173)	KIRP (n=232)	LHIC (n=360)	READ (n=76)	SKCM (n=247)	THYM (n=100)	UCEC (n=600)	UCS (n=62)	UVM (n=53)
Gender														
Female	92 (25.5%)	247 (100%)	19 (54.3%)	104 (46.8%)	21 (56.8%)	24 (13.9%)	61 (26.3%)	116 (32.8%)	33 (43.4%)	102 (41.3%)	47 (47.0%)	500 (100%)	52 (100%)	25 (47.2%)
Male	269 (74.5%)	-	16 (45.7%)	118 (53.1%)	16 (43.2%)	149 (86.1%)	171 (73.7%)	242 (67.2%)	43 (56.6%)	145 (58.7%)	53 (53.0%)	-	-	28 (52.8%)
BMI (Kg/m²)														
Normal	140 (38.8%)	85 (34.4%)	9 (25.7%)	72 (32.4%)	18 (48.6%)	87 (50.3%)	47 (20.2%)	165 (45.8%)	18 (23.7%)	79 (32.0%)	35 (35.0%)	87 (17.4%)	17 (32.7%)	16 (30.2%)
Underweight	9 (2.5%)	11 (4.4%)	1 (2.8%)	1 (0.4%)	2 (5.4%)	8 (4.6%)	6 (2.6%)	25 (6.9%)	4 (5.3%)	5 (2.0%)	3 (3.0%)	4 (0.8%)	-	1 (1.9%)
Overweight	125 (34.6%)	69 (27.9%)	16 (45.7%)	79 (35.6%)	12 (32.4%)	49 (28.3%)	94 (40.5%)	99 (27.5%)	37 (48.7%)	87 (35.2%)	25 (25.0%)	110 (22.0%)	15 (28.8%)	22 (41.5%)
Obese	87 (24.0%)	82 (33.2%)	9 (25.7%)	70 (31.5%)	5 (13.5%)	29 (16.8%)	85 (36.6%)	71 (19.7%)	17 (22.3%)	76 (30.8%)	37 (37.0%)	299 (59.8%)	20 (38.5%)	14 (26.4%)
Tumor Stage														
I	2 (0.5%)	130 (53.4%)	19 (54.3%)	31 (14.5%)	6 (18.7%)	17 (11.3%)	131 (62.0%)	173 (50.9%)	11 (15.7%)	25 (10.8%)	-	312 (62.4%)	22 (42.3%)	-
II	117 (32.6%)	59 (24.3%)	8 (22.9%)	84 (39.7%)	15 (46.9%)	72 (48.0%)	15 (7.1%)	79 (23.1%)	19 (28.5%)	88 (38.0%)	-	49 (9.8%)	5 (9.6%)	17 (32.1%)
III	119 (33.1%)	35 (14.4%)	1 (2.8%)	74 (34.6%)	3 (9.3%)	52 (34.7%)	49 (23.2%)	86 (25.1%)	30 (42.8%)	93 (40.2%)	-	112 (22.4%)	16 (30.7%)	33 (62.2%)
IV	121 (33.7%)	18 (7.4%)	7 (20.0%)	24 (11.2%)	8 (25.0%)	9 (6.0%)	16 (7.6%)	4 (1.6%)	10 (14.3%)	11 (4.8%)	-	27 (5.4%)	9 (17.3%)	3 (5.7%)
Race														
American indian or Alaska native	-	7 (3.1%)	-	1 (0.45%)	-	-	1 (0.4%)	-	-	-	-	4 (0.8%)	-	-
Asian	43 (12.5%)	18 (8.0%)	3 (8.6%)	8 (3.6%)	17 (46.0%)	45 (29.2%)	4 (1.8%)	157 (45.0%)	-	11 (4.5%)	12 (12.2%)	20 (4.2%)	3 (5.8%)	-
Black or African american	18 (5.2%)	22 (9.8%)	2 (5.7%)	48 (21.6%)	-	3 (2.0%)	48 (21.6%)	18 (5.1%)	6 (7.9%)	1 (0.4%)	3 (3.1%)	93 (19.6%)	8 (15.7%)	-
Native Hawaiian or other Pacific Islander	-	1 (0.4%)	-	-	-	-	-	-	-	-	-	9 (1.9%)	-	-
White	283 (82.3%)	176 (78.6%)	30 (85.7%)	165 (74.3%)	20 (54.0%)	106 (68.8%)	169 (76.1%)	174 (49.8%)	70 (92.1%)	232 (95.0%)	83 (84.7%)	348 (73.4%)	40 (78.4%)	53 (100%)
Ethnicity														
Hispanic or Latino	8 (2.4%)	23 (14.5%)	2 (5.9%)	3 (1.4%)	12 (32.4%)	3 (3.4%)	11 (5.7%)	17 (5.0%)	-	9 (3.7%)	6 (6.9%)	14 (3.8%)	1 (2.5%)	1 (2%)
Not Hispanic or Latino	325 (97.6%)	136 (85.5%)	32 (94.1%)	212 (98.6%)	25 (67.6%)	85 (96.6%)	182 (94.3%)	324 (95.0%)	72 (100%)	234 (93.3%)	81 (93.1%)	352 (96.1%)	38 (97.4%)	50 (98%)