

**ANÁLISE DE ASSINATURAS MUTACIONAIS  
NO PROGNÓSTICO DE PACIENTES COM  
ADENOCARCINOMA GÁSTRICO E INFECÇÃO  
POR EBV**

**IRINA GUEROLDOVNA BOBROVNITCHAIA**

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

**Área de concentração: Oncologia**

**Orientador: Dr. Israel Tojal da Silva**

**São Paulo  
2019**

**FICHA CATALOGRÁFICA**  
**Preparada pela Biblioteca da Fundação Antônio Prudente**

Bobrovnitchaia, Irina Gueroldovna

**Análise de assinaturas mutacionais no prognóstico de pacientes com adenocarcinoma gástrico e infecção por EBV / Irina Gueroldovna Bobrovnitchaia – São Paulo, 2019.**

57p.

Tese (Doutorado)-Fundação Antônio Prudente.

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

Orientador: Israel Tojal da Silva

Descritores: 1. Neoplasias Gástricas/Stomach Neoplasms. 2. Adenocarcinoma/ Adenocarcinoma. 3. Prognóstico/Prognosis. 4. Infecções por Vírus Epstein-Barr/ Epstein-Barr Virus Infections. 5. Taxa de Mutação/Mutation Rate. 6. Base de Dados/Database. 7. Desaminases APOBEC/APOBEC deaminases. 8. Estudos Retrospectivos/Retrospective Studies

“Abre a mente ao que eu te revelo / e retém bem o que eu te digo,  
pois não é ciência / ouvir sem reter o que se escuta.”

Dante Alighieri

“Não é sobre ganhar, é sobre não desistir. Se você tem um sonho,  
lute por ele. Existe uma disciplina. Não é sobre quantas vezes você  
foi rejeitado, caiu e teve que levantar. É quantas vezes você fica em  
pé, levanta a cabeça e segue em frente.”

Lady Gaga

## DEDICATÓRIA

Dedico essa tese, primeiramente aos pacientes que participaram desse estudo e, também a todos os pacientes oncológicos que de alguma forma, possam se beneficiar do resultado dessa pesquisa.

Dedico também aos meus pais – Irina Sedova e Guerold Bobrovnitchii (*in memoriam*), que se dedicaram de corpo e alma à minha formação, e cujo sonho será realizado agora!

## AGRADECIMENTOS

Há muitas pessoas que, direta ou indiretamente, me ajudaram nessa jornada!

A principal pessoa que devo agradecer, é o meu orientador, Dr. Israel Tojal da Silva, que me recebeu em seu laboratório sabendo do desafio de orientar uma aluna sem base em Bioinformática. Agradeço a sua paciência comigo e com as minhas limitações, e pela oportunidade de fazer parte do seu grupo e participar da sua linha de pesquisa.

Agradeço aos colegas do Laboratório de Bioinformática e Biologia Computacional.

Agradeço também o Dr. Emmanuel Dias Neto, que permitiu que eu fizesse parte do projeto temático, do qual esse trabalho faz parte, e que abriu as portas do seu laboratório para que eu possa desenvolver a parte prática do projeto.

Também agradeço aos integrantes do Laboratório de Genômica Médica, liderado pelo Dr. Emmanuel. Em especial à Maria Amorim e Thais Bartelli, que me ajudaram com os experimentos.

Agatha (Gatita) e Mayara (Amora), minhas filhotas do coração, agradeço pela amizade/fraternidade genuína, cafés, broncas, ajudas, passeios e tudo mais! Vocês são um presente que a vida me deu!

Jennice (Je), sem a qual não saberia onde ir, o que assinar, pra onde correr! Além de ser uma amiga querida!

À todos os colegas do CIPE, pelo companheirismo e pelos cafés na copa!

À todos os funcionários da Pós-Graduação, pela ajuda e disposição em ajudar sempre!

Ao A.C.Camargo Cancer Center, por disponibilizar esse centro de pesquisa único!

Às agências de fomento, CAPES e FAPESP, sem as quais esse estudo não seria possível.

Um agradecimento mais que especial à D. Myrian, a qual abriu a sua casa e seu coração para mim, e me trata como filha, desde que eu vim para São Paulo! É a família que não tenho presente aqui. E agradeço aos ronrons gostosos da Micha, que sempre me acalma quando mais preciso!

E, além de dedicar, agradeço aos meus pais, pela paciência, dedicação, educação, empenho e investimento que fizeram em mim! Ter pais como vocês é uma bênção! Meu pai estaria mega feliz e orgulhoso em ver mais essa etapa da minha vida concluída, um sonho realizado! Essa Tese é para vocês e de vocês!

## RESUMO

Bobrovnitchaia IG. **Análise de assinaturas mutacionais no prognóstico de pacientes com adenocarcinoma gástrico e infecção por EBV.** São Paulo; 2019. [Tese de Doutorado-Fundação Antônio Prudente].

Adenocarcinoma gástrico (AdG) possui uma alta incidência na população mundial e brasileira. É um tipo tumoral cujos sintomas são bastante inespecíficos no seu início, resultando no diagnóstico já em estádios avançados. Recentemente, foi proposta uma nova classificação, estratificando o AdG em quatro subtipos, sendo um deles caracterizado pela infecção pelo Epstein Barr vírus (EBV). Neste subtipo, por sua vez, não são totalmente conhecidos os mecanismos subjacentes à infecção, os quais podem estar relacionados à instabilidade genômica e mecanismos adicionais de tumorigênese. Deste modo, o presente estudo fez uso de abordagens de Bioinformática para análise de dados públicos (TCGA) e sua validação no painel gênico de pacientes com AdG do A.C.Camargo Cancer Center, buscando assinaturas mutacionais relacionadas à instabilidade genômica na presença de EBV. A partir dos dados obtidos do consórcio TCGA, observamos expressão aumentada de APOBEC3s no subtipo molecular de AdG EBV positivo. Também foi possível ver maior taxa mutacional, no contexto TCW, nesse subtipo molecular, condizente com a ação de APOBECs. Na coorte do A.C.Camargo Cancer Center, foram obtidas amostras de tecido e de suco gástrico de 240 pacientes com AdG (caso) e 137 controles saudáveis. Nos pacientes caso, foi possível observar 6% de amplificação de fragmento de EBV na biópsia e 11,1% no Suco Gástrico. Os achados do TCGA foram validados na coorte do hospital, utilizando painel gênico. Foi possível confirmar que a ação da família APOBEC3 está mais presente em amostras EBV positivas, indicando a ação viral na instabilidade gênica em AdG, através da família APOBEC.

**Descritores:** Neoplasias Gástricas. Adenocarcinoma. Prognóstico. Infecções por Vírus Epstein-Barr. Taxa de Mutação. Base de Dados. Desaminases APOBEC. Estudos Retrospectivos.

## SUMMARY

Bobrovnitchaia IG. [**Mutational signature analysis in gastric cancer patients' prognosis and EBV infection**]. São Paulo; 2019. [Tese de Doutorado-Fundação Antônio Prudente].

Gastric adenocarcinoma (GA) is highly frequent in the world and in Brazilian population. It's symptoms are rather nonspecific in the beginning, being diagnosed in advanced stages. Recently, a new classification was proposed, stratifying GA in four subtypes, one characterized by Epstein Barr Virus (EBV) infection. Yet, in this subtype, the underlying mechanisms of infection are not fully understood, which may be related to genomic instability and additional mechanisms of tumorigenesis. Thus, the present study made use of Bioinformatics approaches for public data analysis (TCGA) and its validation in the gene panel of A.C.Camargo Cancer Center patients with GA, searching for mutational signatures related to genomic instability in the presence of EBV. Analysing data from TCGA consortium, we observed increased expression of APOBEC3s in the EBV positive GA molecular subtype. It was also possible to observe a higher mutational rate, in TCW context, in GA with EBV infection, consistent with APOBECs action. In A.C.Camargo Cancer Center cohort, tissue samples and gastric juice were collected from 240 GA patients (case) and 137 healthy controls. It was possible to observe 6% EBV fragment amplification in biopsy and 11.1% in Gastric Juice. TCGA findings were validated in the hospital cohort, by gene panel. It was possible to confirm that APOBEC3 family members are more abundant in EBV positive samples, indicating a viral action in GA genomic instability through APOBEC3 members.

**Key-words:** Stomach Neoplasms. Adenocarcinoma. Prognosis. Epstein-Barr Virus Infections. Mutation Rate. Database. APOBEC deaminases. Retrospective Studies



## LISTA DE FIGURAS E QUADROS

<b>Figura 1</b>	Incidência de câncer na população brasileira, para o biênio 2018-2019 .....	3
<b>Figura 2</b>	Principais características e alterações genômicas que têm sido associadas com cada subtipo molecular de CG propostas pelo TCGA	6
<b>Figura 3</b>	Representação da infecção de células B e células epiteliais por EBV	8
<b>Figura 4</b>	Ilustração da ação desaminase das enzimas APOBEC .....	11
<b>Figura 5</b>	Comparação dos níveis de expressão dos membros de APOBEC3 entre os subtipos tumorais de câncer gástrico .....	29
<b>Figura 6</b>	Frequência relativa de mutações nos quatro subtipos de CG correlacionada com expressão de APOBEC3 .....	32
<b>Figura 7</b>	Frequência relativa do motivo TCW correlacionada com expressão de APOBEC3 nos quatro subtipos de CG .....	33
<b>Figura 8</b>	Análise do enriquecimento TCW entre as amostras EBV+ e EBV-....	34
<b>Figura 9</b>	Comparação dos níveis de expressão dos genes APOBEC3 e mutações relacionadas a APOBEC3 .....	36

<b>Figura 10</b>	<i>Heatmap</i> para todas as correlações entre os níveis de expressão dos genes da família APOBEC3 e o enriquecimento TCW .....	37
<b>Figura 11</b>	Correlação entre a pureza do tumor e a expressão dos genes APOBEC3.....	39
<b>Figura 12</b>	Taxa mutacional no contexto TCW é proporcional aos níveis de transcrição de vários genes nas amostras EBV+.....	41
<b>Figura 13</b>	Frequência relativa da taxa de mutações no contexto TCW e taxa de todos os genes mutados, em cinco classes de transcrição gênica, entre pacientes com CG EBV+ e EBV-.....	43
<b>Figura 14</b>	Imagem ilustrativa do gel de agarose, com amplificação de $\beta$ -actina e de BamH1W.....	45
<b>Figura 15</b>	Boxplot ilustrando o enriquecimento do motivo TCW em amostras de AdG EBV+ e EBV- da coorte do A.C.Camargo Cancer Center .....	49
<b>Quadro 1</b>	Dados clínicos dos pacientes do ACCC (coorte de validação) e dos pacientes do TCGA .....	21
<b>Quadro 2</b>	Dados dos oligonucleotídeos utilizados na análise de MSI.....	25

## LISTA DE SIGLAS E ABREVIATURAS

<b>ABraOM</b>	Arquivo Brasileiro Online de Mutações
<b>AD</b>	Adenocarcinoma Gástrico Difuso
<b>AdG</b>	Adenocarcinoma Gástrico
<b>AI</b>	Adenocarcinoma Gástrico Intestinal
<b>AICDA</b>	Activation Induced Cytidine Deaminase
<b>AID</b>	Activation Induced Cytidine Deaminase
<b>APC</b>	Polipose Adenomatosa Coli
<b>APOBEC</b>	Apolipoprotein B mRNA Editing Catalytic Polypeptide-like
<b>ARID1A</b>	AT-Rich Interaction Domain 1A
<b>BAM</b>	Binary Alignment Map
<b>BARF-1</b>	BamH1-A Reading Frame-1
<b>BART</b>	Binding Analysis for Regulation of Transcription
<b>BWA</b>	Burrows-Wheeler Alignment Tool
<b>CD274</b>	Cluster of Differentiation 274
<b>CDH1</b>	E-caderina 1
<b>CDKN2A</b>	Cyclin Dependent Kinase Inhibitor 2A
<b>CG</b>	Câncer Gástrico
<b>COSMIC</b>	Catalogue Of Somatic Mutations In Cancer
<b>CIN</b>	Chromosome Instability
<b>CNV</b>	Copy Number Variation
<b>DNA</b>	Ácido Desoxirribonucléico
<b>EBER1/2</b>	Epstein–Barr virus-encoded small RNA1/2
<b>EBNA2, 3A, 3B, 3C</b>	Epstein–Barr Virus Nuclear Antigen 2/3A/3B/3C
<b>EBV</b>	Epstein Barr Virus
<b>EGFR</b>	Epidermal growth factor receptor
<b>ERBB2/3</b>	Human Epidermal growth factor Receptor 2/3
<b>ExAC</b>	The Exome Aggregation Consortium
<b>FPKM</b>	Fragments Per Kilobase Million
<b>GATK</b>	Genome Analysis Tool Kit
<b>GS</b>	Genome Stable
<b>HBV</b>	Vírus da Hepatite B

<b>HCV</b>	Vírus da Hepatite C
<b>HHV-4</b>	Human Herpes Virus-4
<b>HPV</b>	Human Papilloma Virus
<b>HTLV-1</b>	Human T lymphotropic virus
<b>IG</b>	Instabilidade Genômica
<b>IgH</b>	Immunoglobulin Heavy Chain
<b>InDel</b>	Insertions or Deletions
<b>JAK2</b>	Janus Kinase 2
<b>KRAS</b>	Kirsten rat sarcoma viral oncogene homolog
<b>KSHV</b>	Kaposi's sarcoma-associated herpesvirus
<b>LMP-1</b>	Latent Membrane Protein-1
<b>EBNALP</b>	Epstein-Barr virus Nuclear Antigen Leader Protein
<b>MCV</b>	Merkel Cell Polyomavirus
<b>MMRD</b>	Mismatch Repair System Deficiency
<b>MSI</b>	Microsatellite Instability
<b>MSI-H</b>	Microsatellite Instability High
<b>MSI-L</b>	Microsatellite Instability Low
<b>PDCD1LG2</b>	Programmed Cell Death 1 Ligand 2
<b>PIK3CA</b>	Phosphatidylinositol-4,5-Bisphosphate 3-Kinase Catalytic Subunit Alpha
<b>PTEN</b>	Phosphatase and Tensin Homolog
<b>RHOA</b>	Ras Homolog Gene Family, Member A
<b>RNF43</b>	Ring Finger Protein 43
<b>SAM</b>	Sequence Alignment Map
<b>SG</b>	Suco Gástrico
<b>SMAD4</b>	SMAD family member 4
<b>SNV</b>	Single-Nucleotide Variant
<b>TCGA</b>	The Cancer Genome Atlas
<b>TP53</b>	Tumor Protein 53
<b>WES</b>	Whole Exome Sequencing
<b>WHO</b>	World Health Organization

# ÍNDICE

<b>1</b>	<b>INTRODUÇÃO</b> .....	<b>1</b>
1.1	Câncer Gástrico.....	1
1.2	Epstein-Barr Vírus (EBV).....	6
1.3	Instabilidade Genômica.....	8
1.4	APOBECs .....	10
<b>2</b>	<b>JUSTIFICATIVA</b> .....	<b>14</b>
<b>3</b>	<b>OBJETIVO</b> .....	<b>16</b>
3.1	Objetivo Geral.....	16
3.2	Objetivos Específicos.....	16
<b>4</b>	<b>MATERIAIS E MÉTODOS</b> .....	<b>17</b>
4.1	TCGA.....	17
4.1.1	Dados clínicos e genômicos.....	17
4.1.2	Deteção do enriquecimento das mutações APOBEC.....	17
4.1.3	Enriquecimento de mutações nos genes expressos .....	18
4.1.4	Associação com pureza do tumor .....	19
4.1.5	Análise estatística.....	19
4.2	Coorte de Validação .....	20
4.2.1	Extração do material genômico.....	22
4.2.2	Painel gênico .....	22
4.2.3	Chamada de Variante .....	22
4.2.4	Genotipagem de EBV .....	23
4.2.5	Análise de MSI .....	24
4.3	Amostras do Projeto Temático.....	25
4.3.1	Extração do material genômico.....	25
<b>5</b>	<b>RESULTADOS E DISCUSSÃO</b> .....	<b>26</b>

5.1	Análise de dados públicos (TCGA) .....	27
5.2	Análise de dados da coorte do A.C.Camargo Cancer Center .....	44
5.2.1	Genotipagem de EBV .....	44
5.2.2	Hibridização <i>in situ</i> .....	47
5.2.3	Análise de Instabilidade de Microssatélites .....	48
<b>6</b>	<b>CONCLUSÕES .....</b>	<b>50</b>
<b>7</b>	<b>REFERÊNCIAS BIBLIOGRÁFICAS .....</b>	<b>51</b>

## **ANEXO**

**Anexo 1** Carta de aprovação do Comitê de Ética em Pesquisa-CEP

# 1 INTRODUÇÃO

## 1.1 CÂNCER GÁSTRICO (CG)

A epidemiologia do câncer está relacionada com diversos fatores de risco, entre eles, sexo, idade, tabagismo e etilismo, dieta, infecção por *Helicobacter pylori*, entre outros (KARIMI et al. 2014). A infecção viral também é um fator de risco para o desenvolvimento tumoral, e está relacionada com um quinto dos casos de cânceres a nível mundial. Atualmente são conhecidos sete vírus considerados oncogênicos: papiloma vírus humano (HPV), vírus das hepatites B e C (HBV, HCV), vírus linfotrópico de células T humano tipo 1 (HTLV-1), polioma vírus de células de Merkel (MCV), herpes vírus associado ao Sarcoma de Kaposi (KSHV) e o Epstein-Barr vírus (EBV) (LUO e OU 2015).

Epstein-Barr vírus está etiologicamente associado com mononucleose infecciosa, linfoma de Burkitt, carcinoma de nasofaringe, linfoma de Hodgkin, síndrome de linfocitose hemofagocítica e alguns tumores gástricos (JHA et al. 2016). EBV foi o primeiro vírus a ser associado diretamente à carcinogênese em humanos, infectando cerca de 90% da população mundial, sendo a grande maioria assintomática (THOMPSON e KURZROCK 2004).

Os mecanismos conhecidos da oncogênese viral incluem a geração da instabilidade genômica, aumento da taxa de proliferação celular, resistência à apoptose, alterações nos mecanismos de reparo de DNA e mudanças na polaridade celular. O desenvolvimento tumoral através da infecção viral pode também ocorrer

devido à imunossupressão ou inflamação crônica (MORALES-SÁNCHEZ e FUENTES-PANANÁ 2014; MESRI et al. 2014).

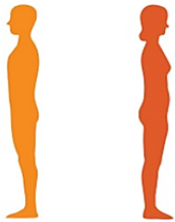
Recente relatório de incidência mundial de câncer do GLOBOCAN estimou mais de 17 milhões de novos casos e mais de 9,5 milhões de mortes por câncer (excluindo o câncer de pele não melanoma), sendo o câncer de estômago o quinto mais incidente e o terceiro em mortalidade entre cânceres. Este tumor acomete duas vezes mais homens que mulheres (BRAY et al. 2018).

A Ásia Oriental possui taxas mais elevadas de Câncer Gástrico (CG), com aproximadamente 60% dos casos, sendo o Japão o país com até 78 acometidos a cada 100 mil habitantes. Por outro lado, países da América do Norte e África possuem taxas mais baixas, cerca de 4 acometidos a cada 100 mil habitantes (Ministério da Saúde 2016).

No Brasil, estima-se para o biênio 2018-2019, a ocorrência de 420 mil casos novos de câncer, por ano, excetuando-se o câncer de pele não melanoma, sendo o câncer de estômago o quarto mais incidente entre os homens e, o sexto entre as mulheres (Figura 1). Em homens, a maior incidência ocorre nas Regiões Norte e Nordeste e a menor, na Região Sudeste. Já para mulheres, a maior incidência é nas Regiões Sul, Centro-Oeste e Norte, e a menor, no Sudeste (Ministério da Saúde 2018).



Distribuição proporcional dos dez tipos de câncer mais incidentes estimados para 2018 por sexo, exceto pele não melanoma\*

Localização Primária	Casos	%			Localização Primária	Casos	%
Próstata	68.220	31,7%		<b>Homens</b> <b>Mulheres</b>	Mama Feminina	59.700	29,5%
Traqueia, Brônquio e Pulmão	18.740	8,7%			Cólon e Reto	18.980	9,4%
Cólon e Reto	17.380	8,1%			Colo do Útero	16.370	8,1%
Estômago	13.540	6,3%			Traqueia, Brônquio e Pulmão	12.530	6,2%
Cavidade Oral	11.200	5,2%			Glândula Tireoide	8.040	4,0%
Esôfago	8.240	3,8%			Estômago	7.750	3,8%
Bexiga	6.690	3,1%			Corpo do Útero	6.600	3,3%
Laringe	6.390	3,0%			Ovário	6.150	3,0%
Leucemias	5.940	2,8%			Sistema Nervoso Central	5.510	2,7%
Sistema Nervoso Central	5.810	2,7%			Leucemias	4.860	2,4%

\*Números arredondados para múltiplos de 10.

Fonte: Adaptado de Ministério da Saúde (2018)

**Figura 1** - Incidência de câncer na população brasileira, para o biênio 2018-2019

O **CG** é uma doença heterogênea e multifatorial, caracterizada pela diversidade patológica e epidemiológica. Entre os vários tipos de **CG** podemos citar os carcinomas escamosos, coriocarcinomas, carcinomas com estroma linfóide, tumores estromais gastrointestinais (GIST), entre outros (HU et al. 2012; PARK et al. 1999). No entanto, a maioria dos **CG** (mais de 90%) é de adenocarcinomas gástricos (**AdG**), podendo estes ser classificados de acordo com o sítio anatômico como tumores proximais (originários na cárdia) ou tumores distais (com origem distal à cárdia). Pacientes com **AdG** proximal possuem pior prognóstico, independentemente do estadió TNM (PINTO DE SOUSA et al. 2001).

Outros sistemas de classificação utilizados na prática clínica são a de Lauren e a de *World Health Organization* (WHO). A classificação de Lauren caracteriza o **AdG** em duas variantes patológicas: intestinal (**AI**) e difusa (**AD**). O **AI** representa 50% dos casos e é um tipo tumoral com arquitetura tubular bem desenvolvida, associado com gastrite atrófica, metaplasia intestinal e infecção por *Helicobacter pylori* (*H. pylori*). Já o **AD** possui uma frequência de aproximadamente 35%, é caracterizado pela infiltração de células individuais e possui maior contribuição dos

fatores hereditários, enquanto 15% dos AdG são do tipo indeterminado (HU et al. 2012). O sistema de classificação de WHO compreende quatro subtipos: papilar, tubular, mucinoso (colóide) e pouco coesivo (CHIA e TAN 2016; GRABSCH e TAN 2013).

Os tipos **AI** e **AD** podem ser classificados de acordo com o prognóstico em precoces e avançados. Os **AdGs** precoces possuem uma taxa de sobrevivência em 5 anos de 85 a 100%, enquanto para os avançados, esta taxa é de 5 a 20%. O tipo precoce é definido como limitado à mucosa ou submucosa, sem invasão linfonodal. Já o **AdG** avançado é classificado de acordo com a aparência geral: polipóide, ulcerado com bordas bem definidas ou bordas irregulares e infiltrativas e, finalmente, infiltrativo difuso sem evidência de massa ou úlcera (PIAZUELO e CORREA 2013).

Apesar de vários tipos de classificação de **CG**, ainda há limitações no entendimento de sua heterogeneidade, necessitando de classificadores mais robustos que permitam fornecer informações precisas para, assim, conduzir uma terapia alvo que considere as características moleculares de cada subtipo (LIN et al. 2015). Neste cenário, o consórcio *The Cancer Genome Atlas* (TCGA) gerou dados genéticos de **CG** com o objetivo caracterizar as bases moleculares desse tumor utilizando, para tanto, amostras dos tecidos normal e tumoral de 295 pacientes (BASS et al. 2014). Neste estudo, foram empregadas diversas ferramentas de análise de dados moleculares, resultando em informações em diferentes níveis, incluindo: o número de cópias de DNA (CNV, *Copy Number Variation*); a instabilidade de microssatélites (MSI, *Microsatellite instability*); o perfil de metilação de DNA; o sequenciamento total de exoma (WES, *Whole Exome Sequencing*); o array do perfil de metilação de DNA; o sequenciamento de RNA mensageiro; o sequenciamento de

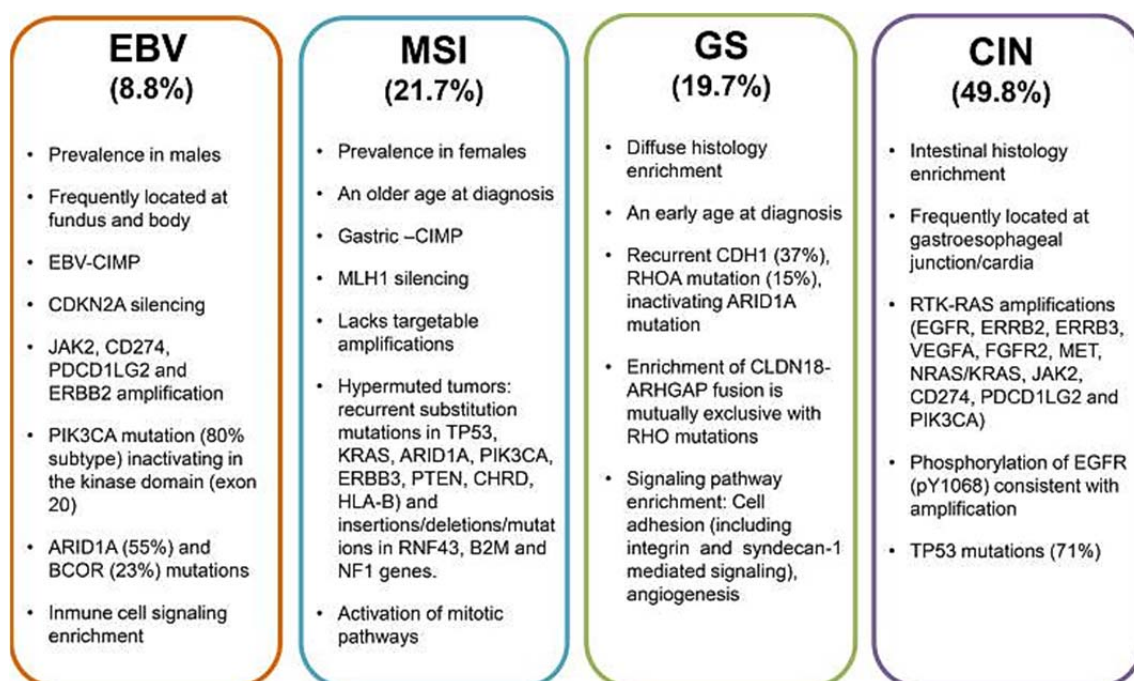
microRNA e o array protéico de fase reversa (RPPA). Como resultado, foi proposta uma nova classificação do CG, estratificando esse tumor nos seguintes marcadores moleculares: tumores infectados por Epstein-Barr Virus (**EBV+**), tumores com instabilidade de microssatélites (MSI), tumores geneticamente estáveis (GS) e tumores cromossomicamente instáveis (CIN).

O subtipo **EBV+** compreende aproximadamente 9% de **CG**, dos quais a maior parte da população acometida é masculina (GULLEY 2015). É um subtipo de localização de fundo ou corpo do estômago. Aproximadamente 80% dos **CGs EBV+** carregam mutações no gene *PIK3CA*, apresentando alta prevalência de hipermetilação do DNA. Também foi detectada hipermetilação do promotor de *CDKN2A* em todas as amostras de **CG EBV+**. Além disso, foram detectadas ampliações dos genes *JAK2*, *CD274*, *PDCD1LG2* e *ERBB2* e deleções de *PTEN*, *SMAD4*, *CDKN2A* e *ARID1A* (BASS et al. 2014).

O subtipo MSI representa 21% dos **CG** e geralmente é caracterizado pelo acúmulo de mutações nos genes *PIK3CA*, *ERBB3*, *ERBB2* e *EGFR*, assim como nos outros genes considerados com mutações recorrentes (*hotspots*) em outros tumores. De acordo com o grau de instabilidade, os MSIs podem ser de nível alto (MSI-H) ou baixo (MSI-L). O MSI-H tem sido associado com **AI** e a um melhor prognóstico (BASS et al. 2014).

O tumores **GS** representam 20% dos **CG** e são tipicamente diagnosticados em pacientes jovens. Este subtipo possui frequência de mutações de *TP53* relativamente baixa e baixo grau de aneuploidia; neste subtipo foram detectados 13-37% de mutações em *CDHI* e 14-25% em *RHOA* (BASS et al. 2014).

Finalmente, o quarto subtipo de CG descrito foi o CIN, fortemente associado com junção gastresofágica. Esse subtipo é caracterizado por mutações recorrentes nos genes *TP53*, *ARID1A*, *KRAS*, *PIK3CA*, *RNF43*, *ERBB2* e *APC*, além de fosforilação significativamente elevada de EGFR. Este subtipo de também possui frequente amplificação de genes que codificam para receptores de tirosinas quinase (Figura 2) (BASS et al. 2014; RIQUELME et al. 2015).



Fonte: Adaptado de RIQUELME et al. (2015).

**Figura 2** - Principais características e alterações genômicas que têm sido associadas com cada subtipo molecular de CG propostas pelo consórcio TCGA

## 1.2 EPSTEIN-BARR VIRUS (EBV)

Epstein-Barr Vírus (EBV), ou herpes vírus humano 4 (HHV-4), é classificado como  $\gamma$ -herpes vírus e contém uma molécula de DNA linear de aproximadamente

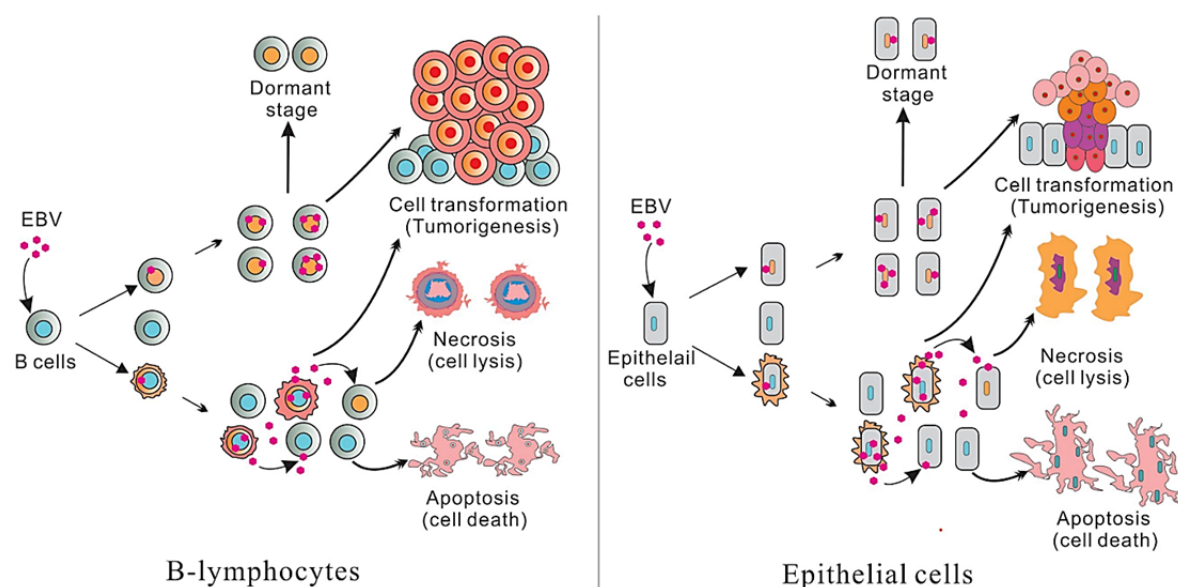
172Kb que codifica >85 genes (THOMPSON e KURZROCK 2004). O alvo principal do EBV são as células B, mas em algumas circunstâncias EBV também pode infectar células T e células epiteliais (Figura 3) (KÜPPERS 2003).

A principal forma de disseminação de EBV é através de secreções orais. A replicação viral acontece na orofaringe, estando o vírus presente na saliva de quase todos os afetados (ELIGIO et al. 2010). Após a infecção primária, EBV estabelece uma infecção latente, que expressa um conjunto limitado de genes virais necessários para a manutenção do episoma viral, conferindo assim vantagem de sobrevivência para a célula infectada. Linfoma de Burkitt, cerca de 50% dos carcinomas de nasofaringe e o CG associado ao EBV pertencem à latência I que, por sua vez, é caracterizada pela expressão do ácido ribonucleico codificado por EBV 1 e 2 (*EBER1/2*) e transcritos de RNA da região BamHI-A (*BART*). Já as neoplasias da latência II (restante dos carcinomas de nasofaringe e linfoma de Hodgkin) são caracterizadas pela expressão da proteína latente de membrana (LMP1), além *EBER1/2* e do *BART*. Neoplasias da latência III, típicas dos linfomas observados em pacientes imunocomprometidos, adicionalmente expressam *EBNA2*, 3A, 3B, 3C e LP (IIZASA et al. 2012).

O papel patogênico de EBV em CG ainda é pouco conhecido. Estudos mostram que no CG do tipo linfoepitelioma, EBV espalha da nasofaringe para o estômago. Já no AdG, EBV pode entrar nas células epiteliais gástricas através da ligação de IgA com partículas do EBV derivadas dos linfócitos B e captação destas partículas pelas células epiteliais gástricas. Além disso, EBV tem mostrado um padrão novo de latência, caracterizado pela produção de BARF-1 e ausência de LMP-1. Desta forma, o EBV pode estar envolvido nos estágios iniciais de CG,

devido a proliferação monoclonal de uma única célula com infecção persistente por EBV (IIZASA et al. 2012; JHA et al. 2016).

Quanto ao efeito prognóstico de EBV em CG, os estudos são controversos. Enquanto alguns grupos conseguiram mostrar um melhor prognóstico em CG EBV+ (van BEEK et al. 2004; SONG e KIM 2011; CAMARGO et al. 2014), outros falharam nessa observação (CHANG et al. 2001; PARK et al. 2009).



Fonte: Adaptado de JHA et al. (2016)

**Figura 3** - Representação da infecção de células B e células epiteliais por EBV

### 1.3 INSTABILIDADE GENÔMICA

Mudanças no DNA genômico podem ser alterações na sequência primária ou rearranjos que alteram a estrutura cromossômica. A maioria das mutações e rearranjos surgem ao acaso durante o processo de duplicação genômica ou em resposta à dano no DNA, e múltiplas vias de reparo estão envolvidas na manutenção

de tais alterações. As consequências de alta instabilidade genética são evidentes, onde a perda da capacidade de reparo é associada com síndrome de predisposição ao câncer e com envelhecimento (KIM e JINKS-ROBERTSON 2012).

Instabilidade genômica (**IG**) pode ser definida como tendência aumentada do genoma em adquirir mutações. Ela ocorre quando vários processos, envolvidos na manutenção e replicação do genoma, perdem função ou quando há exposição aumentada aos carcinógenos. Os mecanismos que levam à **IG** incluem defeitos herdados ou adquiridos durante o reparo de DNA, replicação de DNA, controle do ciclo celular ou da segregação cromossômica resultando nas mudanças na sequência de nucleotídeos de uma única base (SNV, *Single Nucleotide Variation*), quebra de DNA, translocações cromossômicas, inversões e deleções (InDel) e aneuploidia (CHARAMES e BAPAT 2003; LANGIE et al. 2015).

Apesar do padrão mutacional de cada tumor fornecer informação útil sobre os mecanismos da carcinogênese, vários tipos de assinaturas mutacionais frequentemente coexistem num tipo de tumor específico, variando de acordo com os subtipos histológicos (ALEXANDROV et al. 2013). Por exemplo, no **CG**, os tumores com estabilidade de microssatélites exibem variações de T:A>G:C no contexto CpTpT; tumores com MSI possuem padrões de estabilidade cromossômica e um grande número de substituições T:A>G:C; já os tumores do tipo difuso (**AD**) possuem poucas variantes de nucleotídeos únicos e alterações do número de cópias (CNV) (MATSUMOTO et al. 2015).

O cálculo da **IG** pode ser realizado em função da carga mutacional, distinguindo os tumores estáveis de moderadamente instáveis e altamente instáveis.

Essa estimativa, por sua vez, permite avaliar o risco desse tumor adquirir resistência à intervenção citotóxica (EPSTEIN 2015).

A infecção por EBV pode levar à instabilidade genética nas células do hospedeiro através da introdução de dano de DNA e desregulação dos mecanismos de reparo de DNA. Por exemplo, EBNA-1 pode induzir aberrações cromossômicas e DSB, já a LMP1 pode induzir a formação de micronúcleos e reprimir o reparo de DNA em células epiteliais. Além disso, LMP1 pode estimular a via de PI3K/AKT, resultando na fosforilação de FOXO3, levando sua retenção no núcleo, impedindo a ativação dos componentes das vias de reparo. A proteína serina/treonina quinase BGLF4, codificada por EBV, também é associada à instabilidade genética, sua expressão leva a ativação de DDR (*DNA Damage Response*) e atrasa a progressão através da fase S. Já a expressão de BGLF5, em células epiteliais, pode induzir DSB, levando ao aumento de MSI e mutações genéticas (HOLLINGWORTH e GRAND 2015).

Estas informações fortalecem a necessidade de que os tumores sejam classificados em subtipos genéticos com base nas suas assinaturas mutacionais e outras características de aberrações genéticas além da classificação tradicional (MATSUMOTO et al. 2015), além de suportar a hipótese de envolvimento da infecção viral como uma das etapas da carcinogênese.

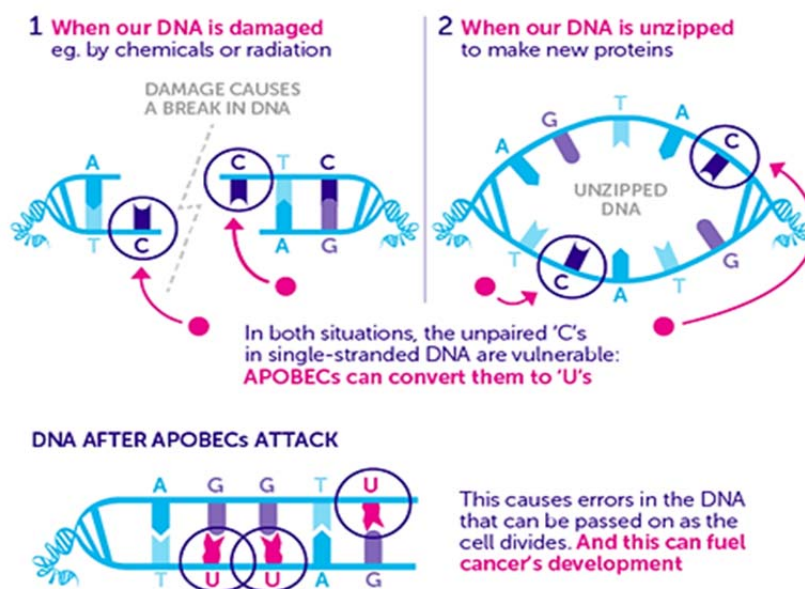
#### **1.4 APOBECs**

Estudos recentes mostram que uma classe de genes de edição de ácidos nucléicos podem causar mutações no DNA. Os genes da família *Apolipoprotein B*



*mRNA Editing Enzyme, Catalytic Polypeptide* (APOBEC) possuem uma atividade catalítica onde induz a desaminação no DNA, resultando em alterações T:A>C:G no contexto TpCp[AT]. O agrupamento regional dessas mutações, conhecido como kataegis, podem ser indicativos de rearranjos cromossômicos sugerindo que esse tipo de mutação está ligada com o desenvolvimento tumoral (ROBERTS et al. 2013; MATSUMOTO et al. 2015).

A atividade dos genes APOBECs faz parte do repertório do sistema imune, importante tanto na resposta imune inata quanto adaptativa, sendo ativada para restringir a replicação viral e de elementos transponíveis. Estas enzimas possuem a capacidade única de editar RNA e/ou DNA tanto do hospedeiro quanto do patógeno, atividade relacionada à sua habilidade de desaminação da citidina em uridina (Figura 4) (MORIS et al. 2014).



**Fonte:** Adaptado do blog do Cancer Research UK Science

**Figura 4** - Figura representativa da ação do APOBEC.

A família APOBEC é formada por 11 membros: APOBEC1 (A1) no cromossomo 6, APOBEC2 (A2) no cromossomo 12, APOBEC3A (A3A), B (A3B), C (A3C), D (A3D), F (A3F), G (A3G), H (A3H) codificados em tandem no cromossomo 22, APOBEC4 (A4) no cromossomo 1 e *Activation Induced Cytosine Deaminase* (AID/AICDA) no cromossomo 6. Todas as enzimas APOBECs, exceto APOBEC2 e APOBEC4, são capazes de converter citosina para uracila onde ssDNA é exposto e acessível para desaminação (BURNS et al. 2015).

Todas as enzimas da família APOBEC/AID possuem pelo menos um domínio catalítico dependente de zinco, o qual contém uma sequência consenso de aminoácidos H-X-E-X<sub>23</sub>-28-P-CX<sub>2</sub>-4-C (onde X pode ser qualquer aminoácido). A3A, A3C, A3H, AID e A1 possuem domínio único zinco-dependente conservado, enquanto A3B, A3D, A3F e A3G possuem dois domínios conservados zinco-coordenado. A família APOBEC inclui dois membros de oito éxons (A3B e A3G), um membro de 7 éxons (A3F), um de cinco éxons (A3A), dois membros de quatro éxons (A3C e A3E) e um membro de três éxons (A3D) (ZOU et al. 2017).

Os diferentes processos mutacionais podem gerar combinações de mutações, chamados “assinaturas”. Nessa perspectiva, ALEXANDROV et al. (2013) utilizou cerca de cinco milhões de substituições somáticas e pequenos InDels de 30 tumores diferentes na detecção de padrões de mutações. Ao final, foi possível identificar 21 assinaturas mutacionais, sendo que uma delas estava associada à super-atividade dos membros da família APOBEC. Uma vez que a ativação de APOBEC faz parte da resposta imune inata a vírus e retrotransposons, é possível que a atividade anormal dessas enzimas leve a danos colaterais (*off-targets*) no genoma humano durante a infecção viral. A confirmação desta hipótese poderia estabelecer um possível

mecanismo importante para o início e progressão da carcinogênese. Desta forma, a atividade elevada dos níveis de expressão dos genes APOBECs pode resultar em uma carga mutacional no contexto TCW, característica dessas enzimas.

## 2 JUSTIFICATIVA

O adenocarcinoma gástrico (**AdG**) é um dos tumores mais frequentes no mundo. Apesar de ter apresentado diminuição na sua incidência, este tipo de tumor ainda se mostra como um grave problema de saúde pública. O **AdG** muitas vezes apresenta sintomas inespecíficos e, comumente, é diagnosticado em estádios avançados. Desta forma, há a necessidade de um conhecimento mais aprofundado dos fatores genéticos, bem como os fatores de risco incluindo as infecções por *H. pylori* e EBV.

Como descrito no trabalho de ROBBIANI et al. (2015), em colaboração com nosso grupo, foi investigada a relação entre infecção crônica por *Plasmodium falciparum* e linfoma de Burkitt. Neste trabalho, foi observado que a infecção crônica promove a expansão prolongada dos centros germinativos, onde as células B sofrem rápida expansão clonal e expressam AID. Como resultado, a super expressão da AID induziu mutações *off-targets* (fora da região IgH) precipitando, assim, translocações cromossômicas em oncogenes, incluindo o gene c-Myc. Desta forma, pôde-se concluir que a malária crônica favoreceu a linfomagenese em tumores de células B maduras pela expressão prolongada do gene AID.

Esses achados nos levam a pensar que, em tumores do subtipo EBV, a atividade das enzimas APOBEC tem um efeito secundário na geração de dano no DNA desse subtipo e, conseqüentemente, na carcinogênese. Desta forma, o nosso trabalho tem por objetivo contribuir com melhor conhecimento da ação dos APOBECs em tumores gástricos infectados por EBV. Essa informação pode ser

relevante para uma melhor estratificação dos pacientes e tratamento mais direcionado.

### 3 OBJETIVOS

#### 3.1 OBJETIVO GERAL

Identificar fontes adicionais de dano ao DNA resultante da ação das enzimas APOBECs em AdG.

#### 3.2 OBJETIVOS ESPECÍFICOS

- Realizar a genotipagem dos **AdG** para a presença do EBV;
- Avaliar a presença de instabilidade de microssatélites (MSI) nas amostras tumorais;
- Identificar as assinaturas mutacionais induzidas por APOBECs nas amostras de **AdG EBV+** e **EBV-**;
- Correlacionar a expressão dos genes da família APOBEC com mutações no contexto TCW;
- Validar os achados em uma coorte brasileira.

## 4 MATERIAIS E MÉTODOS

### 4.1 TCGA

#### 4.1.1 Dados clínicos e genômicos

Foram recuperados do banco de dados disponibilizados pelo consórcio *The Cancer Genome Atlas* (TCGA - <https://portal.gdc.cancer.gov>) mutações de variantes somáticas de um único nucleotídeo (SNV), níveis de expressão de genes normalizados (*Fragments Per Kilobase of exon per Million fragments mapped* - FPKM) e os dados clínicos de 240 casos de câncer gástrico. Estes *datasets* provém de quatro subtipos tumorais distintos: tumor positivo para Epstein–Barr virus (EBV+, N=23), tumor com instabilidade de microssatélite (MSI, N=47), tumor geneticamente estável (GS, N=48) e tumor com instabilidade cromossômica (CIN, N=122).

#### 4.1.2 Detecção do enriquecimento das mutações APOBEC

O motivo TCW (W = A ou T) foi escolhido por ser o mais estrigente e preferido no contexto de desaminação da citosina mediada por APOBEC3. A taxa de enriquecimento no contexto TCW foi considerada como a medida da prevalência da mutação APOBEC (como descrito em ROBERTS et al 2013) e calculada como:

$$(M_{tcw} * C_{cg}) / (M_{cg} * C_{tcw})$$

onde  $M_{tcw}$  é o número de citosinas/guaninas mutadas localizadas no motivo TCW,  $M_{cg}$  é o número total de citosinas/guaninas mutadas,  $C_{tcw}$  é o número total de TCW dentro de +/- 20 nucleotídeos em volta das citosinas/guaninas mutadas e  $C_{cg}$  é o número total de citosinas/guaninas dentro de uma janela de +/- 20 nucleotídeos ao redor de citosinas/guaninas mutadas.

#### 4.1.3 Enriquecimento de mutações nos genes expressos

Os níveis de expressão de mRNA foram avaliados para cada um dos sete genes da família AC3: A3A, A3B, A3C, A3D, A3F, A3G, e A3H. Para tanto, foram considerados os níveis de expressão gênica disponibilizados pelos pipelines do TCGA (<https://portal.gdc.cancer.gov/>).

Os níveis de expressão quantitativa encontram-se normalizados pela contagem de *reads* mapeados e dos *reads* sequenciados através do método FPKM (*Fragments Per Kilobase of exon per Million fragments mapped*).

Essa abordagem considera unicamente os *reads* mapeados do experimento de RNA-Seq (profundidade de sequenciamento), o tamanho dos transcritos, assim como o número de contagens para todos os éxons de cada transcrito. Em seguida, após a separação dos transcritos sem expressão (silenciados), nós estratificamos os transcritos ativos em quatro níveis (alto, os top 25%; médio, 75 a 50%; baixo, 50 a 25% e trace, aqueles abaixo dos 25%) com base nos valores de FPKM, usando o modelo de mistura normal implementado pelos métodos *Mclust quantile* disponíveis na linguagem R, como previamente descrito (KLEIN et al 2011). Então, calculamos o valor numérico para avaliar a densidade de mutações induzidas por APOBEC nessas respectivas classes, para cada amostra de câncer gástrico, como segue:



$$(10^6/\text{TG}) * (\text{TCW}/\text{TM})$$

onde **TG** é o total de genes com qualquer mutação, **TCW** é o total de mutações no contexto APOBEC para cada classe e, finalmente, **TM** é o total de mutações para cada classe.

#### 4.1.4 Associação com pureza do tumor

Uma vez que o microambiente tumoral tipicamente compreende também de uma mistura de populações de células, incluindo células do estroma, do infiltrado inflamatório e células tumorais, foi avaliado se a pureza do tumor (porcentagem de células tumorais) encontra-se correlacionada com expressão de APOBEC. Nesta etapa, aproveitamos a informação da pureza tumoral inferida pelo software *Estimation of STromal and Immune cells in MAlignant Tumor tissues using Expression data* (ESTIMATE) (YOSHIHARA et al. 2013) disponível no portal do MD Anderson Cancer Center (<https://bioinformatics.mdanderson.org/estimate/>). O coeficiente de correlação de Pearson foi usado como teste de associação entre a pureza do tumor e a expressão gênica analisada no presente estudo.

#### 4.1.5 Análise Estatística

A comparação da expressão gênica e o enriquecimento de mutações no contexto TCW entre as amostras **EBV+** e **EBV-** foram realizadas através do teste não-paramétrico de Wilcoxon (teste não pareado unilateral) ou pelo teste de permutação de igualdade, conforme apropriado.

## 4.2 COORTE DE VALIDAÇÃO

Foram coletadas 112 biópsias de pacientes com **AdG** do A.C.Camargo Cancer Center. Destas amostras, 71 (63,4%) foram de pacientes do sexo masculino e 41 (36,6%), do sexo feminino. Quanto à localização anatômica do tumor: 11,6% estava localizado na junção esofagogástrica; 50% no corpo, 30,4% no antro, 6,2% no linite e 1,8% no coto gástrico. As amostras foram classificadas quanto ao tipo Lauren em: intestinal (41,1%), difuso (47,3%), misto (2,7%) e indeterminado (8,9%).

**Quadro 1** - Dados clínicos dos pacientes do A.C.Camargo Cancer Center (coorte de validação) e dos pacientes do TCGA. NA – sem informação

	A.C.Camargo Cancer Center (N=112)	TCGA (N=240)
<b>Gênero</b>		
Masculino	72 (64.3%)	149 (62.1%)
Feminino	40 (35.7%)	91 (37.9%)
<b>Idade</b>	37-91	39-90
Média	63 (56.3%)	65,35 (27.2%)
Mediana	64 (57.1%)	67,5 (28.1%)
Moda	49 (43.8%)	58 (24.2%)
<b>Lauren</b>		
Intestinal	46 (41.1%)	83 (34.6%)
Difuso	53 (47.3%)	43 (17.9%)
Misto	3 (2.7%)	NA
Inclassificável	10 (8.9%)	111 (46.3%)
<b>Localização</b>		
TEG	13 (11.6%)	14 (5.8%)
Corpo	56 (50.0%)	91 (37.9%)
Antro	34 (30.4%)	93 (38.8%)
Cárdia	NA	32 (13.3%)
NA	9 (8.0%)	10 (4.3%)
<b>Estadio</b>		
I	NA	1 (0.4%)
IA	17 (15.2%)	7 (2.9%)
IB	8 (7.1%)	23 (9.6%)
II	NA	21 (8.8%)
IIA	12 (10.7%)	27 (11.3%)
IIB	11 (9.8%)	37 (15.4%)
III	NA	2 (0.8%)
IIIA	12 (10.7%)	33 (13.8%)
IIIB	17 (15.2%)	29 (12.1%)
IIIC	19 (17.0%)	23 (9.6%)
IV	16 (14.3%)	20 (8.3%)
<b>Recidiva</b>		
Sim	51 (45.5%)	26 (10.8%)
Não	61 (54.5%)	173 (72.1%)
<b>Óbito</b>		
Sim	62 (55.4%)	18 (7.5%)
Não	50 (44.6%)	222 (92.5%)
<b>EBV</b>		
Sim	5 (4.5%)	23 (9.6%)
Não	107 (95.5%)	217 (90.4%)
<b>Classificação Molecular</b>		
EBV	NA	23 (9.6%)
MSI	NA	47 (19.6%)
GS	NA	48 (20.0%)
CIN	NA	122 (50.8%)

#### 4.2.1 Extração do material genômico

A extração do DNA genômico das amostras tumorais a fresco foi realizada através da metodologia padrão de extração de DNA com fenol-clorofórmio, utilizando a proteinase K (BioAmerica Inc.), pelo Biobanco do A.C. Camargo Cancer Center. Resumidamente, após a digestão com a proteinase K, o DNA foi precipitado com fenol-clorofórmio (1:1), seguido pela segunda precipitação com acetato de sódio e etanol 100%. A qualidade e quantidade do DNA obtido foi avaliada através da quantificação em Qubit® (Invitrogen), usando o kit Qubit™ dsDNA HS Assay (Invitrogen), de acordo com as recomendações do fabricante.

#### 4.2.2 Painel gênico

Foi usado DNA genômico (100ng) de 112 amostras de adenocarcinoma gástrico, para enriquecimento baseado em captura de 781,132 pb de um painel gênico customizado incluindo 99 dos genes mais frequentemente mutados no câncer gástrico. Bibliotecas foram preparadas seguindo instruções do fabricante (SeqCap EZ Library SR User's Guide Version 5.1 - NimbleGen, Roche), e foram sequenciadas na plataforma NextSeq 500 (Illumina), usando *paired-end reads* (2x75bp).

#### 4.2.3 Chamada de Variante

Para a coorte de validação, SNVs somáticas foram chamadas através do protocolo *in-house*, de acordo com as boas práticas de GATK de *Broad Institute* (VAN DER AUWERA et al. 2013). As *reads* brutas foram alinhadas ao genoma de referência hg19/GRCh37, usando BWA-mem (*Burrows Wheeler Aligner*) (LI e DURBIN 2010). Os arquivos alinhados, em formato SAM foram convertidos em

arquivos BAM, selecionados e filtrados a fim de excluir as *reads* com *score* < 15. As *reads* retidas foram processadas usando SAMtools (LI H et al. 2009) e Picard (<http://broadinstitute.github.io/picard/>), os quais excluem as *reads* de baixa qualidade e duplicatas de PCR. Finalmente, chamada de SNVs somáticas foi realizada para os dados do painel a partir dos dados BAM, usando GATK-UnifiedGenotyper (v3.8). Filtragem extensiva foi aplicada para remover qualidade de mapeamento baixa, assim como viés de fita e posição. As variantes germinativas residuais foram filtradas usando a base de dados de mutações do *Exome Aggregation Consortium* (ExAC) (LEK et al. 2016) e o Arquivo Brasileiro Online de Mutações (ABraOM) (NASLAVSKY et al. 2017).

#### 4.2.4 Genotipagem de EBV

A presença do vírus nas 112 amostras tumorais foi realizada através da técnica de PCR em Tempo Real, usando um método multiplex, onde oligonucleotídeos específicos para a região BamH1W do EBV (F: 5'-GCAGCCGCCAGTCTCT-3'; R: 5'-ACAGACAGTGCACAGGAGCCT-3') junto com a sonda marcada (FAM-AAAAGCTGGCGCCCTTGCCTG-TAMRA), amplificando aproximadamente 84 pb do genoma viral simultaneamente com o gene endógeno humano  $\beta$ -actina (F: 5'-CCATCTACGAGGGGTATGC-3'; R: 5'-GGTGAGGATCTTCATGAGGTA-3'), marcada com a sonda (VIC-CCTGCGTCTGGACCTGGCTG-NFQ). Esta reação foi realizada no equipamento ABI 7500 Fast Real Time. Foi usada condição padrão de ciclagem: 95°C por 20 segundos, 95°C por 3 segundos e 60°C por 30 segundos por 40 ciclos. Cada reação de 20  $\mu$ l continha: TaqMan® Fast Universal Master Mix (2x) (Applied

Biosystems®), oligonucleotídeos *forward* e *reverse* para cada sequência (oligonucleotídeo 18 µM; sonda 5 µM). Para verificar a reação quanto à contaminação, cada corrida continha ao menos dois controles negativos (sem *template*). Finalmente, a análise para a presença da infecção viral por EBV foi realizada calculando  $\Delta Ct$ , para o qual o *Ct* (*cycle threshold*) de EBV foi subtraído do *Ct* de  $\beta$ -actina, e o resultado calculado através da fórmula  $2^{(-\Delta Ct)}$ . As amostras foram consideradas positivas para EBV quando a razão EBV/ $\beta$ -actina era maior que 1.

#### 4.2.5 Análise de MSI

Microssatélites são sequências curtas e repetitivas de DNA randomicamente distribuídas no genoma. A deficiência do sistema de reparo de incompatibilidade (*Mismatch Repair System Deficiency* - MMRD) geralmente é causada por mutações germinativas ou silenciamento epigenético esporádico que levam à inserção ou deleção de nucleotídeos nas regiões de microssatélites durante a replicação de DNA; este fenômeno é conhecido como Instabilidade de Microssatélites (MSI). Microssatélites são sequências de DNA com comprimento entre 1 a 6 repetições de nucleotídeos (geralmente de 10 a 60 vezes). Esses motivos de DNA estão espalhados por regiões codificadoras e não-codificantes do genoma, altamente polimórficas entre as populações, mas estáveis em cada indivíduo (RATTI et al. 2018).

A análise de instabilidade de microssatélites (MSI) foi realizada, de forma pareada entre DNA extraído de biópsias e DNA extraído de leucócitos, no sequenciador ABI 3130xl, utilizando oligonucleotídeos marcados com fluoróforos específicos (Quadro 1); GeneScan 350 ROX foi usado como padrão de tamanho.

**Quadro 2** - Dados dos oligonucleotídeos utilizados na análise de MSI.

Marker	Gene	GenBank No	Localizatin	Primer Sequence (5'-3')	Fluorescent dye (antisense primer)
NR-27	<i>Inhibitor of Apoptosis Protein-1</i>	AF070674	5'UTR	AACCATGCTTGCAAACCACT CGATAATACTAGCAATGACC	HEX
NR-21	<i>SLC7A8</i>	XM_033393	5'UTR	GAGTCGCTGGCACAGTTCTA CTGGTCACTCGCGTTTACAA	FAM
NR-24	<i>Zinc finger 2</i>	X60152	3'UTR	GCTGAATTTTACCTCCTGAC ATTGTGCCATTGCATTCCAA	NED
BAT-25	<i>c-kit</i>	X69313	intron 16	TACCAGGTGGCAAAGGGCA TCTGCATTTTAACTATGGCTC	HEX
BAT-26	<i>hMSH2</i>	AY601851	intron 5	CTGCGGTAATCAAGTTTTTATAG AACCATTCAACATTTTAAACCC	FAM

### 4.3 AMOSTRAS DO PROJETO TEMÁTICO (14/26897-0)

Foram coletadas biópsias de pacientes com AdG localmente avançado coletado antes de tratamento da Quimioterapia Neoadjuvante. Além disso, amostras do suco gástrico (SG) também foram coletadas. Estas amostras também foram analisadas quanto à presença de EBV e quanto à instabilidade de microssatélites, como descrito nos tópicos 4.2.3 e 4.2.4 acima.

#### 4.3.1 Extração do material genômico

A extração do material genômico das biópsias dos pacientes com AdG do projeto temático foi realizada como descrito no tópico 4.2.1.

Extração de DNA a partir do suco gástrico foi feita de acordo com o método descrito por CHOMEZYNSKI e SACCHI (1987), baseado em fenol-clorofórmio. A

qualidade e a quantidade do DNA obtido foram medidas através do espectrofotômetro NanoDrop™ 2000 (Thermo Scientific™).



## 5 RESULTADOS e DISCUSSÃO

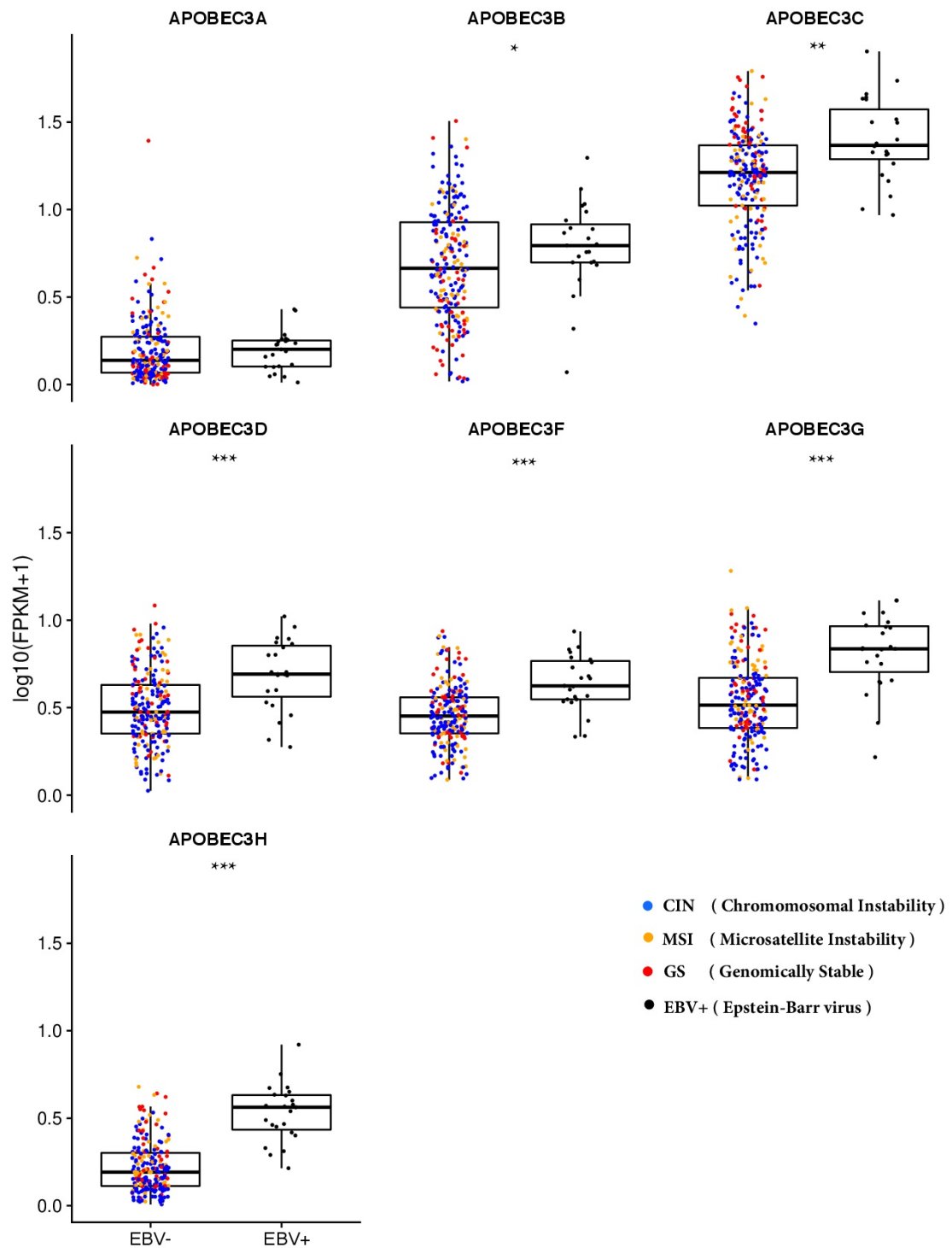
### 5.1 ANÁLISE DE DADOS PÚBLICOS (TCGA)

Em paralelo aos experimentos descritos acima, foi realizada uma análise integrativa utilizando, para tanto, SNVs somáticos, dados clínicos e dos níveis de expressão gênica normalizados de 240 casos de câncer gástrico, disponíveis no repositório *The Cancer Genome Atlas* (TCGA) a fim de avaliar possíveis danos *off-targets* no DNA gerados pelos genes APOBEC3s. Esses genes atuam na edição de ácidos nucleicos que, em condições fisiológicas, agem na restrição de vírus e elementos transponíveis. Nós, assim, hipotetizamos que, em condições anormais decorrentes da infecção viral por EBV, genes APOBECs podem gerar danos no DNA humano, resultando em mutações pontuais que favorecem a tumorigênese. Para confirmar essa hipótese, realizamos uma abordagem integrativa a fim de avaliar a expressão e a carga mutacional no contexto de APOBEC nessas amostras.

Primeiro, definimos dois grupos de pacientes de acordo com o status EBV, denotados como **EBV+** e **EBV-**. No grupo **EBV-**, encontram-se as classes CIN, GS e MSI. Em seguida, avaliamos os níveis de expressão de cada um dos 7 APOBEC3s (A3A, A3B, A3C, A3D, A3F, A3G e A3H) nos respectivos grupos, a fim de determinar o padrão de mutação destes genes na nossa coorte. As análises consideraram todos os casos onde havia informação do status de EBV, dados de RNA-Seq e SNVs detectados no exoma.

Ao comparar os níveis de expressão entre os grupos **EBV+** e **EBV-**, foi possível observar que seis dos sete genes APOBEC3 estavam upregulados nas 23 amostras **EBV+** quando comparados às amostras **EBV-** (Figura 5), o que está em conformidade com a resposta fisiológica à infecção viral. Microrganismos como bactérias e vírus instigam o processo inflamatório mediado por citocinas como interleucina-2 (IL-2), IL-15 e IL-7, as quais aumentam a expressão de APOBECs (STOPAK et al 2007). Além disso, os nossos dados descrevem a interação de cada um dos membros da família APOBEC com a carcinogênese gástrica.

Dentre os sete membros da família APOBEC3, somente A3A não teve diferença significativa entre os subtipos de câncer gástrico.



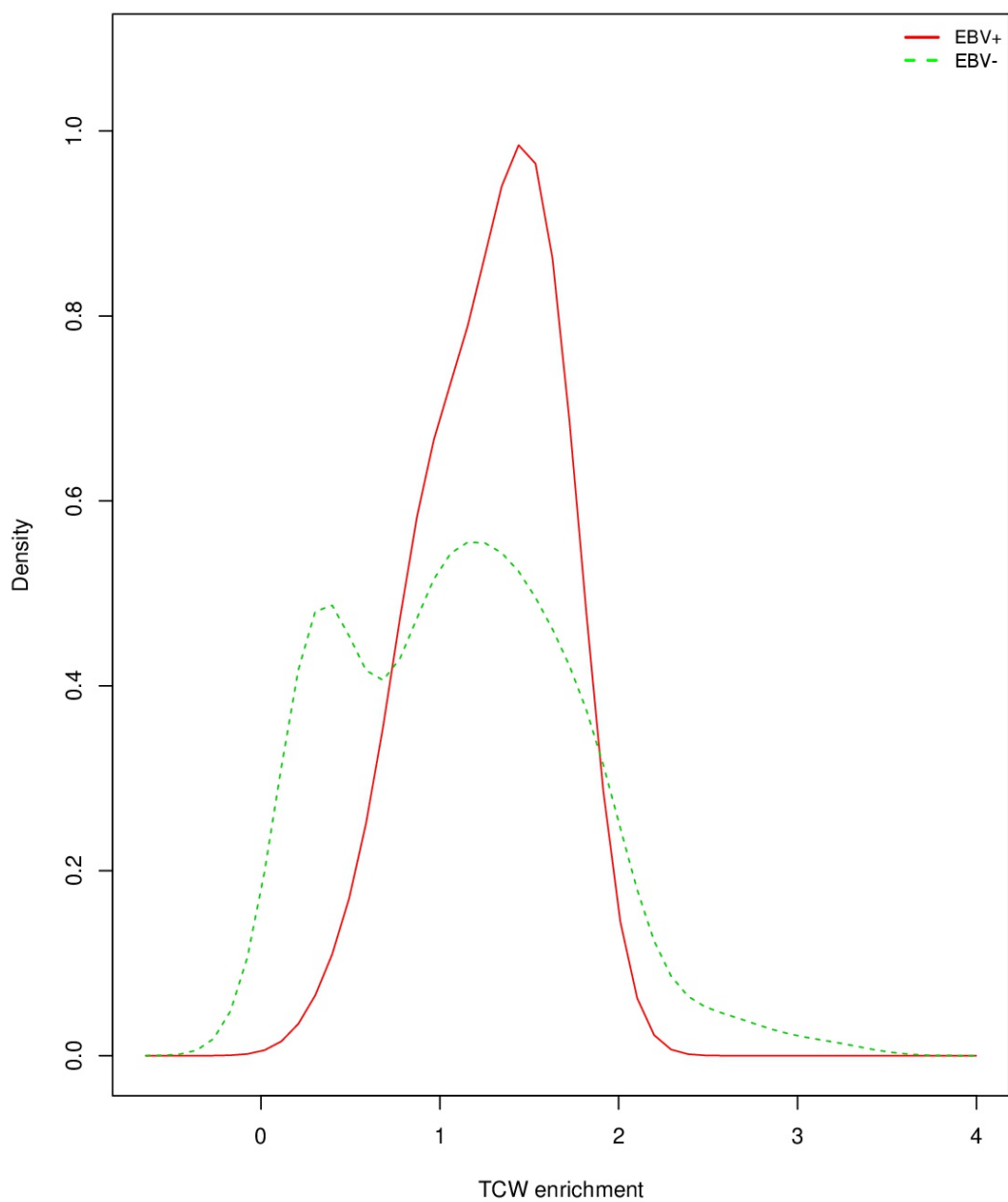
**Figura 5** - Comparação dos níveis de expressão dos membros de APOBEC3 entre os subtipos tumorais de **CG EBV+** e **CG EBV-**. (CIN, GS, MSI), onde CIN está indicado pela cor azul, GS pela cor vermelha, MSI pela laranja e EBV pela cor preta. Os *p-values* obtidos através do teste de Wilcoxon foram:  $* \leq 0,05$ ;  $** \leq 0,0001$  e  $*** \leq 0,00001$ .

Para suportar nossos achados, podemos citar o estudo de HU et al. (2016), onde os autores avaliaram a infecção por EBV em câncer de mama, tanto em cultura celular quanto em biópsia, concluindo que o vírus tem um papel nas primeiras etapas da carcinogênese da mama. Também foi observada uma assinatura mutacional condizente com tumores com maior instabilidade genômica, associada à família de enzimas APOBECs. Além disso, VARTANIAN et al. (2018), ao avaliar a infecção por HPV em lesões cervicais pré-cancerosas e benignas, também observou que a superexpressão de qualquer um dos quatro genes da família APOBEC (A1D, A3A, A3C e A3H) pode ser suficiente para criar um amplo espectro mutante inicial a partir do qual o genoma do câncer pode emergir.

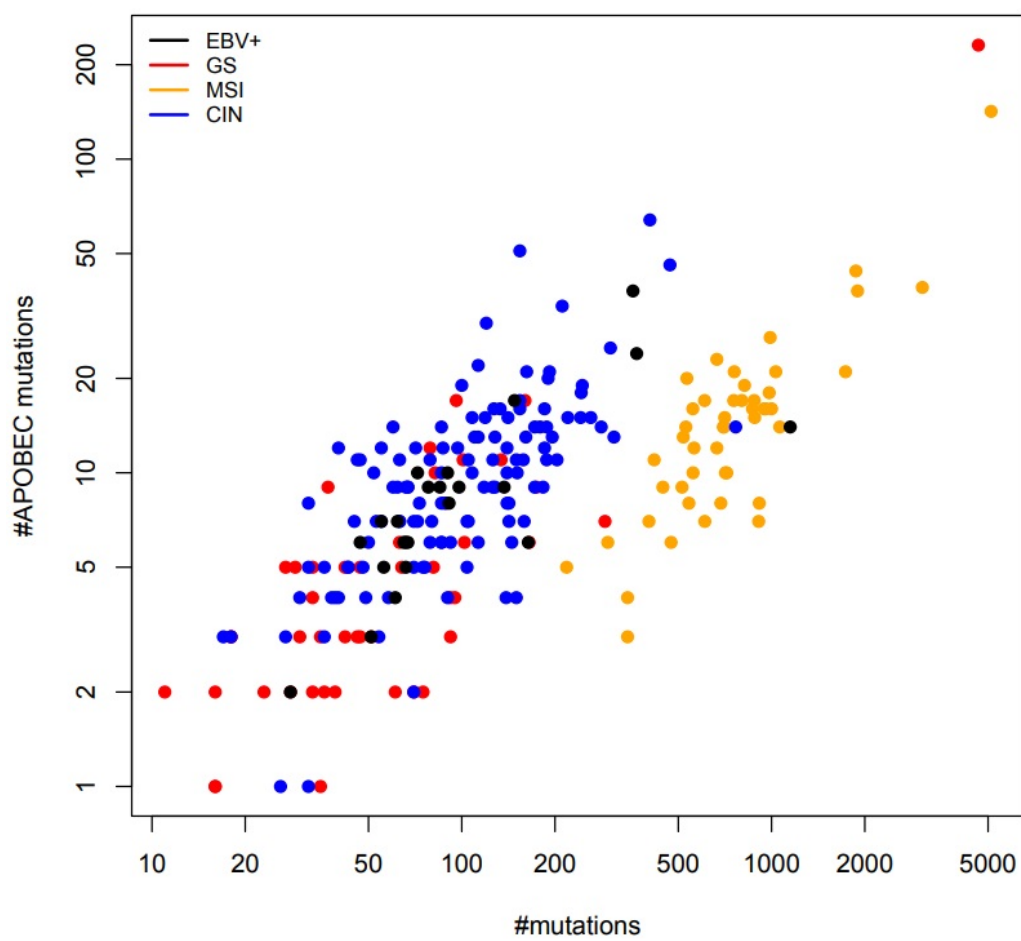
Um dos resultados da ação mutagênica dos membros da família APOBEC3 é o acúmulo de mutações *off-target* caracterizada de uma forma específica, através do motivo TCW→TTW ou TCW→TGW (nucleotídeo mutado está sublinhado, W=A ou T). Apesar da sua importante função fisiológica, este mecanismo de defesa gerado pelas imunidades inata e adaptativa, implica um alto risco de carcinogênese ao induzir diferentes tipos de mutações (PETTERSEN et al 2014). Visto isso, nosso próximo passo foi avaliar se havia um perfil de mutação maior, no contexto TCW, em pacientes com câncer gástrico **EBV+**, em relação aos pacientes **EBV-**. Como pode ser visto na Figura 6, o padrão de mutação TCW se mostrou enriquecido nas amostras **EBV+** quando comparado com as amostras **EBV-** (p=0.01). Este resultado foi validado numa coorte independente da população brasileira de pacientes com câncer gástrico (**EBV+** vs **EBV-**, p=0.042). Mais uma vez, esses resultados suportam o envolvimento de APOBECs na indução de mutações no contexto TCW na

associação com a infecção por EBV, com diferença significativa entre os pacientes **EBV+** e **EBV-**.

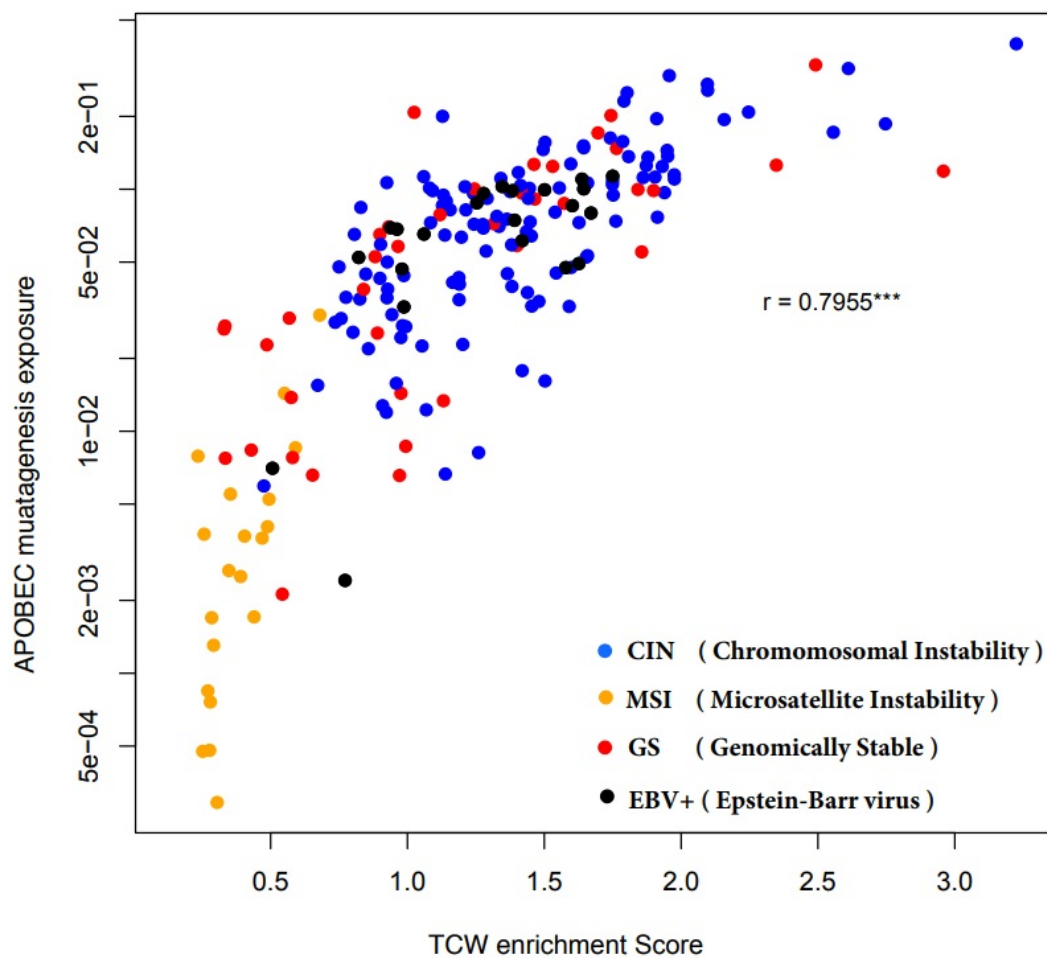
De maneira complementar, ao avaliar a exposição das assinaturas APOBEC descritas no COSMIC, observamos uma alta correlação que corrobora com a aplicabilidade do nosso método e sua contribuição para a carga mutacional em amostras **EBV+** (Figuras 7 e 8). Essas figuras mostram a frequência relativa de mutações no genoma correlacionadas com mutações APOBEC, para cada um dos quatro subtipos de **CG**. Para comprovar que a expressão de APOBECs da Figura 8 é, de fato, advinda do APOBEC, mostramos que quanto maior o score de enriquecimento de TCW, maior a frequência relativa de mutações no contexto APOBEC.



**Figura 6** - Análise do enriquecimento de mutações no contexto TCW nas amostras **EBV+** e **EBV-**. Foi realizado um teste de permutação de igualdade entre as duas densidades indicando que as distribuições são significativamente diferentes ( $p < 0.01$ ).



**Figura 7** - Frequência relativa de mutações nos quatro subtipos de **CG** correlacionada com mutação de APOBEC3 ( $r=0.85$ ).

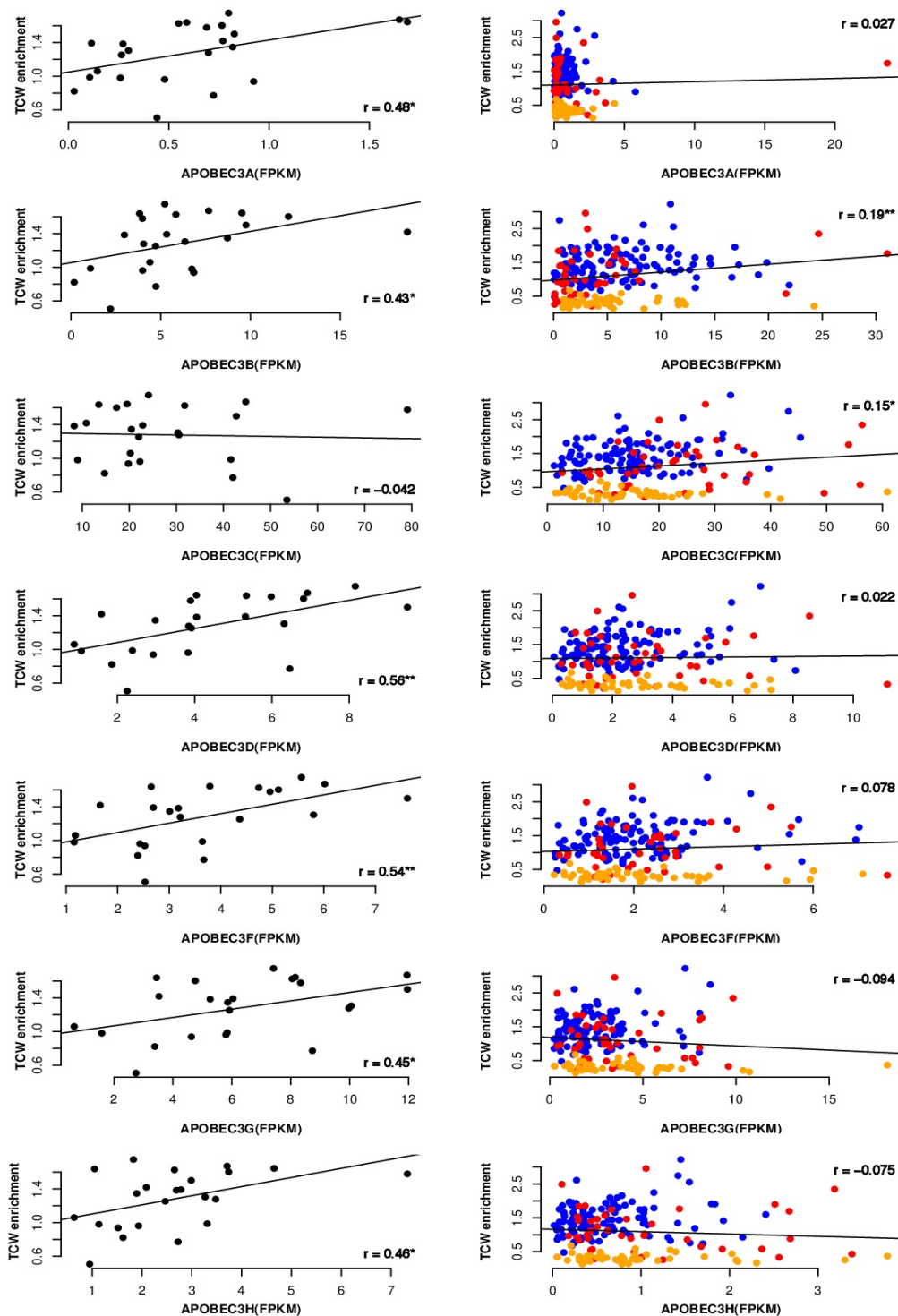


**Figura 8** - Frequência relativa do enriquecimento no contexto TCW correlacionada com expressão de APOBEC3 nos quatro subtipos de **CG**.

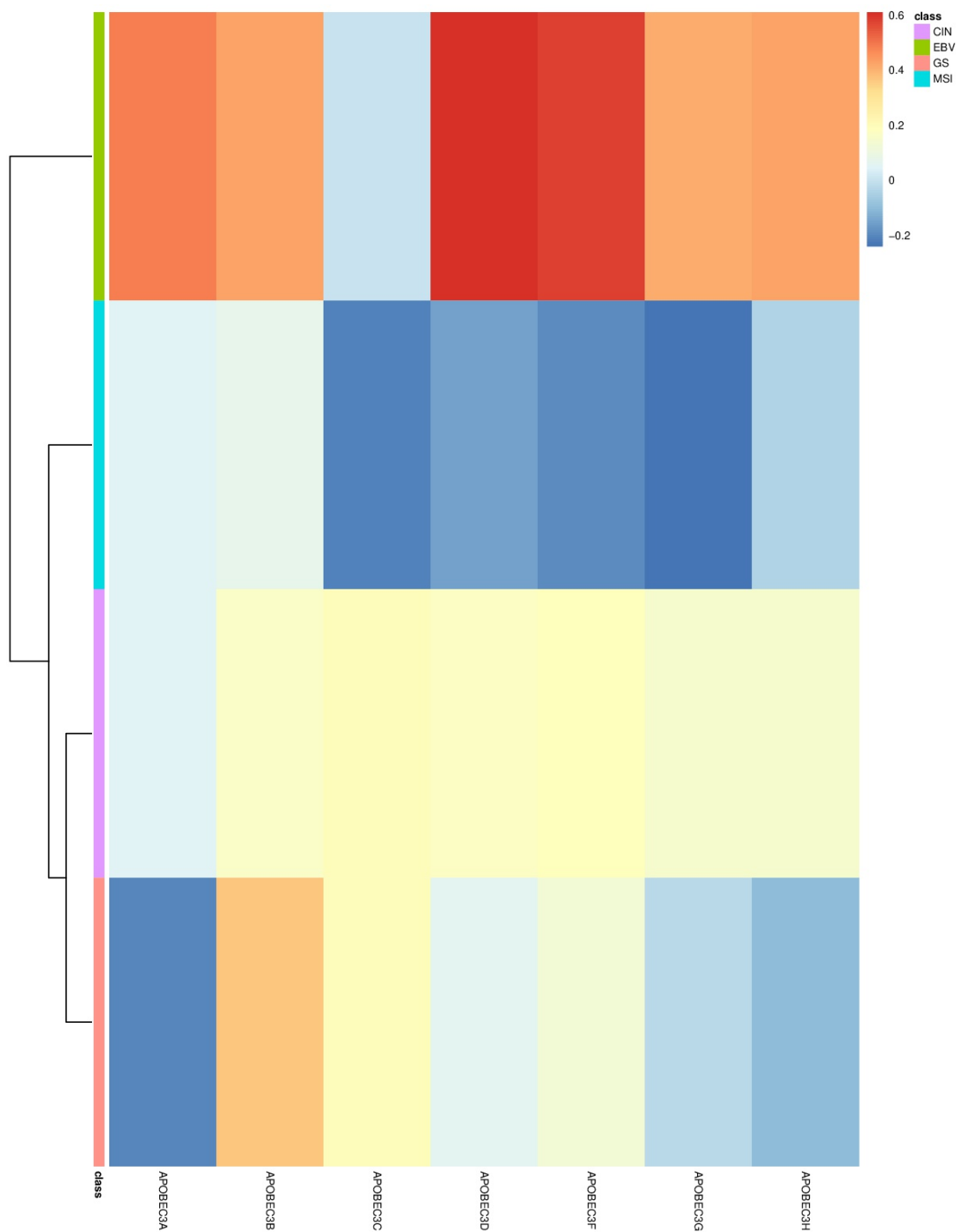


Visto que as amostras de tumor gástrico positivas para o vírus possuíam níveis de expressão de APOBEC3 superior às amostras **EBV-** e, o enriquecimento de mutações no contexto TCW, quando comparadas às amostras negativas para EBV é superior, nós então avaliamos se essas duas variáveis apresentam correlação. Os níveis de expressão foram avaliados em conjunto com o enriquecimento de mutações no contexto TCW para cada amostra. Foi possível observar que 86% dos genes APOBEC3 tiveram correlação positiva entre o enriquecimento da mutação TCW e a sua expressão, nas amostras **EBV+**, de maneira complementar, ao analisar as amostras **EBV-**, não foi observada nenhuma correlação (Figura 9).

Para melhor visualização da correlação entre a expressão dos genes APOBEC3 e mutações no contexto TCW, nós estratificamos os subgrupos de **CG** e realizamos uma clusterização dos valores de correlação da expressão de cada APOBEC com a carga mutacional no contexto TCW. O *heatmap* (Figura 10) sumariza todas as correlações entre os níveis dos transcritos de APOBEC3 e as mutações no contexto TCW. Vale destacar que o agrupamento hierárquico gerou um grupo externo (*outgroup*) que, por sua vez, representa o grupo **EBV+**.



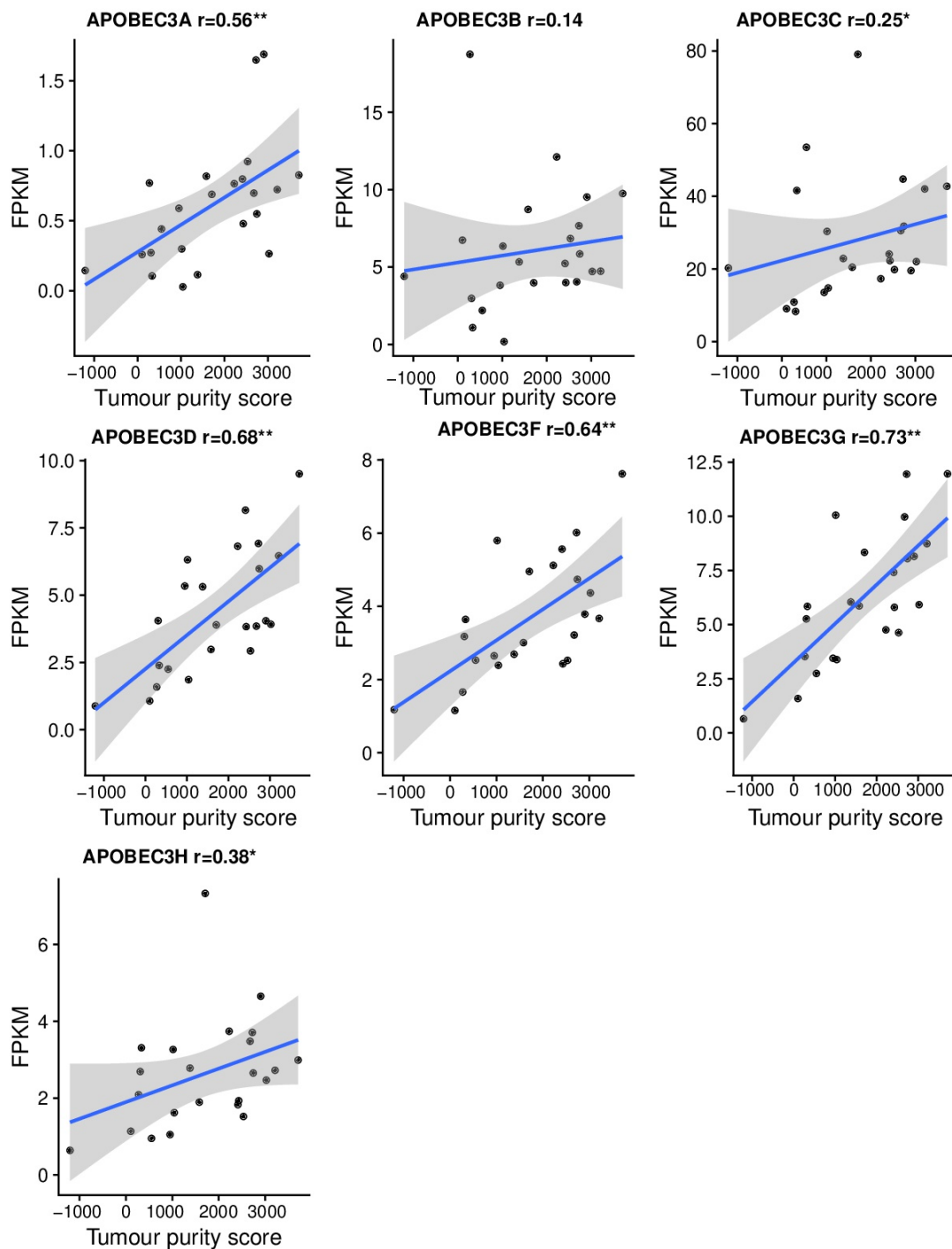
**Figura 9** - Comparação dos níveis de expressão dos genes APOBEC3 e mutações relacionadas a APOBEC3 (enriquecimento TCW). O gráfico de dispersão mostra que a elevada expressão da maioria dos genes APOBEC3 está associada com mais mutações relacionadas a APOBEC3 nas amostras EBV+. Os subtipos moleculares de câncer gástrico estão indicados por pontos coloridos: CIN está indicado pela cor azul, GS pela cor vermelha, MSI pela laranja e EBV pela cor preta.



**Figura 10** - *Heatmap* para todas as correlações entre os níveis de expressão dos genes da família APOBEC3 e o enriquecimento TCW de mutações relacionadas à APOBEC3 através dos subtipos tumorais. Os valores de expressão gênica para os membros da família APOBEC3 são dados como FPKM.

Esses resultados nos levaram a questionar se a expressão de APOBEC3 observada é advinda do próprio tumor ou das células do sistema imune, as quais também expressam APOBEC.

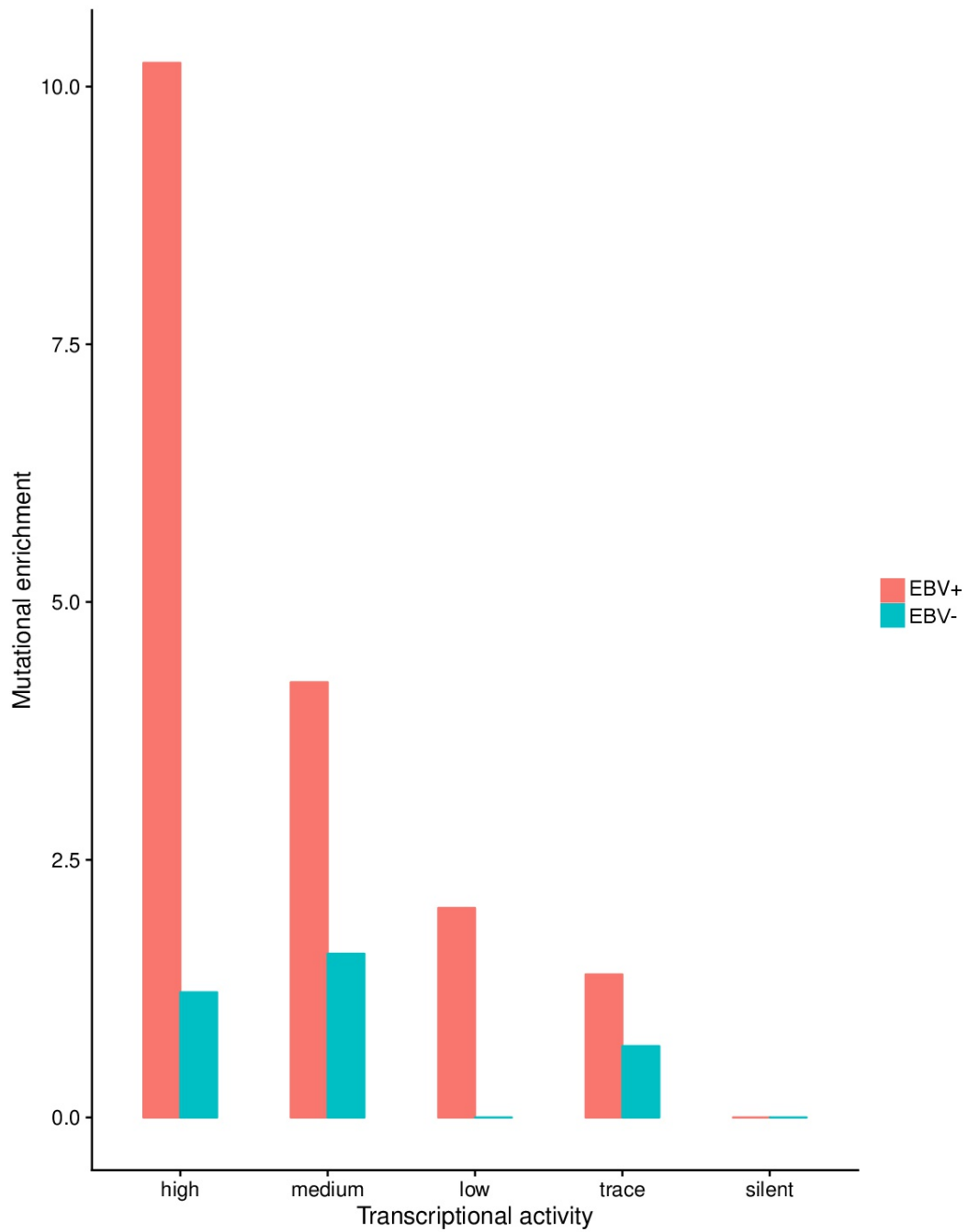
Dessa forma, para confirmar a origem da expressão dos genes APOBEC, nós analisamos a correlação entre os níveis de expressão dos genes APOBEC3s e a pureza do tumor, usando o score de pureza tumoral disponível no TCGA para as respectivas amostras de **CG** previamente calculada. Como pode ser observado na Figura 11, a expressão dos APOBEC3 possui correlação positiva com a pureza tumoral, o que sugere que a atividade desses genes vem, em sua maioria, das células tumorais e não das células do sistema imune.



**Figura 11** - Correlação entre a pureza do tumor e a expressão dos genes APOBEC3.

Os p-valores resultantes dos testes de associação entre amostras pareadas foram:  $* \leq 0,05$ ;  $** \leq 0,001$ .

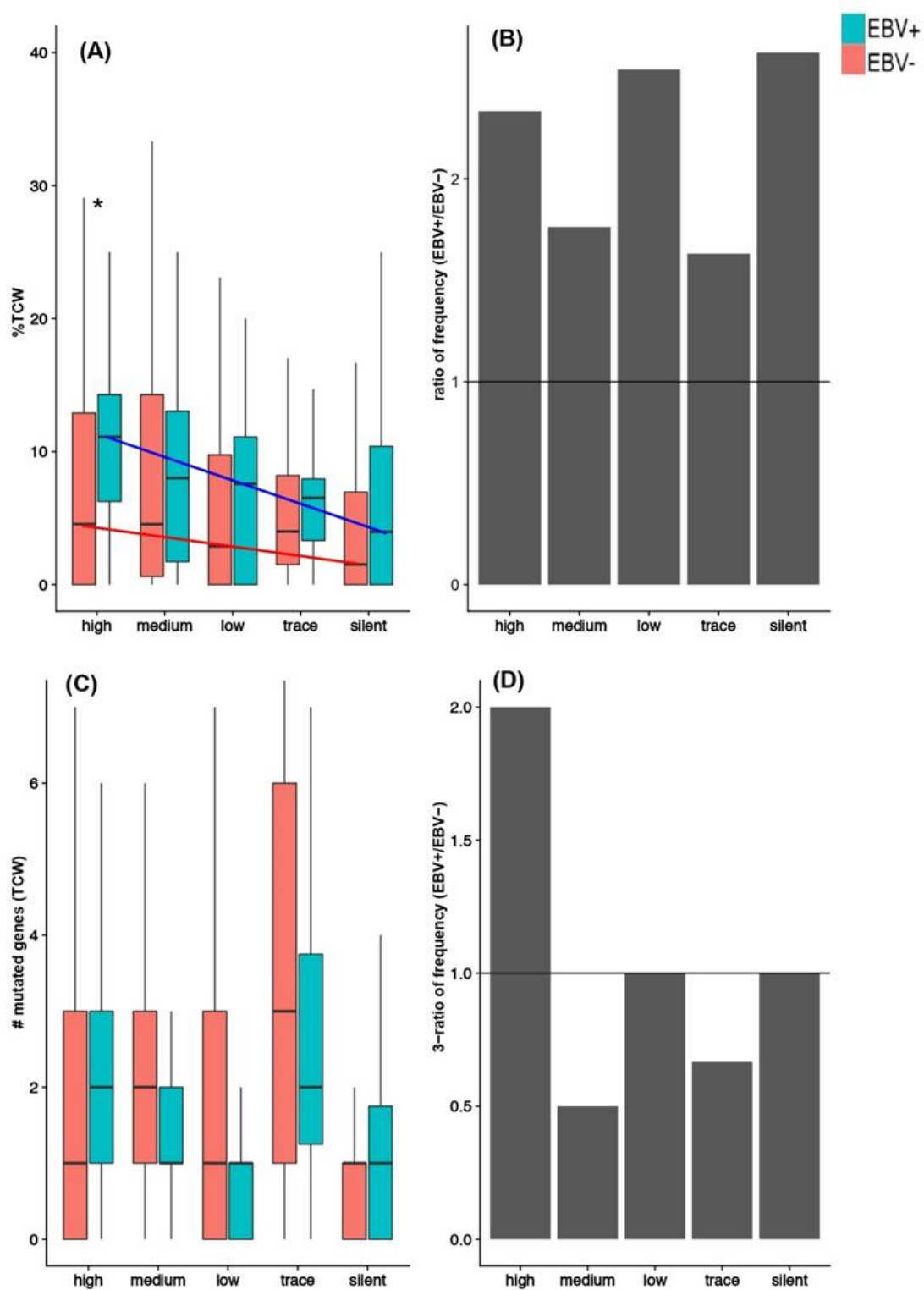
Como foi observada correlação entre a taxa mutacional no contexto TCW e o aumento de expressão de mRNA APOBEC3 nas amostras de câncer gástrico **EBV+**, nós avaliamos também se as mutações, no contexto contexto TCW, ocorriam de maneira uniforme ao longo do genoma. Assim, realizamos uma análise em genes, de acordo com sua atividade transcricional. Desta forma, todos os transcritos detectados nas amostras de câncer gástrico do TCGA foram divididos em cinco categorias de quantis de atividade transcricional, além de avaliar a frequência de mutações TCW em cada uma das categorias. Como observado na Figura 12, a média do enriquecimento de mutações no contexto TCW é diretamente relacionada à atividade transcricional nas amostras **EBV+**, sendo que a mesma tendência não foi observada nas amostras **EBV-**. Do ponto de vista bioquímico, a desaminação das bases nitrogenadas pode ser dependente da transcrição do DNA alvo. Visto que amostras **EBV+** possuem taxas de transcrição mais altas, podemos supor que a desaminação foi facilitada por esse evento, pausando a ação da RNA polimerase, formando *R-loops* e estruturas secundárias, o que aumentaria o tamanho e tempo de transcrição da região ainda em fita simples, deixando o DNA vulnerável para a ação dos APOBECs. Além disso, podemos supor que a ação do vírus está atuando de forma que inibe as vias de reparo de DNA, como sugerido nos trabalhos de e QADRI et al. (2011) e NIKITIN e LUFTIG (2012), onde os autores observaram que durante a infecção por oncovírus, havia inibição da via de *Transcription-capture repair* (TCR), a qual pode ter um papel mais forte em genes com taxas de transcrição mais altas, levando à maior tempo de exposição do ssDNA à ação de mutagênicos.



**Figura 12** - Taxa mutacional no contexto TCW é proporcional aos níveis de transcrição de vários genes nas amostras **EBV+**.

A atividade transcricional em cada uma das classes analisadas pode ser uma evidência de que a infecção está modulando a via de reparo de DNA. Tendo isso em vista, fizemos a deconvolução da taxa mutacional mostrada na Figura 12, avaliamos a taxa de mutação no contexto TCW entre as amostras EBV+ e EBV-, quanto às classes de transcrição gênica; a taxa de genes mutados em cada classe; e a taxa de genes mutados no contexto TCW (Figura 13). A análise também avaliou o razão da frequência relativa entre as amostras **EBV+** e **EBV-**, onde podemos observar que em amostras **EBV+**, no contexto TCW, há maior frequência de genes mutados. Assim, observamos uma correlação negativa, onde quanto mais expresso é o gene, menor a taxa de mutação apresentada, e vice-versa. Sugerindo a ação da via de reparo TCR, a qual atua em genes com alta taxa de transcrição, levando à correlação inversa entre a taxa de expressão gênica e sua densidade mutacional, como pode ser visto nos nossos dados. No estudo de MATHONNET et al. (2004), os autores relatam que a expressão da proteína X da Hepatite B (HBx), envolvida em carcinogênese hepática, pode reduzir a eficácia da via de reparo acoplada à transcrição (TCR). O mesmo processo pode estar ocorrendo na nossa análise, ao avaliarmos a Figura 13, onde podemos supor que a infecção por EBV esteja modulando o sistema de reparo de DNA, em **CG**, através do mecanismo descrito por GOODWIN et al. (2015), o qual descreve que vírus está modulando o sistema imune para “usurpar” a maquinaria energética do hospedeiro para garantir a sua persistência no organismo. Além disso, a via TCR também já foi descrita apresentando taxas de polimorfismos, os quais podem afetar o prognóstico de pacientes com câncer gástrico ou de bexiga (XING et al 2012; LIU et al. 2016).





**Figura 13** - Frequência relativa da taxa de mutações no contexto TCW e taxa de todos os genes mutados, em cinco classes de transcrição gênica (High, Medium, Low, Trace e Silent), entre pacientes com CG EBV+ e EBV-.

## **5.2 ANÁLISE DA COORTE DO A.C.CAMARGO CANCER CENTER**

### **5.2.1 Genotipagem de EBV**

A fim de padronizar a técnica a ser usada na detecção de EBV nos pacientes com AdG do presente estudo, utilizamos amostra de **AdG** positivo para EBV gentilmente cedida pela Dra. Maria Dirlei Begnami. Primeiramente, a presença de EBV nessa amostra foi realizada por meio de imunohistoquímica pelo departamento de Patologia, e em seguida, confirmada por genotipagem através da PCR em tempo real pelo presente estudo. Para tanto, foi feita extração de DNA da amostra EBV positiva, conservada em bloco de parafina, usando dois kits comerciais ReliaPrep™ FFPE gDNA Miniprep System (Promega) e QIAamp DNA FFPE Tissue kit (Qiagen), usando duas quantidades diferentes de amostra, para fins de padronização da técnica.

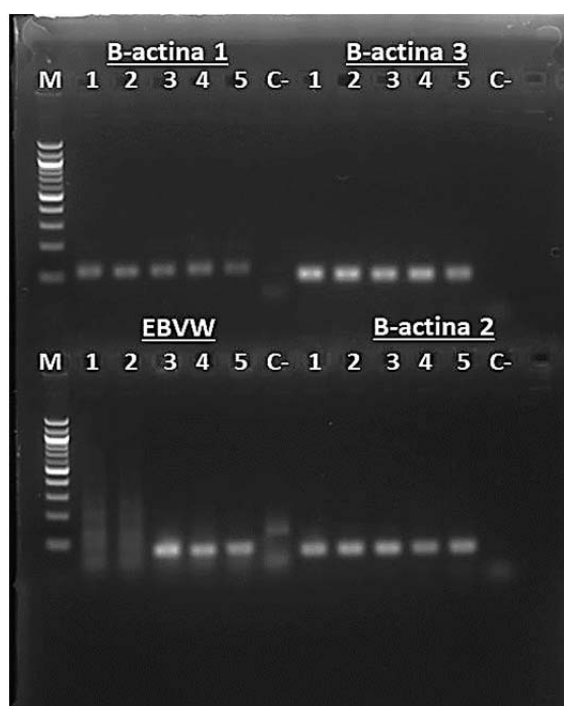
A quantidade de DNA foi comparada com a amostra de DNA, do mesmo paciente, obtida de biópsia a fresco, extraída pelo Biobanco do A.C.Camargo Cancer Center, podendo confirmar a eficiência de extração dos kits comerciais para a amostra em análise e a reprodutibilidade da técnica.

A genotipagem dos tumores para a presença de EBV foi realizada no presente estudo, pois esta prática não faz parte da rotina clínica. Assim, todas as amostras usadas neste projeto foram analisadas quanto à presença do vírus EBV.

Para a genotipagem, primeiramente foi realizado o levantamento de artigos científicos que realizaram a quantificação viral de EBV. Foram selecionadas sequências de oligonucleotídeos específicas para a região conservada do vírus e, em

seguida, foi realizada um análise de bioinformática para escolher um oligonucleotídeo cuja sequência possua múltiplas cópias no genoma, para garantir uma maior sensibilidade e, assim, eficiência durante a amplificação.

Os testes para a eficiência, tanto da extração de DNA do tumor em bloco de parafina, quanto dos oligonucleotídeos escolhidos, foram realizados através da PCR convencional, e o resultado foi visualizado em gel de agarose (Figura 14). Foram usados duas amostras de DNA, comercial e do projeto de esclerose múltipla, como amostras controle. Como esperado, não houve amplificação da sequência de BamH1W (EBVW) nas amostras controle (DNA comercial).



**Figura 14** - Imagem ilustrativa do gel de agarose 2%, com amplificação de três sequências diferentes de  $\beta$ -actina ( $\beta$ -actina1,  $\beta$ -actina2 e  $\beta$ -actina3) e da sequência de BamH1W (EBVW). 1. DNA comercial; 2. DNA de paciente com esclerose múltipla; 3. DNA fornecido pelo Biobanco do paciente com CG EBV+; 4. DNA do paciente com CG EBV+, extraído pelo kit da Promega; 5. DNA do paciente com CG EBV+, extraído pelo kit da Quagen; C- Controle negativo.

Foi feito levantamento epidemiológico de 32 DNAs de amostras pareadas (biópsia e **SG**), onde podemos observar que a maioria das amostras de **AdG** foi do tipo difuso (22/32). Ao ser realizada a genotipagem das biópsias quanto à presença de EBV, 9% das amostras tiveram razão acima de 1 (positivas), sendo a maioria do tipo intestinal. Já a análise de **SG** mostrou positividade para EBV em 22% das amostras. No entanto, somente amostras de um paciente se mostraram positivas para o vírus tanto na biópsia quanto no **SG**.

Foi realizada genotipagem para EBV de um outro grupo de amostras composto de biópsias pareadas com leucócitos. No total, foram 23 DNAs de biópsias e 12 de leucócitos. Não foi observada amplificação da região referente ao vírus nas amostras de leucócitos, e somente duas (8,9%) biópsias foram consideradas positivas para EBV.

Além das biópsias pareadas com **SG**, também avaliamos 115 biópsias não pareadas, das quais 4,3% amplificaram fragmento do vírus; estando abaixo da média de 10% descrita na literatura (IIZASA et al. 2012).

Ao todo, 504 amostras de DNA obtidas a partir de **SG** (240 pacientes caso e 137 pacientes controle) foram analisadas quanto à presença de EBV, das quais, 11,1% foram positivas. Entre as amostras de **SG** de pacientes controle, somente quatro (2,9%) tiveram razão EBV/humano maior que 1. Já, entre as amostras de pacientes caso, 16 (12,7%) foram consideradas positivas para a presença do vírus, sendo que o subtipo misto apresentou 16,7% de amplificação do vírus (2/12), o subtipo intestinal - 13,6% (6/44) e o difuso teve 11,4% das amostras positivas para EBV (8/70).

Do total de amostras tumorais analisadas (pareadas com leucócito, pareadas com **SG** e sem par), 6% foram positivas para EBV, próximo aos valores revisados no trabalho de NASEEM et al. (2018).

Esse conjunto de dados, no entanto, não foi suficiente para correlacionar a presença do EBV no **AdG** da nossa coorte. Esse dado pode ser explicado pelo número reduzido de pacientes. O comportamento do vírus também pode ser um fator por trás da falta de correlação entre os diferentes tipos de amostras (**SG**, biópsia e leucócito).

A presença de EBV no **SG** pode ser indício de uma infecção mais recente, consequentemente, uma infecção mais prolongada teria a presença do vírus nas células epiteliais do estômago e não no **SG** ou leucócitos. Para comprovar essa hipótese, seria necessário um número amostral maior.

### **5.2.2 Hibridização *In Situ***

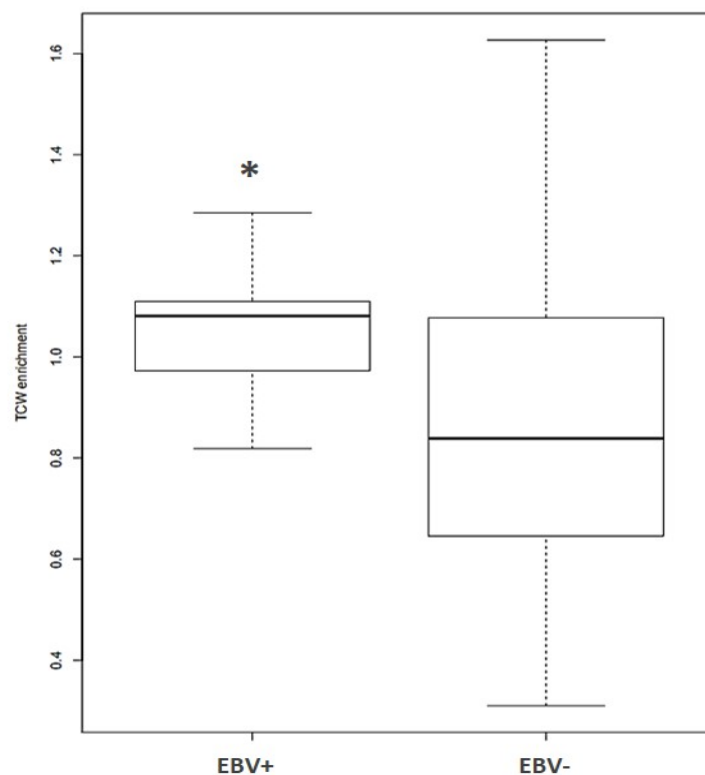
A fim de confirmar os achados da presença de EBV através da PCR em Tempo Real, foram separadas onze amostras para análise por hibridização in situ (ISH). As amostras foram separadas em três categorias, onde cinco amostras foram consideradas totalmente negativas para a presença do vírus através da técnica de PCR em Tempo Real (qRT-PCR), três amostras com baixa amplificação de EBV, e três amostras com alta amplificação de EBV. Do total de onze amostras, três foram positivas para EBV através do ISH. Das amostras negativas por qRT-PCR todas foram negativas também por ISH. Comprovando a eficiência da genotipagem de EBV por PCR em tempo real, como já descrito em outros trabalhos (WHITEHURST et al. 2009; MA et al. 2011; KHODDAMI et al. 2015).

### 5.2.3 Análise de Instabilidade de Microssatélites

Visto que infecções virais podem levar à instabilidade de microssatélites (MSI), nós avaliamos 44 amostras tumorais pareadas com DNA de leucócitos. As reações foram realizadas usando cinco conjuntos de oligonucleotídeos (NR-27, NR-21, NR24, BAT-25 e BAT-26) marcados com fluoróforos, no sequenciador API 3130XL (Applied Biosystems). Como controle interno, foi usado GeneScan 350 ROX. As análises foram feitas na plataforma Thermo Fisher Cloud para Microsatellite Analysis. Considerando zero oligonucleotídeos, alterados como MSI-S, de 1 a 2 oligonucleotídeos alterados como MSI-L, e pelo menos três oligonucleotídeos alterados como MSI-H, obtivemos os seguintes resultados: 20,7% (N=12) das amostras foram MSI-S (MSI-Stable), 63,8% (N=37) MSI-L (MSI-Low) e 15,5% (N=9) MSI-H (MSI-High). De acordo com o estudo de CORTES-CIRIANO et al. (2017), a frequência de MSI-H em amostras de câncer de estômago do TCGA está em torno de 21.9%, frequência próxima a encontrada (19.09%) no estudo de BONNEVILLE et al. (2017), onde foi analisada a instabilidade de microssatélites em 39 tipos tumorais. No entanto, o estudo de KIM e et al. (2016), realizado com 434 pacientes de CG da Faculdade de Medicina de Konkuk, na Coréia, descreve uma frequência de 9.4% de MSI-H, a qual é próxima a encontrada no nosso estudo. Dessa forma, podemos concluir que os valores encontrados no nosso trabalho estão próximos aos descritos em outros trabalhos encontrados na literatura.

Para finalizar, relatamos que foram realizadas análises de correlação entre a infecção por EBV e dados clínicos dos pacientes da nossa instituição, no entanto, devido ao número restrito de casos **EBV+**, não foi possível observar correlação entre a taxa de mutações no contexto TCW e o desfecho clínico. Contudo, o projeto

institucional do qual o presente estudo faz parte, continua recrutando pacientes e dados mais apurados deverão ser obtidos com um *follow-up* mais longo, possibilitando realizar uma correlação positiva entre as variáveis analisadas. Por outro lado, é descrito na literatura (JAMAL-HANJANI et al. 2017; CESCA TCGA, dados não publicados) que a atividade de APOBECs pode prevalecer em estádios mais tardios em tumores como câncer cervical associado a HPV, adenocarcinoma de estadios tardios ou tumores de células escamosas, associados à infecção viral. Na coorte do A.C.Camargo Cancer Center também pudemos observar padrão semelhante, onde enriquecimento de mutações TCW foi observado em **CG EBV+** ( $p < 0.05$ ) (Figura 15).



**Figura 15** – Boxplot ilustrando o enriquecimento do motivo TCW em amostras de AdG EBV+ e EBV- da coorte do A.C.Camargo Cancer Center ( $p < 0.05$ ).

## 6 CONCLUSÕES

Visto que a infecção viral ativa o nosso sistema imune, e uma das linhas de defesa é um conjunto de genes chamados APOBECs, os quais têm como um dos objetivos, inibir a ação viral nas células do hospedeiro. Essa resposta, em condições anormais, pode levar ao dano de DNA. Dessa forma, nós avaliamos se é possível observar um padrão de mutação associado aos APOBECs, em CG com a infecção por EBV, e se esse mecanismo pode ser uma fonte de dano adicional na tumorigênese.

Para provar a nossa hipótese, fizemos uso de um conjunto de dados públicos (TCGA), a partir do qual concluímos que as amostras de CG classificadas como **EBV+** possuem expressão de APOBEC3 significativamente aumentada, quando comparadas com as amostras de subtipos tumorais **EBV-**.

As análises laboratoriais de genotipagem de EBV não encontraram nenhuma correlação entre a presença do vírus no suco gástrico e/ou na biópsia, e o desenvolvimento do tumor. No entanto, esses dados são interessantes para análises mais robustas e com maior número amostral, podendo ser de grande valia para estratificação dos pacientes para tratamentos individualizados num futuro próximo.

Concluímos que a alta expressão de APOBEC em câncer gástrico com infecção por EBV pode aumentar as taxas mutacionais durante a tumorigênese deste tipo tumoral. Além disso, a infecção por EBV e consequente alta expressão de APOBEC3 pode ser um evento secundário no desenvolvimento do câncer gástrico, levando assim a novas oportunidades para investigar estas causas no que se refere ao prognóstico dos pacientes.



## 7 REFERÊNCIAS BIBLIOGRÁFICAS

Alexandrov LB, Nik-Zainal S, Wedge DC, et al. Signatures of mutational processes in human cancer. **Nature** 2013; 500:415-21.

Bass AJ, Thorsson V, Shmulevich I, et al. Comprehensive molecular characterization of gastric adenocarcinoma. **Nature** 2014; 513:202-9.

Bonneville R, Krook MA, Kautto EA, et al. Landscape of Microsatellite Instability Across 39 Cancer Types. **JCO Precis Oncol** 2017; 1-15.

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

Burns MB, Leonard B, Harris RS. APOBEC3B: pathological consequences of an innate immune DNA mutator. **Biomed J** 2015; 38:102-10.

Camargo MC, Kim WH, Chiaravalli AM, et al. Improved survival of gastric cancer with tumour Epstein-Barr virus positivity: an international pooled analysis. **Gut** 2014; 63:236-43.

Chang MS, Lee HS, Kim CW, Kim YI, Kim WH. Clinicopathologic characteristics of Epstein-Barr virus-incorporated gastric cancers in Korea. **Pathol Res Pract** 2001; 197:395-400.

Charames GS, Bapat B. Genomic instability and cancer. **Curr Mol Med** 2003; 3:589-96.

Chia NY, Tan P. Molecular classification of gastric cancer. **Ann Oncol** 2016; 27:763-9.

Eligio P, Delia R, Valeria G. EBV Chronic Infections. **Mediterr J Hematol Infect Dis** 2010; 2:e2010022.

Epstein RJ. A periodic table for cancer. **Future Oncol** 2015; 11:785-800.

Goodwin CM, Xu S, Munger J. Stealing the keys to the kitchen: viral manipulation of the host cell metabolic network. **Trends Microbiol** 2015; 23:789-98.

Grabsch HI, Tan P. Gastric cancer pathology and underlying molecular mechanisms. **Dig Surg** 2013; 30:150-8.

Gulley ML. Genomic assays for Epstein-Barr virus-positive gastric adenocarcinoma. **Exp Mol Med** 2015; 47:e134.

Henderson S, Chakravarthy A, Su X, Boshoff C, Fenton TR. APOBEC-mediated cytosine deamination links PIK3CA helical domain mutations to human papillomavirus-driven tumor development. **Cell Rep** 2014; 7:1833-41.

Hollingworth R, Grand RJ. Modulation of DNA damage and repair pathways by human tumour viruses. **Viruses** 2015; 7:2542-91.

Hu B, El Hajj N, Sittler S, Lammert N, Barnes R, Meloni-Ehrig A. Gastric cancer: Classification, histology and application of molecular pathology. **J Gastrointest Oncol** 2012; 3:251-61.

Hu H, Luo ML, Desmedt C, et al. Epstein-Barr virus infection of mammary epithelial cells promotes malignant transformation. **EBioMedicine** 2016; 9:148-60.

Iizasa H, Nanbo A, Nishikawa J, Jinushi M, Yoshiyama H. Epstein-Barr Virus (EBV)-associated gastric carcinoma. **Viruses** 2012; 4:3420-39.

Jamal-Hanjani M, Wilson GA, McGranahan N, et al. Tracking the evolution of non-small-cell lung cancer. **N Engl J Med** 2017; 376:2109-21.

Jha HC, Pei Y, Robertson ES. Epstein-barr virus: Diseases linked to infection and transformation. **Front Microbiol** 2016; 7:1-16.

Karimi P, Islami F, Anandasabapathy S, Freedman ND, Kamangar F. Gastric cancer: descriptive epidemiology, risk factors, screening, and prevention. **Cancer Epidemiol Biomarkers Prev** 2014; 23:700-13.

Khoddami M, Nadji SA, Dehghanian P, Vahdatinia M, Shamshiri AR. Detection of Epstein-Barr Virus DNA in Langerhans Cell Histiocytosis. **Jundishapur J Microbiol** 2015; 8:e27219.

Kim N, Jinks-Robertson S. Transcription as a source of genome instability. **Nat Rev Genet** 2012; 13:204-14.

Kim YB, Lee SY, Kim JH, et al. Microsatellite instability of gastric and colorectal cancers as a predictor of synchronous gastric or colorectal neoplasms. **Gut Liver** 2016; 10:220-7.

Klein DN, Kotov R, Bufferd SJ. Personality and depression: explanatory models and review of the evidence. **Annu Rev Clin Psychol** 2011; 7:269-95.

Küppers R. B cells under influence: transformation of B cells by Epstein-Barr virus. **Nat Rev Immunol** 2003; 3:801-12.

Langie SA, Koppen G, Desaulniers D, et al. Causes of genome instability: the effect of low dose chemical exposures in modern society. **Carcinogenesis** 2015; 36(Suppl 1):61-88.

Lek M, Karczewski KJ, Minikel EV, et al. Analysis of protein-coding genetic variation in 60,706 humans. **Nature** 2016; 18:536:285-91.

Li H, Durbin R. Fast and accurate short read alignment with Burrows-Wheeler Transform. **Bioinformatics** 2009; 25:1754-60.

Lin Y, Wu Z, Guo W, Li J. Gene mutations in gastric cancer: a review of recent next-generation sequencing studies. **Tumour Biol** 2015; 36:7385-94.

Liu J, Deng N, Xu Q, et al. Polymorphisms of multiple genes involved in NER pathway predict prognosis of gastric cancer. **Oncotarget** 2016; 30:48130-42.

Luo GG, Ou JJ. Oncogenic viruses and cancer. **Viol Sin** 2015; 30:83-4.

Ma SD, Hegde S, Young KH, et al. A new model of Epstein-Barr virus infection reveals an important role for early lytic viral protein expression in the development of lymphomas. **J Virol** 2011; 85:165-77.

Mathonnet G, Lachance S, Alaoui-Jamali M, Drobetsky EA. Expression of hepatitis B virus X oncoprotein inhibits transcription-coupled nucleotide excision repair in human cells. **Mutat Res** 2004; 554:305-18.

Matsumoto T, Shimizu T, Takai A, Marusawa H. Exploring the mechanisms of gastrointestinal cancer development using deep sequencing analysis. **Cancers (Basel)** 2015; 7:1037-51.

Mesri EA, Feitelson MA, Munger K. Human viral oncogenesis: a cancer hallmarks analysis. **Cell Host Microbe** 2014; 15:266-82.

Ministério da Saúde. Instituto Nacional de Cancer José Alencar Gomes da Silva. **Estimativa 2016: incidência de câncer no Brasil**. Rio de Janeiro, INCA; 2015.

Ministério da Saúde. Instituto Nacional do Câncer. **Estimativa 2018: incidência de câncer no Brasil**. Rio de Janeiro, INCA; 2018.

Morales-Sánchez A, Fuentes-Pananá E. Human viruses and cancer. **Viruses** 2014; 6:4047-79.

Moris A, Murray S, Cardinaud S. AID and APOBECs span the gap between innate and adaptive immunity. **Front Microbiol** 2014; 5:534.

Naseem M, Barzi A, Brezden-Masley C, et al. Outlooks on Epstein-Barr virus associated gastric cancer. **Cancer Treat Rev** 2018; 66:15-22.

Naslavsky MS, Yamamoto G, de Almeida TF, et al. Exomic variants of an elderly cohort of Brazilians in the ABraOM database. **Hum Mutat** 2017; 38:751-63.

Nikitin PA, Luftig MA. The DNA damage response in viral-induced cellular transformation. **Br J Cancer** 2012; 106:429-35.

Park ES, Do IG, Park CK, et al. Cyclooxygenase-2 is an independent prognostic factor in gastric carcinoma patients receiving adjuvant chemotherapy and is not associated with EBV infection. **Clin Cancer Res** 2009; 15:291-8.

Park SH, Han JK, Kim TK, et al. Unusual gastric tumors: radiologic-pathologic correlation. **RadioGraphics** 1999; 19:1435-46.

Pettersen HS, Galashevskaya A, Doseth B, et al. AID expression in B-cell lymphomas causes accumulation of genomic uracil and a distinct AID mutational signature. **DNA Repair** 2015; 25:60-71.

Piazuelo MB, Correa P. Gastric cancer: overview. **Colomb Med (Cali)** 2013; 44:192-201.

Pinto-de-Sousa J, David L, Seixas M, Pimenta A. Clinicopathologic profiles and prognosis of gastric carcinomas from the cardia, fundus/body and antrum. **Dig Surg** 2001; 18:102-10.

Qadri I, Fatima K, AbdeL-Hafiz H. Hepatitis B virus X protein impedes the DNA repair via its association with transcription factor, TFIIH. **BMC Microbiol** 2011; 11:48.

Ratti M, Lampis A, Hahne JC, Passalacqua R, Valeri N. Microsatellite instability in gastric cancer: molecular bases, clinical perspectives, and new treatment approaches. **Cell Mol Life Sci** 2018; 75:4151-62.

Riquelme I, Saavedra K, Espinoza JA, et al. Molecular classification of gastric cancer: Towards a pathway-driven targeted therapy. **Oncotarget** 2015; 6:24750-79.

Robbiani DF, Deroubaix S, Feldhahn N, et al. Plasmodium infection promotes genomic instability and AID-dependent B cell lymphoma. **Cell** 2015; 162:727-37.

Roberts SA, Lawrence MS, Klimczak LJ, et al. An APOBEC cytidine deaminase mutagenesis pattern is widespread in human cancers. **Nat Genet** 2013; 45:970-6.

Song HJ, Kim KM. Pathology of Epstein-Barr virus-associated gastric carcinoma and its relationship to prognosis. **Gut Liver** 2011; 5:143-8.

Stopak KS, Chiu YL, Kropp J, Grant RM, Greene WC. Distinct patterns of cytokine regulation of APOBEC3G expression and activity in primary lymphocytes, macrophages, and dendritic cells. **J Biol Chem** 2007; 282:3539-46.

Thompson MP, Kurzrock R. Epstein-Barr virus and cancer. **Clin Cancer Res** 2004; 10:803-21.

Van Beek J, Zur Hausen A, Klein Kranenbarg E, et al. EBV-positive gastric adenocarcinomas: a distinct clinicopathologic entity with a low frequency of lymph node involvement. **J Clin Oncol** 2004; 22:664-70.

Vartanian JP, Guetard D, Henry M, Wain-Hobson S. Evidence for editing of human papillomavirus DNA by APOBEC3 in benign and precancerous lesions. **Science** 2008; 320:230-3.

Whitehurst CB, Ning S, Bentz GL, et al. The Epstein-Barr virus (EBV) deubiquitinating enzyme BPLF1 reduces EBV ribonucleotide reductase activity. **J Virol** 2009; 83:4345-53.

Xing J, Dinney CP, Shete S, et al. Comprehensive pathway-based interrogation of genetic variations in the nucleotide excision DNA repair pathway and risk of bladder cancer. **Cancer** 2012; 118: 205-15.

Yoshihara K, Shahmoradgoli M, Martinez E, et al. Inferring tumour purity and stromal and immune cell admixture from expression data. **Nat Commun** 2013; 4:2612.

Zou J, Wang C, Ma X, Wang E, Peng G. APOBEC3B, a molecular driver of mutagenesis in human cancers. **Cell Biosci** 2017; 7:29.

## Anexo 1 – Carta de aprovação do Comitê de Ética em Pesquisa-CEP



**A.C. Camargo  
Cancer Center**

**Comitê de Ética em  
Pesquisa - CEP**

São Paulo, 19 dezembro de 2016.

Ao  
**Dr. Emmanuel Dias Neto**  
**Orientador: Dr. Israel Tojal da Silva**  
**Aluna: Irina Gueroldovna Bobrovnitchaia (Doutorado)**

**Ref.: Projeto de Pesquisa nº. 2134/15C**  
**“Análise de assinaturas mutacionais no prognóstico de pacientes com adenocarcinoma gástrico e infecção por EBV”.**

Os membros do Comitê de Ética em Pesquisa em Seres Humanos da Fundação Antonio Prudente – A.C. Camargo Cancer Center, em sua última reunião de **06/12/2016**, **tomaram conhecimento e aprovaram** os seguintes documentos:

- Solicitação de dispensa da submissão da documentação obrigatória e análise ética do projeto acima mencionado por se tratar de um projeto afiliado ao temático intitulado “Genômica de Adenocarcinomas Gástricos.”, registrado neste CEP sob nº 2134/15. O projeto afiliado em referência será Doutorado da aluna *Irina Gueroldovna Bobrovnitchaia*
- Tese de Doutorado.

Atenciosamente,

  
**Dr. Jefferson Luiz Gross**

**1º Vice-Coordenador do Comitê de Ética em Pesquisa**

1/1