[[Assinatura gênica em melanoma Braf mutado: análise exploratória e descritiva]]

Palavras-Chave: melanoma, assinatura gênica, *data science*

Autores/as:

Anderson Gabriel dos Reis Ortelan [UNICAMP]

Alessandra Valéria de Sousa Faria (Co-orientadora) [UNICAMP]

Prof./ª Dr./ª Carmen Veríssima Ferreira-Halder (Orientadora) [UNICAMP]

INTRODUÇÃO:

O melanoma causa aproximadamente 55.500 mortes anuais em todo o mundo, sendo responsável por 0,7% de todas as mortes por câncer, que em última análise resultam da metástase do melanoma (Eggermont et al, 2014). As assinaturas de expressão gênica têm o potencial de melhorar a predição do comportamento biológico do melanoma, ao permitir defini-lo/classificá-lo como "alto risco", sob o ponto de vista molecular (Lopes et al, 2020).

No presente estudo utilizou-se dados disponíveis em repositório público com estudos sobre melanoma e trametinib. Os dados disponíveis são referentes a RNAseq que consiste em informações de 26.334 transcritos para cada um dos 95 pacientes, e apresenta as informações de cada paciente: qual tipo de tratamento foi feito, qual o momento da coleta (pré- e pós-tratamento), se a amostra foi coletada de sítio primário ou metastático, qual o sítio metastático (secundário) e presença ou ausência de mutação Braf. Assim, este trabalho traz uma análise exploratória das informações disponíveis com objetivo de construir um mapa com indicação das vias de onco-sinalização relevantes para auxiliar em estratégias de monitoramento e escolha terapêutica.

METODOLOGIA:

Busca de dados de amostras de pacientes com melanoma obtidos através de análises em larga escala disponíveis em repositório público

A análise em GEOdataset foi feita a partir da busca por estudos sobre melanoma e trametinib. Foi encontrado o estudo de Ribas colaboradores(2020) que envolveu a análise global da expressão de RNA, cujos dados estão disponíveis em Series GSE158403. Os dados disponíveis são referentes ao RNAseq e informações de cada paciente, como tipo de tratamento realizado dia da coleta (pré- e pós-tratamento), sítio da coleta (primário ou metastático), local do sítio metastático (secundário) e presença ou ausência de mutação *Braf*. É importante mencionar que o método de coleta realizado foi diferente entre os pacientes. Um mesmo paciente teve avaliação de diferentes sítios secundários (metastático) no pré-tratamento e pós-tratamento. Essa foi uma dificuldade encontrada no momento de avaliação do estudo, a ausência de pareamento das fontes das amostras em alguns pacientes. Os dados relativos a este estudo em questão consistem em dois arquivos (GSE158403_series_matrix e GSE158403_FPKM_matrix). O banco de dados GSE 158403_FPKM_matrix consiste em informações de 26.334 transcritos, para cada um dos 95 pacientes. Enquanto que o arquivo GSE_158403_series_matrix contém informações

mais específicas dos pacientes, como órgão da coleta, presença de mutação *Braf* e regime de tratamento do estudo. O fluxo do delineamento experimental deste estudo está representado no *pipeline* conforme Figura 1.

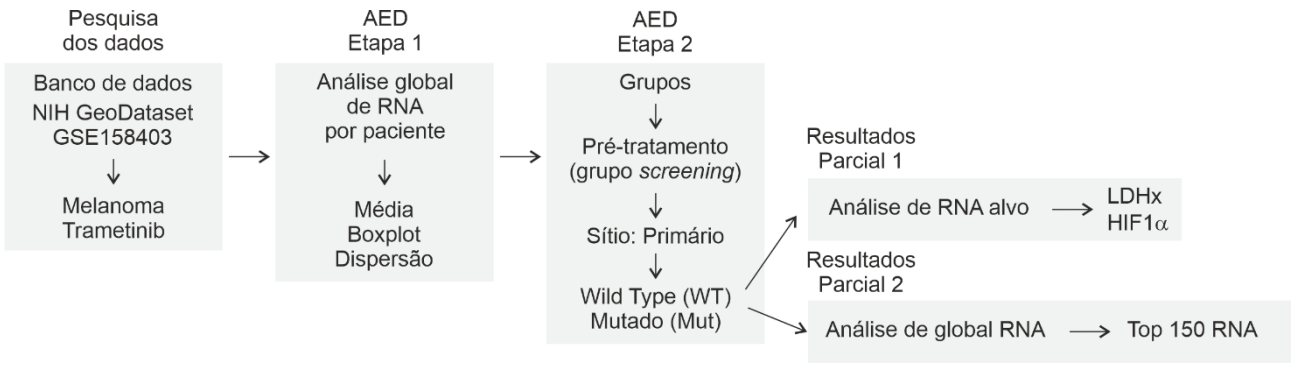

Figura 1. *Pipeline* do desenvolvimento do projeto. Usando o banco de dados NIH GeoDataset, foi feita a pesquisa de um banco de dados de RNAseq, com amostras de paciente com melanoma em regime de tratamento com trametinib, inicialmente a droga escolhida para o projeto. Na primeira etapa da Análise Exploratória de Dados (AED Etapa 1) foi feita uma análise global de RNA por paciente, demonstrado por gráficos de média de expressão, boxplot e dispersão. Na etapa 2 de AED, os pacientes foram divididos em grupos, sendo escolhido para análise pacientes do grupo pré-tratamento, com tumor em sítio primário para comparação entre pacientes com presença e ausência de mutação em Braf. Na primeira parte do projeto, os resultados parcial 1 mostraram a diferença de expressão dos transcritos alvo desse projeto: família LDHx e HIF1a. Na segunda parte do projeto, os resultados parcial 2 foram baseados na busca de uma assinatura gênica desses grupos a partir da análise dos 150 transcritos mais expressos.

Grupos do estudo

Os grupos definidos para investigação neste trabalho foram definidos Braf WT - pacientes que apresentam *Braf wild type* e *Braf* Mut - pacientes que apresentam mutação em *Braf*; mutações em *Braf* acometem 50-70% dos pacientes com melanoma, conforme detalhado na Figura 2.

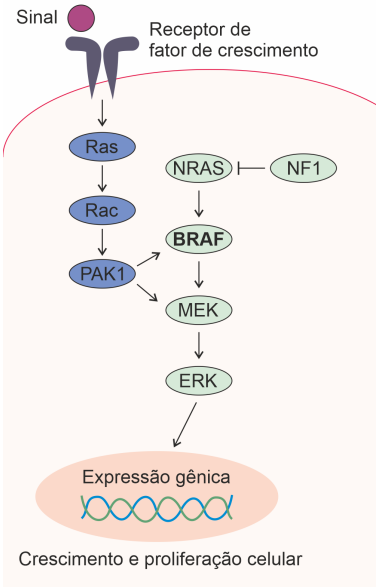

Figura 2. Via de sinalização de *Braf*. Mutações no gene *Braf* acometem de 50-70% dos melanomas, e geralmente ocorre a substituição de valina por ácido glutâmico na posição 600. Essa mutação desencadeia ativação constitutiva da via de sinalização MAPK/ERK, que promove a progressão do tumor pelo aumento da proliferação, sobrevivência, motilidade e capacidade de invasão (Lopes et al, 2020). Figura adaptada de Lopes et al, 2020.

RESULTADOS E DISCUSSÃO:

Uma vez que o estudo apresenta grande quantidade de dados, foi necessário inicialmente fazer uma análise descritiva dos dados, para um melhor direcionamento. Dessa forma, constatou-se que dos 26.334 transcritos, sua grande maioria apresentou valores bem semelhantes. No entanto, percebeu-se também que alguns transcritos apresentaram valores extremamente altos, comparados com os demais. Assim, procurou-se fazer um ranqueamento dos transcritos, a fim de avaliar a sua distribuição nos pacientes, bem como a sua proporção. Constatou-se que o valor da expressão gênica era muito parecido na grande maioria, e optou-se em utilizar os 150 transcritos mais expressos para o refinamento da análise.

Conforme descrito anteriormente, alguns pacientes apresentaram diferenças de transcritos muito elevadas, e portanto, optamos em separá-los em grupos. De acordo com a investigação

anterior do perfil global e individual dos pacientes, foi feito o agrupamento dos pacientes de acordo com os perfis: *Braf* WT e *Braf* Mut.

Análise dos 150 transcritos mais abundantes: estratégia de mapeamento de assinatura gênica

A partir das diferenças encontradas nos transcritos, tanto pela análise global, quanto pela análise dos transcritos das proteínas-alvo do projeto, foi feita uma análise dos 150 transcritos mais abundantes nos dois grupos de pacientes, para identificar um perfil de transcritos nos grupos. Dessa forma, os pacientes foram separados em dois grupos: *Braf* Mut que contém três pacientes e *Braf* WT que contém dois pacientes. Diante dessa separação, foram realizadas análises mais específicas, conforme a Figura 2.

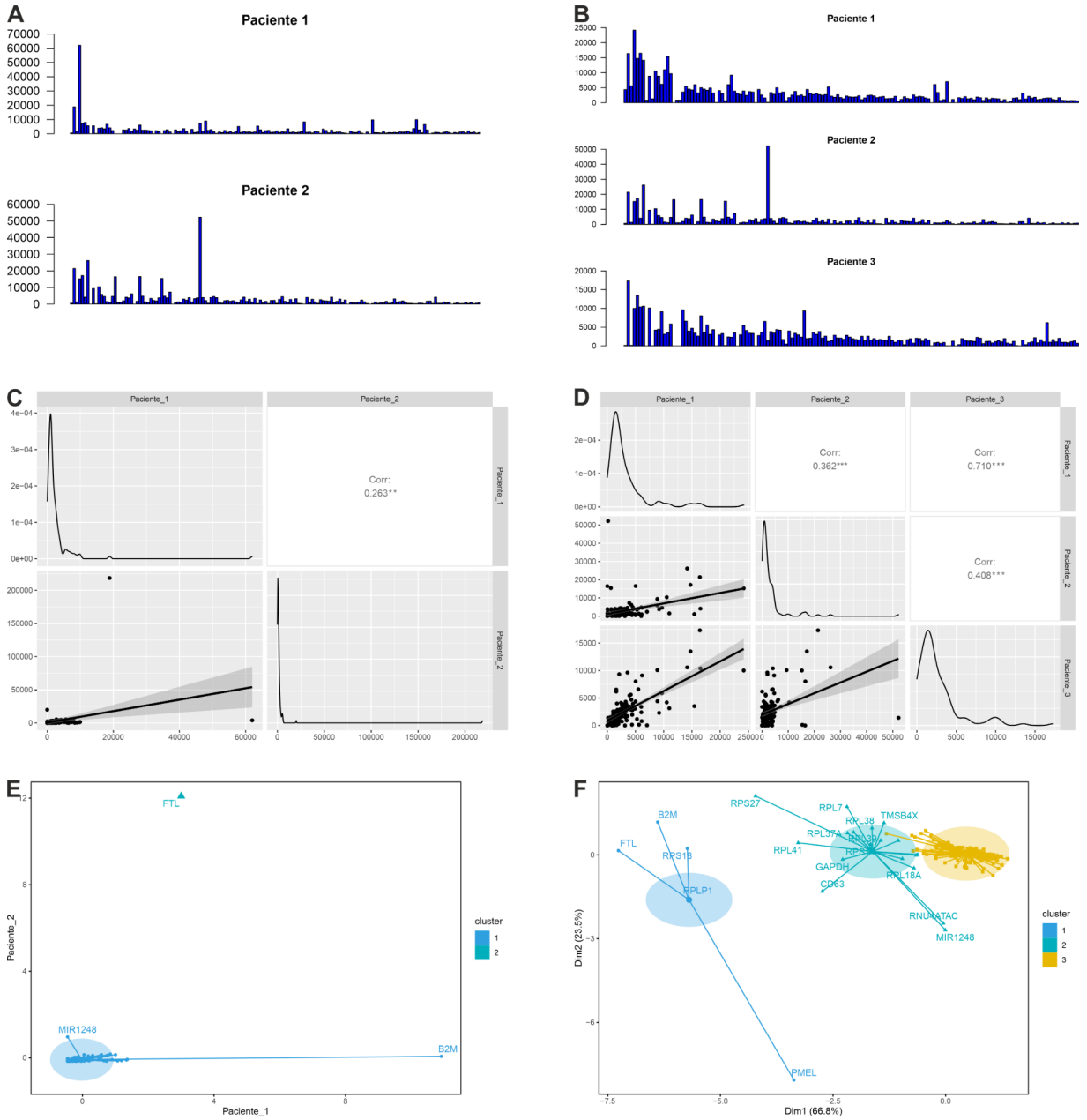

Figura 2. Representação gráfica da comparação entre os pacientes com tumor primário pré-tratamento do grupo *Braf* WT em e *Braf* Mut. (A-B) Gráficos **A** e **B** apresentam a classificação das expressões gênicas para pacientes *Braf* WT (A) e *Braf* Mut (B). (C-D) Gráficos **C** e **D** exibem a correlação feita entre os pacientes em cada grupo *Braf* WT (C) e *Braf* Mut (D). (E-F) Os gráficos **E** e **F** mostram um estudo de *cluster* para explorar a existência de similaridades dos genes nos pacientes *Braf* WT (E) e *Braf* mutado (F).

Ao avaliar o perfil dos 150 transcritos mais abundantes nas amostras, podemos observar que os pacientes apresentam diferenças no padrão de expressão dentro do mesmo grupo. No grupo *Braf* Mut (Figura 2B), o padrão de distribuição do gráfico de barras é mais semelhante entre os pacientes deste grupo do que quando compara-se os pacientes dentro do grupo *Braf* WT (Figura 2A). De fato, ao observar o correlograma dos grupos, o R do grupo *Braf* WT é 0.263 (Figura 2C), enquanto o R no grupo *Braf* Mut Paciente 1 e Paciente 3 apresentam R = 0.710, que confirma maior correlação entre os pacientes (Figura 2D). De forma interessante, ao avaliar o perfil de formação de cluster entre os grupos, o grupo *Braf* WT apresenta dois clusters (Figura 2E), sendo um deles formado unicamente por *FTL* e o outro com os demais transcritos, o que sugere que a menor correlação entre o perfil desses pacientes pode estar relacionada ao transcrito *FTL*. Em relação ao cluster no grupo *Braf* Mut, observa-se três agrupamentos, sendo um deles também com presença do transcrito *FTL*.

Sendo assim, a elucidar potenciais transcritos que comporiam a assinatura gênica desses pacientes, foi feito o *ranking top* 10 dos transcritos mais abundantes nos pacientes (Tabela 1). Os resultados obtidos reforçam a hipótese de que os pacientes com *Braf* Mut apresentam um perfil mais homogêneo em relação à expressão gênica comparado aos pacientes *Braf* WT. E de fato, a assinatura gênica em pacientes *Braf* Mut poderia ser uma estratégia de validação de perfil dos pacientes, bem como abordagem de decisão terapêutica.

De forma mais específica, os transcritos de *LDH* estão presentes no *ranking* dos 150 transcritos mais expressos. Nos pacientes Braf WT, o *LDHB* ocupa a posição 41 no *ranking*. E nos pacientes Braf Mut, *LDHB* ocupa a posição 35, e *LDHA* ocupa a posição 102. Ou seja, a partir da correlação entre os transcritos da família *LDHx* e mutação em *Braf* poderia fornecer informações para tomada de decisão terapêutica, uma vez que mutações em *Braf* somada a potencial switch metabólico podem contribuir para a resistência de tumores a quimioterápicos (Hanahan & Weinberg, 2011; Martinez-Cardús et al, 2015). Além disso, ao analisar o interactoma das proteínas a partir dos transcritos mais abundantes nesses grupos, observa-se que esses transcritos também correlacionam biologicamente, confirmando a relevância do uso dessa ferramenta para identificar a assinatura gênica em pacientes com uso de plataforma de sequenciamento de RNA (Figura 3).

Tabela 1. *Ranking* dos 10 transcritos mais abundantes entre pacientes *Braf* WT e *Braf* Mut

Braf WT				Braf Mut					
Gene_id	*Braf*WT.1	Gene_id	*Braf*WT.2	Gene_id	*Braf*Mut.1	Gene_id	*Braf*Mut.2	Gene_id	*Braf*Mut.3
B2M	62006.19	*FTL*	218557.37	*B2M*	24172.93	*PMEL*	52201.87	*FTL*	17326.48
FTL	18852.66	*DONSON*	33336.76	*FTL*	16395.21	*FTL*	21407.31	*B2M*	10015.55
IFI27	9908.12	*MIR6723*	22210.99	*TMSB4X*	10995.36	*CD63*	16536.08	*S100A6*	9330.99
RNASE1	9764.21	*MIR1248*	20028.31	*GAPDH*	10534.05	*MIR1248*	16462.52	*CD63*	8030.90
IFITM3	8944.76	*MALAT1*	5384.06	*SPP1*	7037.15	*RNU4ATAC*	15406.21	*LGALS1*	6535.54
IFITM1	8310.21	*TMSB4X*	4938.82	*APOC1*	6096.53	*B2M*	15222.96	*S100A4*	6179.86
PMEL	7384.24	*B2M*	4031.26	*HBB*	5612.34	*SNORA66*	12738.16	*GAPDH*	4206.08
TMSB4X	6611.66	*SCARNA22*	3882.91	*CD63*	5014.95	*GAPDH*	10381.85	*TMSB10*	4140.72
MGP	6444.16	*LGALS1*	3288.72	*FCER1G*	4697.67	*SNORA25*	5822.22	*IFITM3*	3409.28
CD63	6040.20	*GAPDH*	2886.87	*HBA2*	4370.42	*SNORD67*	5626.69	*EEF1A1*	3383.96

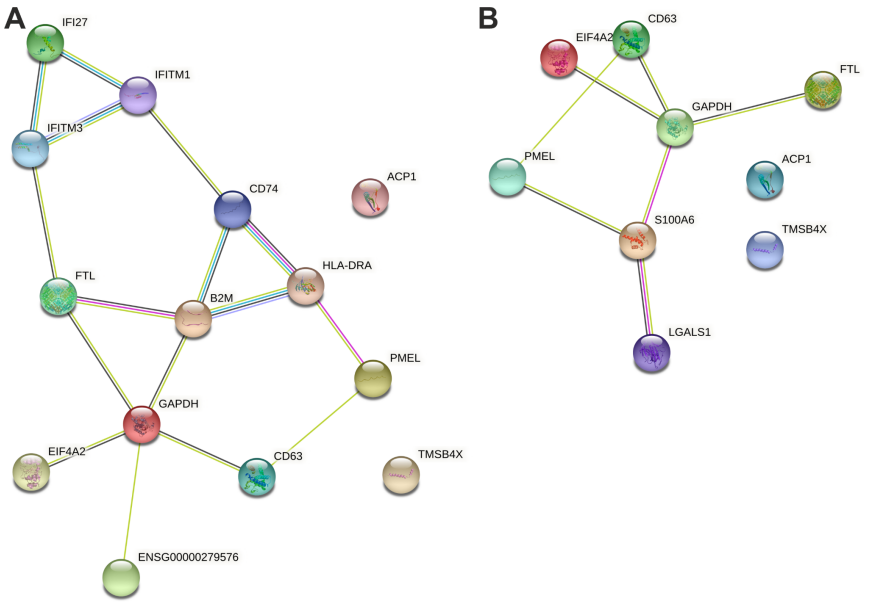

Figura 3. Interactoma dos genes a partir do ranking dos 150 transcritos mais abundantes comparando as amostras de paciente *Braf* WT (A) e *Braf* Mut (B). *Braf* mut: FTL, PMEL, CD63, GAPDH, MIR1248, TMSB4X, S100A6, LGALS1. *Braf* WT: FTL, B2M, MIR1248, TMSB4X, IFI27, IFITM3, PMEL, IFITM1, MALAT1, HLA-DRA, GAPDH, CD74, CD63. Legenda – cores interactoma – interações conhecidas: de databases – ciano, determinadas experimentalmente – magenta; Interações preditas – genes vizinhos – verde, fusão – vermelho, co-ocorrência – azul; Outros – text mining – amarelo, co-expressão – preto, proteína homóloga – violeta.

CONCLUSÕES:

Levando-se em consideração os dados disponíveis inicialmente, foi necessário fazer uma análise exploratória, para estabelecer as estratégias/ferramentas para análises subsequentes. Dessa forma, foi possível classificar/elencar os genes com base no nível de expressão gênica. Conclui-se que existem diferenças entre as expressões gênicas, as quais ficam mais evidentes, quando os pacientes foram divididos em subgrupos. O nível de transcritos presente nos pacientes *Braf* WT é muito superior comparado aos pacientes *Braf* Mut. Além disso, a correlação entre os pacientes identificados na análise exploratória apresenta relevância biológica conforme apresentado pelo interactoma de proteínas.

BIBLIOGRAFIA

Eggermont AM, et al. Cutaneous melanoma. Lancet. 2014;383(9919):816-27.

Martinez-Cardús A, et al. Epigenetic mechanisms involved in melanoma pathogenesis and chemoresistance. Ann Transl Med. 2015;3(15):209.

Hanahan D, Weinberg RA. Hallmarks of cancer: the next generation. Cell. 2011;144(5):646-74.

Lopes BA, et al. Melanoma cell lines: Mutations and impact on signal transduction pathways. Braz J Hea Rev. 2020;3(4):10859-10884.

Ribas A, et al. PD-L1 blockade in combination with inhibition of MAPK oncogenic signaling in patients with advanced melanoma. Nat Commun 2020;11(1):6262.