

UNIVERSIDADE FEDERAL DO RIO GRANDE DO SUL
INSTITUTO DE BIOCÊNCIAS

MATHEUS GIBEKE SIQUEIRA DALMOLIN

Perfil de expressão gênica associado ao prognóstico em
Sarcoma de Ewing

Porto Alegre

2019

MATHEUS GIBEKE SIQUEIRA DALMOLIN

Perfil de expressão gênica associado ao prognóstico em
Sarcoma de Ewing

Trabalho de conclusão de curso de graduação apresentado ao Instituto de Biociências da Universidade Federal do Rio Grande do Sul, como requisito parcial para a obtenção do título de Bacharel em Ciências Biológicas.

Orientadora: Dra. Rita Maria Cunha de Almeida.

Co-orientadora: Dra. Marialva Sinigaglia

Porto Alegre

2019

Dedico esse trabalho à todas as
pessoas que de alguma forma
contribuem para a existência do
Instituto do Câncer Infantil.

AGRADECIMENTOS

Agradeço aos meus pais, Mauro e Tatiana, por proverem todas as condições para que eu pudesse estudar, além de todo amor e carinho. Agradeço também a minha irmã, Alice, por todo amor. Se não fossem vocês eu não teria chegado até aqui.

Agradeço à todos meus amigos que fizeram os momentos difíceis serem os mais divertidos.

Agradeço ao meu tio, Rodrigo, por me apresentar a bioinformática e me dar inúmeros valiosos conselhos.

Agradeço às minhas orientadoras, Marialva e Rita, por toda a dedicação e paciência durante todo esse período. Todas as conversas e reuniões de quinta-feira, foram fundamentais para o meu crescimento como pesquisador. Obrigado pela confiança!

Agradeço, em especial, aos meus colegas de laboratório, Maurício e Ricardo, por terem me auxiliado em cada etapa deste trabalho, tanto na parte teórica como prática. Sem vocês o trabalho não teria nem metade do brilho. Aprendi muito com vocês, obrigado por tudo.

“Em algum lugar, algo incrível
está esperando para ser descoberto”

Carl Sagan

"Este trabalho teve apoio do Instituto do Câncer Infantil do Rio Grande do Sul
(PRONON SIPAR 25000.202.751/2016-65), do Conselho Nacional de Desenvolvimento
Científico e Tecnológico (CNPq) e da Coordenação de Aperfeiçoamento de Pessoal de Nível
Superior (CAPES) "

RESUMO

O sarcoma de Ewing (SE) acomete ossos e tecidos moles e é o segundo tumor ósseo mais frequente em crianças e adolescentes. No tumor de Ewing localizado a taxa de sobrevida é de 70%, entretanto o tumor metastático a taxa de sobrevida é de 30%. Pacientes com recidiva normalmente não respondem ao tratamento. Assim a detecção de genes prognósticos de desfecho do SE no momento do diagnóstico poderia orientar para um protocolo de tratamento mais eficaz e individualizado de acordo com o perfil de agressividade do tumor. Para o presente estudo, foram selecionados dados de expressão gênica de pacientes com SE de três coortes (COG, EuroEwing e Finlandeses) no Gene Expression Omnibus (GEO), com o intuito de selecionar vias e/ou genes marcadores de prognóstico em SE. Para cada uma das coortes, os pacientes foram divididos em dois grupos: sobreviventes, com sobrevida maior do que cinco anos (SOB) e não sobreviventes, com sobrevida menor do que cinco anos (NSOB). As análises da expressão diferencial entre os grupos foram feitas através do *software* Transcriptograder. O conjunto de dados da coorte do COG foi utilizada como referência, onde foram selecionados grupos de genes diferencialmente expressos (GGDE). O GGDE foi utilizado para realizar o enriquecimento funcional. As vias foram analisadas através do *software* Transcriptograder, sendo que 38 apresentaram diferença de expressão significativa ($P < 0,05$) quando comparado SOB vs NSOB. Dentre estas, destacam-se as vias de sinalização de efrina, ERBB, proliferação celular, PI3K-AKT, MAPK, fagocitose entre outras que estavam superexpressas em NSOB. Além disso, foram selecionados os dez genes mais significativos ($p < 0,05$) de cada uma das vias que foram utilizados para avaliar a sobrevida global através dos gráficos de Kaplan-Meier. Foram utilizadas duas coortes independentes (EuroEwing e Finlandeses) para validar os resultados. Como resultado principal foram encontrados 23 genes associados com sobrevida global (Kaplan-Meier significativo, $p < 0,05$) validados nas duas coortes independentes. Dos genes cuja expressão foi associada ao bom prognóstico encontram-se o C5 que faz parte da cascata do sistema complemento e já havia sido associado a bom prognóstico em SE em uma das coortes. Outros genes cuja superexpressão foi associada a bom prognóstico são: *ULK2*, *BECN1* e *ATG4* e estão relacionados a autofagia. Dos genes cuja a superexpressão foi associada ao mau prognóstico destaca-se o *PLAUR* que tem sido associado com invasão migração e metástase em outros tumores. A nossa perspectiva é validar esses resultados em análises de bancada com intuito de contribuir para um tratamento mais personalizado e eficaz do SE.

Palavras-chave:

câncer infantil; sarcoma de ewing; transcriptograder; expressão diferencial; prognóstico;

LISTA DE FIGURAS

- Figura 1.** Mapa do enriquecimento da lista ordenada de 9684 genes por termos do Gene Ontology e por rotas metabólicas do KEGG.....23
- Figura 2.** Transcriptograma relativo comparando as amostras de SE de duas coortes de pacientes EuroEwing e COG.....27
- Figura 3.** Transcriptograma relativo entre não sobreviventes (NSOB) e sobreviventes (SOB), realizado a partir dos dados de expressão gênica da coorte do COG das amostras com baixo conteúdo tumoral (<70%)..... 29
- Figura 4.** Transcriptograma relativo entre não sobreviventes (NSOB) e sobreviventes (SOB), realizado a partir dos dados de expressão gênica da coorte do COG das amostras com alto conteúdo tumoral (> 70%)..... 30
- Figura 5.** Transcriptograma relativo entre não sobreviventes (NSOB) e sobreviventes (SOB), realizado a partir dos dados de expressão gênica de SE da coorte do EuroEwing..... 31
- Figura 6.** Transcriptograma relativo entre não sobreviventes (NSOB) e sobreviventes (SOB), realizado a partir dos dados de expressão gênica de SE da coorte do COG..... 31
- Figura 7.** Transcriptograma relativo entre não sobreviventes (NSOB) e sobreviventes (SOB), realizado a partir dos dados de expressão gênica de SE da coorte dos Finlandeses..... 32
- Figura 8.** Transcriptograma relativo entre não sobreviventes (NSOB) e sobreviventes (SOB), realizado a partir dos dados de expressão gênica da coorte de SE do COG..... 33
- Figura 9.** Curvas de Kaplan-Meier para os 23 genes estimadas com dados de sobrevivência e expressão gênica obtidos em cada uma das três coortes (COG, EuroEwing e Finlandeses), totalizando 69 diagramas..... 43

LISTA DE TABELAS

Tabela 1. Relação de amostras (ID), desfecho (SOB = sobrevivente; NSOB = não sobrevivente) e Sobrevivência global em meses (SG) para as amostras das três coortes (COG, EuroEwing e Filandeses).....	26
Tabela 2. Relação de amostras (ID) e conteúdo tumoral da coorte COG.....	28
Tabela 3. Termos do GO, vias do Kegg e Reactome que apresentaram expressão diferencial significativa quando comparado SOB vs NSOB, amostras de SE (COG).....	34
Tabela 4. Lista de genes mais significativos da coorte COG.....	38
Tabela 5. Relações entre nível de expressão e bom prognóstico dos 23 genes com Kaplan-Meier significativo nas três coortes.....	42

LISTA DE ABREVIATURAS

COG	Children's Oncology Group
EuroEwing	Euro-Ewing99-R1
GEO	Gene Expression Omnibus
GO	Gene Ontology
GGDE	grupo de genes diferentemente expressos
HDGF	Fator de crescimento derivado de hepatoma
HGC	hibridação genômica comparativa
INCA	Instituto Nacional de Câncer
JAK1	Janus Kinase 1
NSOB	Não sobreviventes, com sobrevida menor do que cinco anos
OMS	Organização Mundial da Saúde
R	Raio
RCBP	Registros de Câncer de Base Populacional
RE	Reticulo Endoplasmático
(RMS)	Rabdomiossarcoma
RNA	Ácido ribonucleico
RNA-Seq	Sequenciamento de RNA
RT-PCR	Reação em cadeia da polimerase da transcrição reversa
SE	Sarcoma de Ewing
SG	sobrevida global
SLE	Sobrevida livre de eventos
SOB	Sobreviventes, com sobrevida maior do que cinco anos
TCGA	<i>The Cancer Genome Atlas</i>

SUMÁRIO

1 INTRODUÇÃO.....	13
1.1 Câncer infantil.....	13
1.2 Sarcoma de Ewing.....	13
1.3 Justificativa.....	19
2. OBJETIVO.....	20
2.1 Objetivo Geral.....	20
2.2 Objetivos Específicos.....	20
3 MATERIAL E MÉTODOS.....	21
3.1 Os dados.....	21
3.2 Classificação das amostras.....	21
3.3 Normalização.....	21
3.4 Transcriptogramas.....	21
3.5 Identificação de vias candidatas diferencialmente expressas.....	23
3.6 Análise por diagrama de Kaplan-Meier para genes diferencialmente expressos.....	24
4 RESULTADOS E DISCUSSÃO.....	25
4.1 Organização e escolha dos dados.....	25
4.2 As coortes COG e EuroEwing não podem ser analisadas em conjunto, pois não são homogêneas.....	27
4.3 As amostras dos pacientes do COG possuem quantidades consideráveis de estroma.....	28
4.4 Transcriptograma realizado apenas com amostras contendo baixo conteúdo tumoral não apresenta diferenças evidentes entre os grupos SOB e NSOB.....	29
4.5 Transcriptograma da coorte COG é capaz de detectar diferenças de expressão gênica entre grupos prognósticos.....	30
4.6 Seleção das vias e genes mais significativos.....	32
4.7 Vinte e três genes tiveram Kaplan-Meier significativo em todas as três coortes.....	38
5 PERSPECTIVAS.....	49
6 BIBLIOGRAFIA.....	50

1 INTRODUÇÃO

1.1 Câncer infantil

O câncer em crianças e adolescentes (entre 0 e 19 anos) se refere a um grande conjunto de doenças que apresentam características próprias, não somente de natureza histológicas, mas também relacionadas ao comportamento clínico (Little, 1999). O câncer infantil é raro quando comparado ao câncer de adulto, corresponde entre 1% e 4% de todos os tumores malignos nas diferentes populações humanas. A proporção dos tipos de câncer infantil variam nas diferentes populações. A incidência é de 1% nos países desenvolvidos e de 3 a até 10% nos países em desenvolvimento, onde a população infantil pode chegar a 50% do total. No Brasil, segundo os Registros de Câncer de Base Populacional (RCBP) utilizados no relatório “Incidência, Mortalidade e Morbidade hospitalar por câncer em crianças, adolescentes e adultos jovens no Brasil: Informações dos registros de câncer e do sistema de mortalidade” de 2016, estima-se que uma porcentagem mediana de 3% do total de casos de câncer ocorram em crianças e adolescentes (0-19 anos) (Instituto Nacional de Câncer José Alencar Gomes da Silva, 2016). Ainda de acordo com o Instituto Nacional de Câncer (INCA), a sobrevida estimada para todas as neoplasias em crianças e adolescentes (de 0 a 19 anos) foi em torno de 64%. A sobrevida não é homogênea ao longo das regiões do Brasil, variando de 50% na Região Norte, 60% na Região Nordeste, 65% na Região Centro-Oeste, 70% na Região Sudeste, e 75% na Região Sul. Em outros países o cenário é semelhante. Nos Estados Unidos em 2018, estimou-se 10.590 novos casos de câncer pediátrico (0-14 anos) diagnosticados em 2018, correspondendo a 0,61% do total (1.735.350). O câncer infantil é subdividido em pelo menos 12 grandes grupos cada qual com seus subgrupos. Os tumores malignos cerebrais e do sistema nervoso somados a leucemia correspondem a 55% de todos os casos de câncer em crianças e adolescentes (0-14 anos), tornando outros tipos de cânceres pediátricos ainda mais raros (American Cancer Society, 2018).

1.2 Sarcoma de Ewing

O Sarcoma de Ewing (SE) é um tumor que acomete tanto ossos como partes moles, afeta principalmente crianças, adolescentes e jovens adultos com aproximadamente 1,5 casos por milhão globalmente. Entre 20 e 25% dos pacientes apresentam metástases ao diagnóstico,

quase sempre resistentes à terapia intensiva (GASPAR, *et al*, 2015). O padrão de tratamento para o sarcoma de Ewing consiste em um regime de tratamento multimodal incluindo ressecção cirúrgica e/ou radioterapia local, bem como quimioterapia multiagente intensa (PAPPO & DIRKSEN, 2018). Apesar de comprovada a eficácia para o tratamento da doença localizada, a sobrevida a longo prazo de pacientes com SE metastático ou que apresentam recaída ainda é inaceitavelmente baixa (GASPAR, *et al*, 2015; STAHL, *et al*, 2011).

A classificação de sarcomas da OMS de 2013 definiu que diversos tumores, como por exemplo o Askin, que antes eram considerados como pertencentes a “família de Sarcoma de Ewing”, agora são considerados "sarcoma de Ewing" (DOYLE, 2013). O que caracteriza o sarcoma de Ewing são as fusões gênicas FET-ETS (WATSON, *et al*, 2018; SANKAR, *et al*, 2011)

A primeira descrição do sarcoma de Ewing foi em 1921 (EWING, 1921), e desde então a sua célula de origem ainda não foi confirmada e é motivo de debate (TOOMEY, 2010). Nas últimas décadas, os estudos convergiram para duas células putativas de origem: células-tronco mesenquimais humanas e células da crista neural humana (GRÜNEWALD, *et al*, 2018). Há também estudos que apontam para um fenótipo de células-tronco neuro-mesenquimais (VON LEVETZOW, *et al*, 2011).

Mesmo com essa indefinição histológica, o sarcoma de Ewing é bem caracterizado geneticamente. Em comparação com outros tipos de cânceres sólidos, o número de mutações somáticas recorrentes no SE é limitado (BROHL, *et al*, 2014; CROMPTON, *et al*, 2014; TIRODE, *et al*, 2014; SOLOMON, *et al*, 2011). As suas principais marcas são translocações cromossômicas específicas entre o cromossomo 11 e 21, que fundem um membro da família FET (FUS, EWSR1 e TAF15) com membros da família ETS (FLI1 em 85% dos casos) (DELATTRE, 1992). EWSR1-FLI1 foi relatado como a fusão mais frequente no Sarcoma de Ewing, a qual codifica uma proteína quimérica que funciona como um fator de transcrição aberrante (GRÜNEWALD, *et al*, 2018). A proteína quimérica EWSR1-FLI1 é capaz de alterar o estado da cromatina, desregulando a transcrição (PATEL, *et al*, 2012; RIGGI, *et al*, 2014). Algumas das anormalidades transcricionais no SE têm sido associadas a modificações epigenéticas induzidas pelo EWSR1-FLI1 (GRÜNEWALD, *et al*, 2018).

Em um estudo realizado com 975 pacientes com SE, Cotterill e colaboradores evidenciaram que a maior parte dos pacientes no momento do diagnóstico possui idades entre 10 e 20 anos (mediana de 14 anos, com idades de 8 meses a 47 anos) (COTTERILL, *et al*, 2000). A incidência em homens é maior do que em mulheres (56,9% contra 43,1%) e foi ainda maior em pacientes com 15 anos ou mais, quando comparado com pacientes com idade inferior a 15 anos (62% contra 52%), assim como volume do tumor (mediana, 190 mL contra 112 mL). O único cenário onde a proporção de mulheres com a doença era maior que em homens foi na faixa etária de 0 a 9 anos (53,1%) atingindo a proporção mínima em pacientes com mais de 20 anos de idade (30,2%). Esses dados epidemiológicos estão próximos dos encontrados por outros autores (ESIASHVILI, *et al*, 2008).

Diversas pesquisas estão sendo realizadas para avaliar se o perfil de expressão gênica pode ser usado para estratificar pacientes com câncer no momento do diagnóstico. Primeiro demonstrado como factível no câncer de mama (CARLSON, *et al*, 2013), essa abordagem prognóstica foi avaliada e validada em outros cânceres humanos (ERHO, *et al*, 2013; TRINQUAND, *et al*, 2013), incluindo neoplasias pediátricas, como neuroblastoma (ASGHARZADEH, *et al*, 2006; OBERTHUER, *et al*, 2010; STRICKER, *et al*, 2014), rabdomyosarcoma (DAVICIONI, *et al*, 2010; WILSON, *et al*, 2013; MISSIAGLIA, *et al*, 2012) e leucemia (CLEAVER, *et al*, 2010; KANG, *et al*, 2009).

Vários pequenos estudos de perfis de genoma de Sarcoma de Ewing foram relatados, e biomarcadores candidatos de prognóstico não sobrepostos foram identificados (BENNANI-BAITI, *et al*, 2010; OHALI A, *et al*, 2004; SCHAEFER, *et al*, 2008; SCOTLANDI, *et al*, 2009). No entanto, nenhuma destas assinaturas de genes prognósticos candidatos foi validada prospectivamente em coortes independentes de pacientes tratados de maneira equivalente.

Savola e colaboradores em 2009, sabendo da importância de que, não somente a fusão gênica EWS/FLI1, mas como também outras diversas alterações no número de cópias cromossômicas são importantes para o desenvolvimento neoplásico do SE e para o desfecho do paciente, realizaram um estudo que avaliou com alta resolução o genoma inteiro e procurou identificar regiões genômicas de maior interesse e possíveis genes alvos nessas áreas. Os pesquisadores realizaram matrizes de hibridação genômica comparativa (HGC) que foram usadas para rastrear alterações no número de cópias em 31 amostras de SE, e também

tentaram detectar as diferenças da expressão gênica entre 16 amostras de pacientes com SE. Como resultado os autores descobriram que em 87% dos casos analisados, o número de cópias estava alterado. As mudanças de número de cópias mais recorrentes foram ganhos em 1q, 2, 8 e 12, e perdas em 9p e 16q. Sobrevida livre de eventos (SLE) e sobrevida global (SG) foram significativamente melhores ($P < 0,05$) para tumores primários que continham três alterações ou menos no número de cópias do que para tumores com maior número de cópias aberrantes. No total, 20 novos candidatos a oncogenes e a supressores tumorais associados à SE foram encontrados na análise de integração de dados de HGC e de expressão gênica. Os autores identificaram o gene HDGF (Fator de crescimento derivado de hepatoma) como sendo o mais importante, portanto RT-PCR foi realizado a fim de estudar os níveis de expressão. O gene HDGF apresentou maior expressão no SE (42 amostras avaliadas) quando comparado com tecidos normais (músculo), mas a correlação entre o nível de expressão de HDGF e desfecho não foi detectada. Portanto os autores concluíram que o gene HDGF pode estar envolvido com a progressão tumoral, mas não apresenta valor prognóstico (SAVOLA, *et al*, 2009).

Sabendo que, ao contrário do câncer adulto, o câncer infantil não é tipicamente desencadeado por eventos inflamatórios (DICKENS, *et al*, 2003), Savola e colaboradores em 2011, estudaram se os genes relacionados ao sistema imune são ativados e se têm significado prognóstico na família de tumores de SE. Para isso os autores utilizaram perfis de expressão gênica de pacientes com SE, linhagens de SE e amostras de músculo esquelético. Além disso eles também utilizaram uma lista contendo 238 genes supostamente relacionados à inflamação e uma lista de 299 genes expressos por macrófagos. Compararam o perfil de expressão gênica das amostras de SE com as linhagens celulares de SE com o objetivo de identificar genes com função inflamatória derivados do componente imune associado a doença. Também foi realizada uma comparação entre o perfil de expressão gênica das amostras de SE com os tecidos musculares normais para identificar genes inflamatórios que são funcionais no SE, mas não em tecidos adjacentes. Os pesquisadores encontraram que os genes inflamatórios são ativados em amostras de pacientes com SE, já que 38 de 238 (16%) genes inflamatórios foram supra regulados quando comparado com linhagens celulares. Esta ativação dos genes inflamatórios foi caracterizada pelo significativo enriquecimento da expressão gênica relacionada à macrófagos com 58 de 299 (19%) dos genes regulados. O

gene C5 (componente do complemento 5) foi quem melhor correlacionou com sobrevida livre de eventos e sobrevida global de maneira dose-dependente. Os autores destacaram ainda JAK1 e IL8 como sendo do sistema imunológico e correlacionados ao prognóstico de amostras com SE (SAVOLA, *et al*, 2011).

Outro estudo importante foi realizado por Volchenboum e colaboradores em 2015, que tiveram como objetivo desenvolver biomarcadores que possam prever quais pacientes estão em risco de recaída, para que a terapia e a pós-terapia possam ser ajustadas em função do grupo de risco. Para isso, eles realizaram o perfil de expressão inteiro de duas coortes independentes de tumores de SE que eram clinicamente anotados em um esforço para identificar e validar assinaturas gênicas de prognóstico. As amostras de SE foram obtidos do *Children's Oncology Group* (COG) e do Euro-Ewing99-R1 (EuroEwing) e o perfil de expressão do genoma completo de cada amostra foi feito usando matrizes Affymetrix Human Exon 1.0 ST. Os pesquisadores geraram listas de genes diferencialmente expressos entre sobreviventes e não sobreviventes. Estas listas serviram para identificar assinaturas gênicas prognósticas. Essa primeira parte das análises foi realizada com os dados do COG, então o próximo passo foi validar essas assinaturas gênicas prognósticas em uma coorte independente, o EuroEwing. A primeira abordagem utilizada pelos autores para descobrir os grupos de genes diferencialmente expressos entre sobreviventes e não sobreviventes (dados do COG), foi uma análise não supervisionada que não conseguiu segregar os tumores com base no desfecho. Já quando adotaram uma abordagem supervisionada, encontraram uma pequena quantidade de genes diferencialmente expressos entre sobreviventes e não sobreviventes, os quais foram utilizados para identificar assinatura de genes prognósticos. A análise de enriquecimento específica dos genes demonstrou que os genes da integrina e da quimiocina estavam associados à sobrevivência em tumores onde o estroma compunha mais de 30% da amostra. Além disso, os autores analisaram de forma similar uma segunda coorte de SE do grupo cooperativo EuroEwing entretanto, a análise não conseguiu segregar os tumores baseados no desfecho. Já nos tumores com baixo conteúdo estromal (< 30%), os autores não encontraram genes associados à sobrevivência (coorte EuroEwing).

Recentemente Yin e colaboradores em 2018, com o objetivo de identificar potenciais biomarcadores prognósticos, realizaram uma meta-análise de conjunto de dados de expressão de SE oriundos do *Gene Expression Omnibus* (GEO) (EDGAR, 2002). Eles utilizaram três

conjuntos de dados e selecionaram genes diferencialmente expressos em amostras de SE em comparação com tecidos normais. Esses genes foram submetidos à uma análise de rede. Como resultado, eles obtiveram 1470 genes diferencialmente expressos principalmente envolvidos em processos biológicos associados à resposta imune e regulação da transcrição. Com a análise de rede, eles conseguiram selecionar 22 genes principais. Destes genes, fizeram uma análise de Kaplan-Meier baseada em conjuntos de dados do SE do *The Cancer Genome Atlas* (TCGA; cancergenome.nih.gov/) e identificaram cinco genes, incluindo o PYGM (glicogênio fosforilase, músculo associado), MEF2C (Fator Potenciador de Miócito 2C), TRIM63 (Motivo tripartido contendo 63), BUB1B (BUB1 checkpoint mitótico serina/treonina quinase B), RACGAP1 (Proteína ativadora Ras GTPase 1), cujos perfis de expressão alterados foram significativamente associados à sobrevida. As alterações de expressão foram validadas através de RT-qPCR em linhagens de SE e de tecidos normais. Os autores concluíram que estes cinco genes podem ser considerados como potenciais biomarcadores prognósticos de SE e devem ser úteis tanto para o diagnóstico como para o tratamento precoce do tumor (YIN, *et al*, 2018).

Até o momento, poucos são os candidatos à genes prognósticos que orientem para um tratamento do Sarcoma de Ewing mais personalizado, portanto a identificação de novos marcadores moleculares ainda é uma prioridade. Dentro deste contexto, nós investigamos dados de expressão de amostras de SE obtidas no momento do diagnóstico, e que continham dados sobre o desfecho dos pacientes. Nossa estratégia foi diferente dos trabalhos anteriores, pois procuramos genes diferencialmente expressos entre tumores com prognóstico bom e ruim, e não entre tecido normal e câncer. Acreditamos que desta forma conseguiremos descobrir os genes que são responsáveis pelos diferentes desfechos de SE. Outro motivo por não comprar o tecido normal ao câncer é a incerteza da célula de origem do SE. É de alta relevância definir, através da expressão gênica, genes prognósticos candidatos, que sejam validados em coortes distintas. Estas descobertas, podem auxiliar na personalização do tratamento para aqueles pacientes que apresentarem doença com perfil mais agressivo ao diagnóstico.

1.3 Justificativa

O câncer infantil é considerada uma doença rara quando comparada a população adulta e os incentivos à pesquisa em crianças são escassos. Entretanto, o câncer infantil é a primeira causa de morte por doença no Brasil e novas opções terapêuticas são essenciais para alavancar as taxas de cura na nossa população. O SE é um tumor pediátrico altamente agressivo e tem um impacto importante no sistema de saúde uma vez que ocorre prevalentemente em crianças e adultos jovens. Devido à sua agressividade, o controle local da doença é feito com cirurgia e/ou radioterapia, juntamente com quimioterapia intensiva para tratar micrometástases. Estas terapias multimodais melhoram dramaticamente a sobrevivência de pacientes com SE localizado no momento do diagnóstico de 50% a 70% (BACCI, *et al*, 2004; GRIER, *et al*, 2003; KOLB *et al*, 2003). Salienta-se que o aumento da sobrevida a longo prazo é obtida através de uma elevada toxicidade e risco considerável de desenvolver câncer secundário, ameaças essas com alto impacto a vida principalmente se considerar a idade jovem dos pacientes com SE. Outro ponto a ser levantado refere-se aos pacientes com metástase ao diagnóstico que apresentam um prognóstico de sobrevida global de 20 a 30%, e do grupo de pacientes que apresentam recorrência da doença cujo prognóstico é ainda muito baixo.

Assim a detecção de genes prognósticos de desfecho do SE no momento do diagnóstico poderia orientar para um protocolo de tratamento mais eficaz e individualizado de acordo com o perfil de agressividade do tumor. Essas informações são cruciais para definir subpopulações de pacientes com diferentes perfis biológicos tumorais, por exemplo, maior agressividade e mortalidade. A partir dessas assinaturas é possível selecionar novos alvos potencialmente modulados farmacologicamente que melhorem a sobrevida dos pacientes.

2. OBJETIVO

2.1 Objetivo Geral

Analisar dados de expressão de amostras de Sarcoma de Ewing com diferentes desfechos clínicos disponíveis em banco de dados públicos, como *Gene Expression Omnibus* (GEO) com intuito de selecionar vias e/ou genes marcadores de prognóstico em SE.

2.2 Objetivos Específicos

- Analisar dados de expressão de amostras de pacientes com SE utilizando o software transcriptogramer, comparando SOB (sobreviventes, com sobrevida maior do que cinco anos) vs NSOB (Não sobreviventes, com sobrevida menor do que cinco anos);
- Selecionar as vias diferencialmente expressas envolvidas nos dois desfechos (SOB, NSOB);
- Selecionar os genes mais significativos dentro de cada via;
- Avaliar o valor prognóstico da expressão de cada gene através da curva de sobrevivência de Kaplan Meier;
- Selecionar os genes com valor prognóstico concordante em diferentes coortes de SE.

3 MATERIAL E MÉTODOS

3.1 Os dados

Os dados de expressão utilizados neste trabalho foram obtidos do *Gene Expression Omnibus* (GEO) (EDGAR, 2002) (códigos de acesso GSE63155, GSE63156 e GSE17679). Os conjuntos de dados GSE63155 e GSE63156 compreendem dados de expressão gênica (Affymetrix Human Exon 1.0 ST Array) de biópsias de Sarcoma de Ewing obtidas respectivamente de 39 pacientes do *Children's Oncology Group* (COG) e de 46 pacientes do grupo colaborativo EuroEwing (Volchenboun, *et al*, 2015). O conjunto de dados GSE17679 compreende dados de expressão gênica (Affymetrix Human Genome U133 Plus 2.0 Array) de biópsias de Sarcoma de Ewing e de linhagens celulares de Sarcoma de Ewing (SAVOLA, *et al*, 2009).

3.2 Classificação das amostras

Com intuito de identificar um grupo de genes prognósticos de desfecho dividimos cada coorte em dois grupos: (i) sobreviventes, com sobrevida maior do que cinco anos e (ii) não sobreviventes, com sobrevida menor do que cinco anos. Além disso, os pacientes identificados como não sobreviventes e com sobrevida maior de cinco anos foram excluídos da análise.

3.3 Normalização

A normalização e o controle de qualidade dos dados de microarranjo (GSE63155 e GSE63156, Affymetrix Human Exon 1.0 ST Array, GPL5175) foi realizada com o oligo *Bioconductor/R* (CARVALHO, *et al*, 2010). Para o GSE17679 (Affymetrix Human Genome U133 Plus 2.0 Array, GPL570) a normalização e o controle de qualidade foi realizada com o pacote *Affy Bioconductor/R* (GAUTIER, *et al*, 2004). As anotações para a GPL5175 e GPL570 foram obtidas no banco de dados disponibilizado em <http://lief.if.ufrgs.br/pub/biosoftwares/transcriptogramer/dictionary/Homosapiens/>.

3.4 Transcriptogramas

A técnica do Transcriptograma (RYBARCZYK-FILHO & CASTRO, *et al*, 2011; DA SILVA & PERRONE, *et al*, 2014) consiste em produzir perfis de expressão gênica a partir de

medidas de expressão de genoma inteiro, tais como as obtidas por microarranjos ou RNA-Seq. Os perfis consistem na projeção dos dados de expressão sobre uma lista de genes ordenadas de tal maneira que os genes pertencentes a uma mesma função biológica estão localizados em regiões próximas na lista. A lista ordenada é feita com base na informação sobre interação proteína-proteína, como disponibilizado na base de dados STRING (SZKLARCZYK, MORRIS, *et al*, 2017), por meio de uma simulação de Monte Carlo que minimiza as distâncias na lista de produtos gênicos que participam de uma mesma rota metabólica. Neste trabalho, utilizamos a lista ordenada para *Homo sapiens* como obtida por de Almeida e colaboradores (DE ALMEIDA & CLENDENON, *et al*, 2016). A lista ordena 9684 genes, cujos produtos gênicos foram listados como associados a pelo menos um outro produto gênico com escore de confiança 800.

A figura abaixo, da ref. (DE ALMEIDA & CLENDENON, *et al*, 2016), mostra que a lista ordenada apresenta uma lógica biológica. Algumas rotas metabólicas como listada pela site KEGG (KANEHISA & GOTO, *et al*, 2014) e pelo Gene Ontology (GO) (CARBON & IRELAND, *et al*, 2009) são projetadas sobre a lista ordenada da seguinte maneira. Para cada rota ou conjunto gênico, a cada posição da lista é atribuído o valor 1 ou 0 dependendo se o produto gênico naquela posição faz ou não parte da lista. A seguir para cada posição da lista calcula-se a média de zeros e uns da posição e seus 30 vizinhos à direita e 30 vizinhos à esquerda. O resultado são perfis de probabilidade de se encontrar um gene de um determinado conjunto gênico em uma dada região da lista, isto é, os picos indicam onde os genes da lista estão localizados.

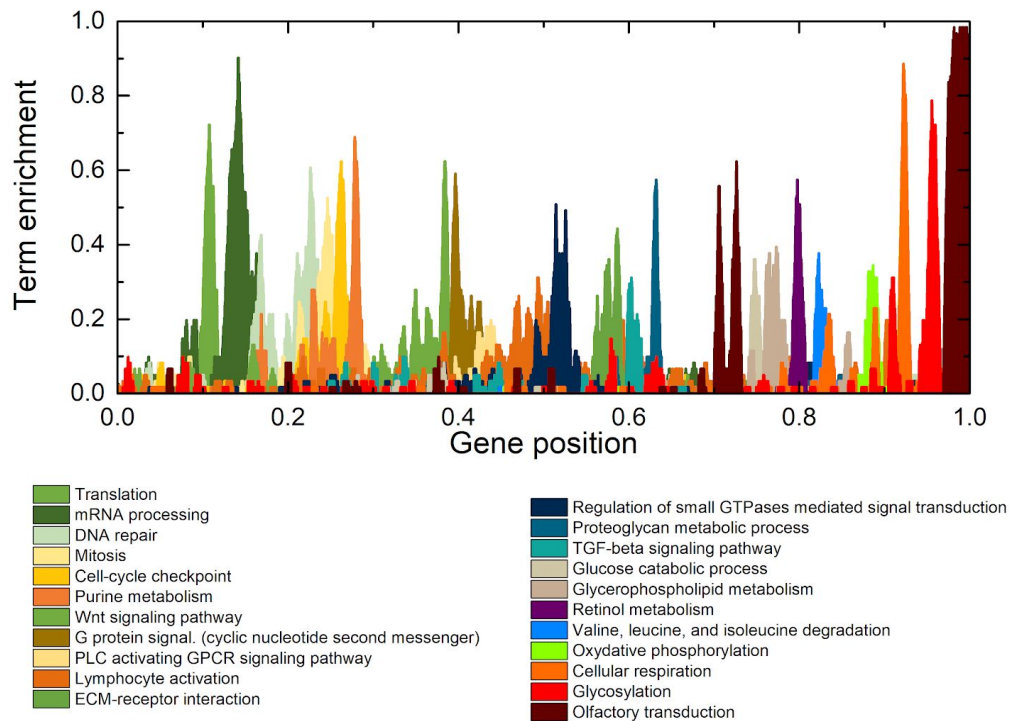


Figura 1. Mapa dos enriquecimento da lista ordenada de 9684 genes por termos do Gene Ontology e por rotas metabólicas do KEGG. Os perfis coloridos, bastante localizados, mostram as distribuições de probabilidade de encontrar os genes de um dado conjunto gênico. Um perfil de valor um em um ponto da lista significa que todos os genes em uma região definida pelo ponto e pelas 30 vizinhos à esquerda e 30 vizinhos à direita pertencem ao conjunto gênico.

Observe que a lista está aumentando o poder do teste (DA SILVA & PERRONE, *et al*, 2014). Além de humanos, esta técnica já foi aplicada com sucesso para *Saccharomyces cerevisiae* (RYBARCZYK-FILHO & CASTRO, *et al*, 2011), *Cryptococcus* (FERRAREZE & STREIT, *et al*, 2017) e *Glycine max* (MIOTTO & DA COSTA, *et al*, 2019).

Com a lista ordenada dados relativos à expressão gênica de cada amostra são projetados de tal maneira que a expressão de cada gene é associada à posição do mesmo e então uma média sobre as expressões dos genes de cada posição e em posições vizinhas de raio R à esquerda e à direita é atribuída ao gene central. Neste trabalho, escolhemos R=60. Esta média janelada reduz o ruído inerente à técnica de medidas de expressão gênica enquanto que a lista ordenada pelas funções biológicas possibilita que a razão sinal-ruído seja melhorada.

3.5 Identificação de vias candidatas diferencialmente expressas

A análise dos dados de expressão entre os dois grupos foi realizada utilizando o software *Transcriptograder* (<https://lief.if.ufrgs.br/pub/biosoftwares/transcriptograder/>) com a opção raio 60. As regiões diferentemente expressas foram selecionadas e os genes contidos nessas regiões foram identificados, resultando num grupo de genes diferentemente expressos (GGDE). O grupo de genes diferencialmente expressos foram utilizados para realizar o enriquecimento funcional no *DAVID Tools 6.8* (<https://david.ncifcrf.gov/home.jsp>) (HUANG, *et al*, 2008a; HUANG, *et al*, 2008b). Por meio dessa análise identificamos os termos do GO (Gene Ontology), as vias do Kegg (KANEHISA, 2000) e Reactome (FABREGAT, *et al*, 2017) significativamente associadas com o GGDE. Foram considerados significantes os termos e vias com p-valor menor a 0,05.

Para cada ontologia selecionada, fizemos o *download* de todos os genes que a compõem por meio do programa “download_onto.py”, de autoria de Ferreira (2019). A análise das vias diferencialmente expressas foi realizada com o software *Transcriptograder*. As vias com p-valor < 0.05 foram selecionadas bem como os dez genes com p-valores mais significantes dentro de cada via. O conjunto de genes resultantes foi utilizado para as próximas análises.

3.6 Análise por diagrama de Kaplan-Meier para genes diferencialmente expressos.

Para investigar a possibilidade de utilizarmos a expressão diferencial destes genes como prognóstico de pacientes, usamos o pacote *Survival Bioconductor / R* para obtermos os diagramas de Kaplan-Meier para a sobrevida global do paciente em relação a cada um dos genes. As amostras foram divididas em dois grupos com base na expressão do gene: (i) expressão alta e (ii) expressão baixa. Para isso, utilizamos o método de escaneamento (*scan*) que testa todos pontos de corte de expressão para determinar o valor que melhor separa as duas classes. Neste procedimento escolhemos o menor tamanho de grupo como sendo oito.

4 RESULTADOS E DISCUSSÃO

O sarcoma de Ewing é o segundo tumor ósseo infantil mais frequente. Pode ocorrer tanto em tecidos ósseos quanto em tecidos moles e está associado a uma alta taxa de recorrência e metástase. Embora a sobrevida tenha aumentado para tumores localizados (70%), a expectativa de vida para tumores metastáticos e com recaída é ainda muito baixa (20%) (GRÜNEWALD, *et al*, 2018). Nos últimos anos muitos esforços foram destinados à encontrar biomarcadores prognósticos, que podem auxiliar no tratamento de maneira significativa. Neste estudo utilizamos dados públicos depositados no GEO de três coortes independentes contendo amostras de pacientes com SE. Através do *software* Transcriptogramer, fomos capazes de identificar grupos de genes diferencialmente expressos que foram utilizados para estimar curvas Kaplan-Meier.

4.1 Organização e escolha dos dados

O objetivo deste trabalho é obtermos um conjunto de genes cuja expressão pode servir de prognóstico de sobrevivência. Para eliminar fatores de confusão e dividir de forma mais precisa nossas amostras, utilizou-se apenas as amostras cujos doadores constavam como sobreviventes e com sobrevida maior de cinco anos (SOB) e cujos doadores constavam como não sobreviventes e com sobrevida menor do que cinco anos (NSOB), foram selecionadas apenas amostras de tumor primário, cujos pacientes ainda não haviam recebido nenhum tipo de tratamento. As amostras que não se encaixaram nestes filtros foram eliminadas (tabela 1). Ao total foram selecionamos 86 amostras distribuídas nas três coortes, onde 51 amostras pertencem ao grupo dos pacientes Sobreviventes com sobrevida maior que cinco anos (SOB) e 35 ao grupo dos Não Sobreviventes com sobrevida menor do que cinco anos (NSOB). Para a coorte do COG selecionamos 32 amostras, sendo 20 (SOB) e 12 (NSOB). A coorte do EuroEwing ficou com um total de 30 amostras, sendo 20 (SOB) e 10 (NSOB). Por fim, a coorte dos finlandeses teve um total de 25 amostras, onde 12 (SOB) e 13 (NSOB).

Tabela 1. Relação de amostras (ID), desfecho e Sobrevida global (SG) em meses para as três coortes (COG, EuroEwing e Filandeses).

COG (GES63155)			EuroEwing (GSE63156)			Finlandeses (GSE17679)		
ID	Desfecho	SG	ID	Desfecho	SG	ID	Desfecho	SG
L309	SOB	90,20	Q171	SOB	113,17	R194	NSOB	15,60
L310	SOB	80,13	Q174	SOB	79,00	R196	NSOB	20,70
L311	SOB	75,10	Q447	SOB	142,03	R29	NSOB	4,60
L312	SOB	73,17	Q455	SOB	80,07	R33	NSOB	16,60
L313	NSOB	20,07	Q456	NSOB	10,90	R38	NSOB	11,30
L388	NSOB	53,03	Q458	SOB	93,50	R39	NSOB	47,00
L389	SOB	86,10	Q460	NSOB	32,50	R40	SOB	117,30
L391	SOB	86,93	Q462	SOB	150,00	R41	SOB	125,10
L392	NSOB	8,37	Q464	SOB	85,03	R42	SOB	192,20
L393	NSOB	30,87	Q465	SOB	86,10	R53	SOB	87,70
L395	SOB	74,17	Q470	SOB	143,87	R55	NSOB	8,50
L676	SOB	78,60	Q908	NSOB	11,30	R57	NSOB	8,70
L678	SOB	66,93	Q914	SOB	125,47	R58	NSOB	13,40
L811	NSOB	29,47	Q919	SOB	67,83	R60	SOB	121,50
L813	NSOB	5,17	Q927	NSOB	22,87	R62	SOB	117,60
L819	SOB	73,70	Q930	NSOB	20,83	R64	NSOB	21,30
M944	NSOB	56,47	Q932	NSOB	47,53	R65	NSOB	11,00
M950	SOB	61,10	Q938	SOB	113,17	R67	NSOB	29,40
N838	SOB	132,90	Q942	SOB	108,70	R69	SOB	62,60
N842	SOB	112,57	Q947	SOB	101,27	R72	SOB	117,70
N845	NSOB	21,00	Q949	SOB	106,20	R74	SOB	110,40
N848	SOB	101,67	Q950	SOB	85,70	R75	NSOB	25,40
N850	SOB	104,17	Q952	SOB	83,47	R78	SOB	69,50
N854	SOB	92,30	Q954	SOB	77,63	R79	SOB	127,10
N856	NSOB	7,57	Q995	NSOB	9,53	R80	SOB	68,50
O436	SOB	71,17	R001	SOB	76,43			
O437	SOB	67,23	R002	SOB	71,33			
O442	NSOB	13,97	R004	NSOB	51,10			
O963	NSOB	60,07	R006	NSOB	22,77			
O965	SOB	75,37	R009	SOB	62,53			
O966	SOB	61,30						
P101	NSOB	57,77						

4.2 As coortes COG e EuroEwing não podem ser analisadas em conjunto, pois não são homogêneas.

Em um primeiro momento foi avaliada se as amostras de SE da coorte dos pacientes do COG e do EuroEwing eram homogêneas, salientando que os dados são da mesma plataforma (GPL5175, Affymetrix Human Exon 1.0 ST Array) e foram realizados pelo mesmo laboratório (VOLCHENBOUM, *et al*, 2015).

A partir da análise do Transcriptograma (Figura 2) foi evidente as diferenças ao longo de toda a expressão gênica. Esta análise mostrou que as coortes apresentam peculiaridades e não podem ser analisadas em conjunto, mesmo sendo da mesma plataforma.

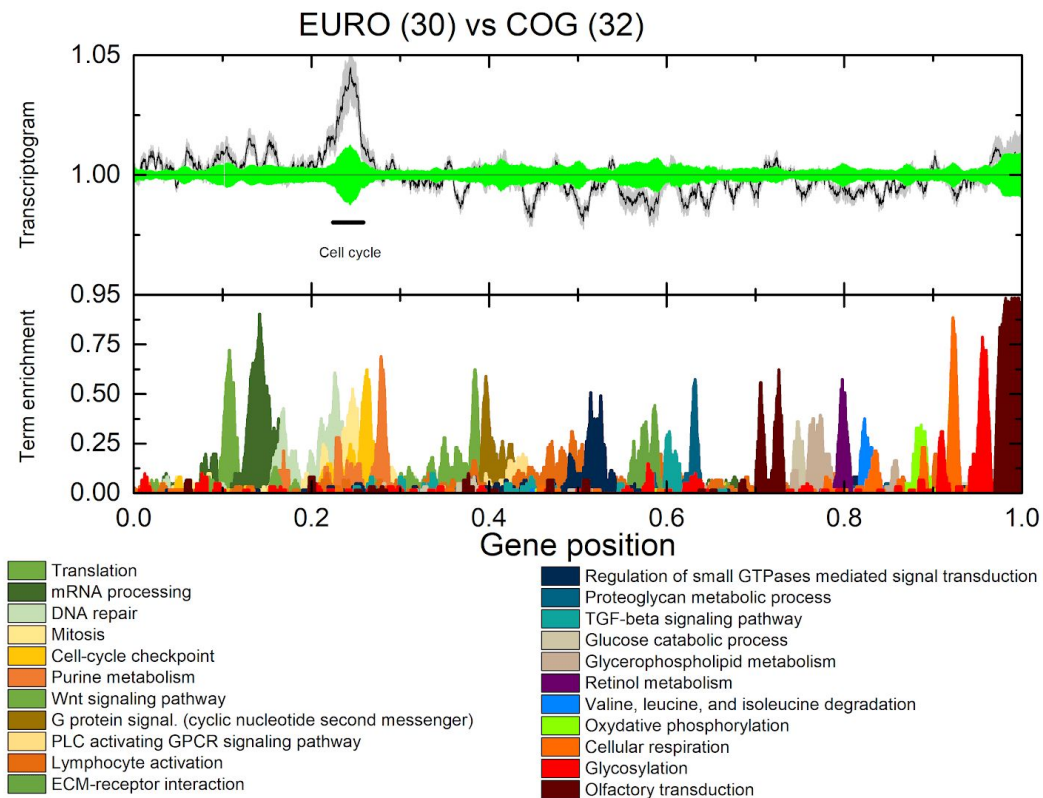


Figura 2. Transcriptograma relativo comparando as amostras de SE de duas coortes de pacientes EuroEwing e COG. Os dados de microarranjo foram obtidos pelo mesmo laboratório e mesma plataforma (GPL5175, Affymetrix Human Exon 1.0 ST Array). A análise evidencia que as coortes apresentam um padrão de expressão gênica diferente ao longo do ordenamento e não podem ser consideradas homogêneas.

4.3 As amostras dos pacientes do COG possuem quantidades consideráveis de estroma

Volchenboum e colaboradores (2015) também analisaram os dados de expressão de SE da coorte do COG e encontraram uma lista de genes diferencialmente expressos entre sobreviventes e não sobreviventes as quais foram utilizadas para identificar assinatura gênica prognóstica. Entretanto, os autores não conseguiram validar a assinatura gênica nas amostras dos pacientes do Grupo Cooperativo Europeu (EuroEwing) e concluíram que as assinaturas estavam presentes no conteúdo estromal do tumor, justificando que as amostras de SE dos pacientes do EuroEwing tinham menor quantidade de estroma.

Com intuito de verificar se a quantidade de estroma poderia estar interferindo na análise dividiu-se a coorte do COG em dois grupos: alto conteúdo tumoral (>70%) e baixo conteúdo tumoral (<70%) como pode ser visualizado na Tabela 2.

Tabela 2. Relação de amostras (ID) e conteúdo tumoral da coorte COG.

COG (GSE63155)					
ID	Conteúdo tumoral	ID	Conteúdo tumoral	ID	Conteúdo tumoral
L309	>70%	L395	>70%	N850	>70%
L310	<70%	L676	>70%	N854	>70%
L311	<70%	L678	<70%	N856	>70%
L312	>70%	L819	<70%	O436	>70%
L313	<70%	M944	>70%	O437	>70%
L388	<70%	M950	>70%	O963	>70%
L389	>70%	N838	>70%	O965	>70%
L391	>70%	N842	<70%	O966	>70%
L392	>70%	N845	>70%	P101	>70%
L393	<70%	N848	>70%		

4.4 Transcriptograma realizado apenas com amostras contendo baixo conteúdo tumoral não apresenta diferenças evidentes entre os grupos SOB e NSOB.

A comparação entre os grupos SOB (N=5) vs NSOB (N=3) de amostras com baixo conteúdo tumoral foi feita através do software Transcriptogramer. Não foi encontrada diferença significativa entre esses dois grupos, pode-se observar um desvio padrão alto, devido ao pequeno número amostral (Figura 3).

As amostras com baixo conteúdo estromal e com conteúdo tumoral maior do que 70% (21), também foram avaliadas. Comparou-se SOB (N=06) vs NSOB (N=15) gerando um transcriptograma, onde foi evidente as diferenças entre os dois grupos (figura 4). Portanto, as diferenças foram observadas entre SOB e NSOB quando temos alto conteúdo tumoral. As amostras dos pacientes do EuroEwing e dos Finlandeses possuíam conteúdo tumoral >75% e >85%, respectivamente, portanto as três coortes foram utilizadas nas análises posteriores.

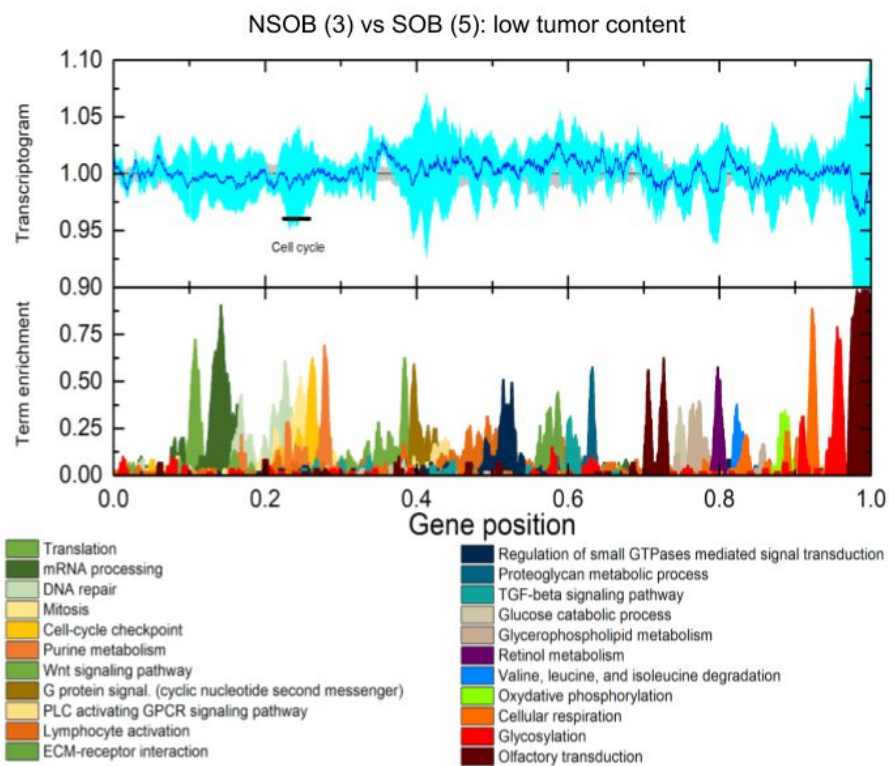


Figura 3. Transcriptograma relativo entre não sobreviventes (NSOB) e sobreviventes (SOB), realizado a partir dos dados de expressão gênica da coorte do COG das amostras com baixo conteúdo tumoral (< 70%). O perfil de expressão relativo não apresentou regiões onde as diferenças entre as categorias fossem maiores do que o desvio padrão das categorias.

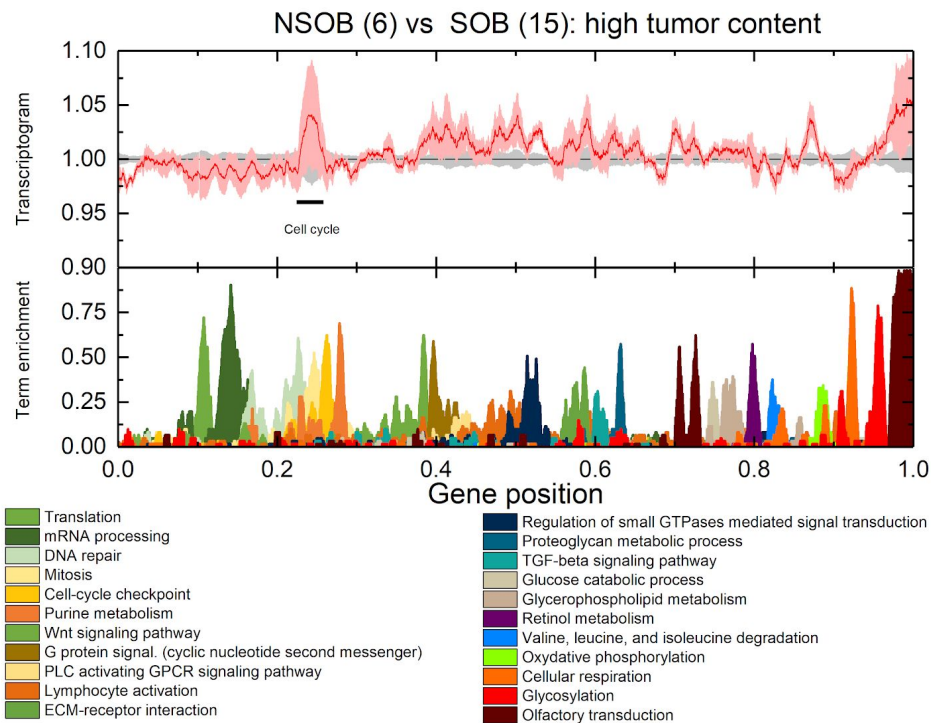


Figura 4. Transcriptograma relativo entre não sobreviventes (NSOB) e sobreviventes (SOB), realizado a partir dos dados de expressão gênica da coorte do COG das amostras com alto conteúdo tumoral (> 70%). O perfil de expressão relativo apresentou regiões onde as diferenças entre as categorias foram maiores do que o desvio padrão das categorias.

4.5 Transcriptograma da coorte COG é capaz de detectar diferenças de expressão gênica entre grupos prognósticos.

As diferenças no nível de expressão entre as categorias SOB e NSOB, foram avaliadas nas três coortes. Para a coorte do EuroEwing não foi possível detectar diferenças significativas entre os grupos (Figura 5). A difícil detecção se deve ao fato de que os sinais (expressão gênica média) dos grupos apresentaram um desvio padrão alto. Já na coorte do COG e dos Finlandeses observa-se uma clara diferença entre os sinais ao longo da expressão. Podemos detectar diversas regiões onde a expressão entre os dois grupos é diferente e o desvio padrão é menor (Figura 6 e 7).

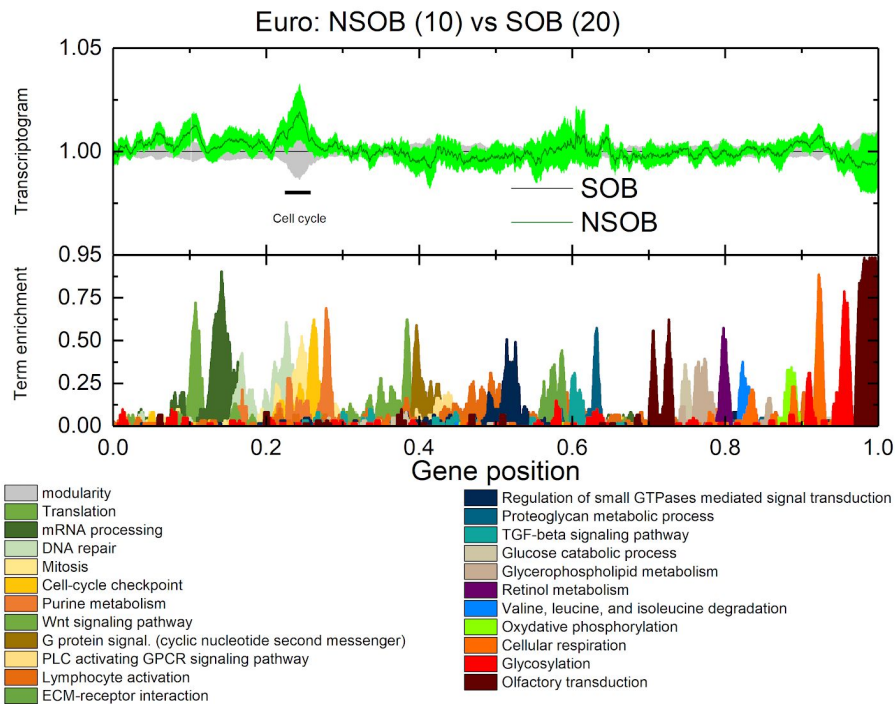


Figura 5. Transcriptograma relativo entre não sobreviventes (NSOB) e sobreviventes (SOB), realizado a partir dos dados de expressão gênica de SE da coorte do EuroEwing. O perfil de expressão relativo não apresentou regiões onde as diferenças entre as categorias fossem maiores do que o desvio padrão das categorias.

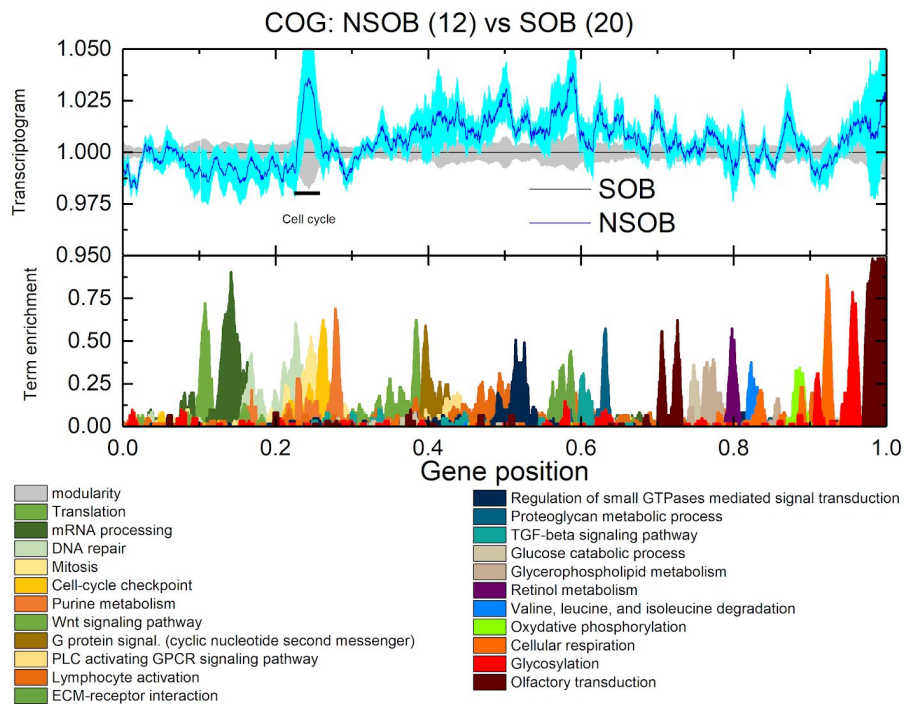


Figura 6. Transcriptograma relativo entre não sobreviventes (NSOB) e sobreviventes (SOB), realizado a partir dos dados de expressão gênica de SE da coorte do COG. O perfil de expressão relativo possui diversas regiões onde as diferenças entre as categorias foram maiores do que o desvio padrão das categorias.

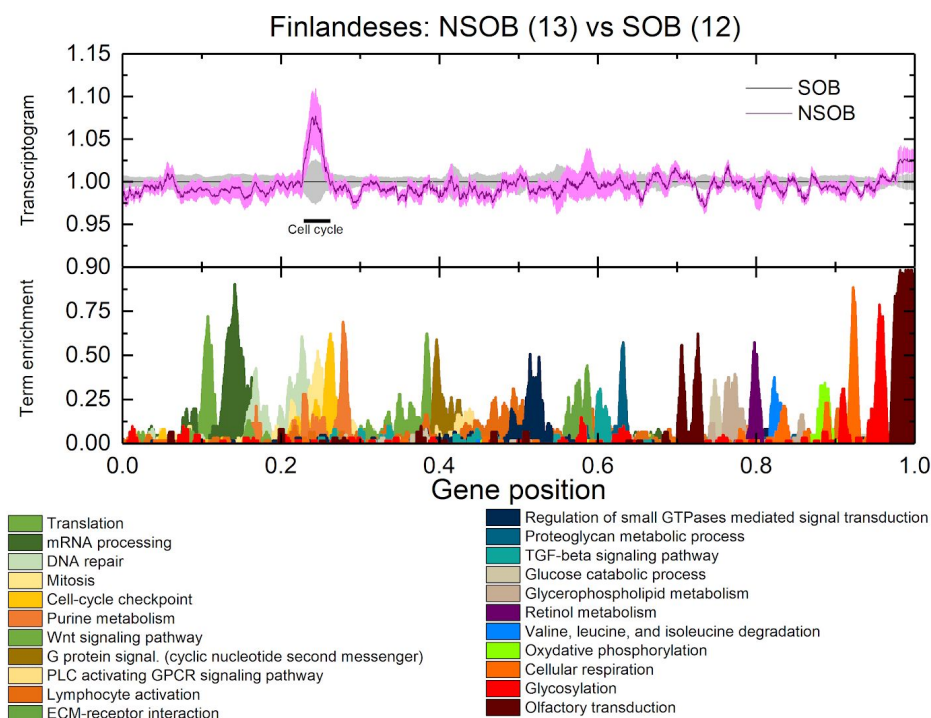


Figura 7. Transcriptograma relativo entre não sobreviventes (NSOB) e sobreviventes (SOB), realizado a partir dos dados de expressão gênica de SE da coorte dos Finlândeses. O perfil de expressão relativo possui diversas regiões onde as diferenças entre as categorias foram maiores do que o desvio padrão das categorias.

4.6 Seleção das vias e genes mais significativos.

Elegemos o COG para ser nossa coorte de referência, as análises referentes aos finlândeses serão feitas em trabalho futuro. A seleção do grupo de genes diferencialmente expressos (GGDE) entre NSOB vs SOB, foi realizada com base no seu Transcriptograma (Figura 8). Na figura 8 estão marcados as cinco regiões de GGDE que foram significativas, com p -valor $< 0,05$.

O GGDE foi utilizado para realizar o enriquecimento funcional no DAVID TOOLS. Obtivemos um total de 160 vias. Todas as vias foram baixadas e analisadas através do Transcriptogramer, sendo que 38 apresentaram diferença de expressão significativa ($P < 0,05$) quando comparado SOB vs NSOB (tabela 3). Além disso, foram selecionados os dez genes mais significativos de cada uma das vias, totalizando 111 genes ($p < 0,05$) (tabela 4).

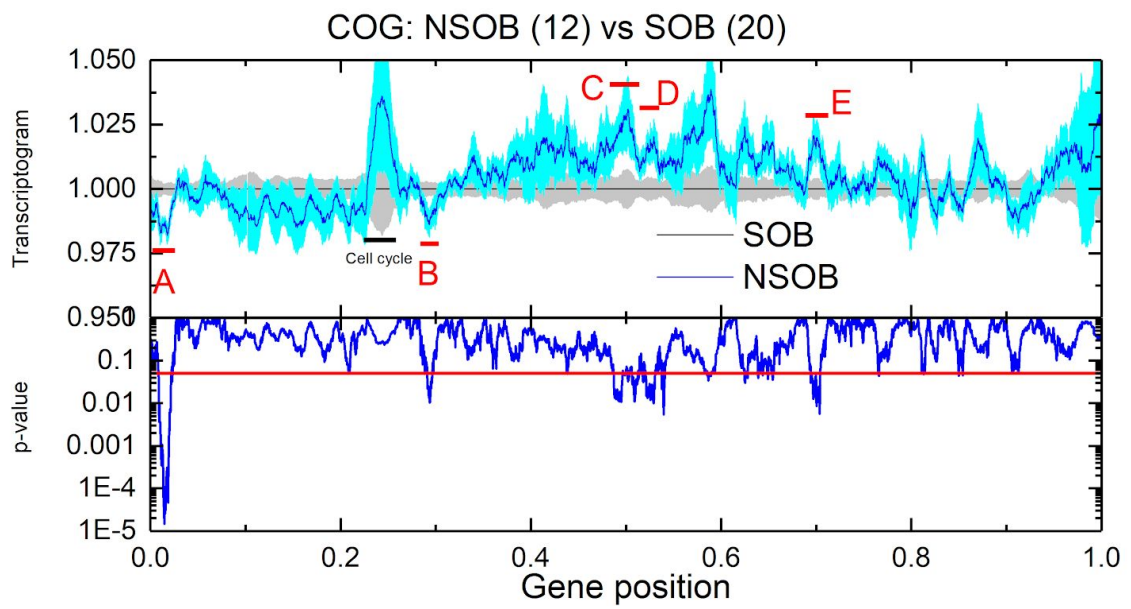


Figura 8. Transcriptograma relativo entre não sobreviventes (NSOB) e sobreviventes (SOB), realizado a partir dos dados de expressão gênica da coorte de SE do COG. Na figura é apresentado o gráfico do p-valor (linha azul marinho), com corte de p-valor $<0,05$ linha horizontal vermelha. As letras (A, B, C, D, E) indicam as regiões onde o grupo de genes diferencialmente expressos (GGDE) foram significativas e selecionado para as futuras análises.

Tabela 3. Termos do GO, vias do Kegg e Reactome que apresentaram expressão diferencial significativa quando comparado SOB vs NSOB, amostras de SE (COG).

Termos GO e vias do Kegg e Reactome	ID	p-valor
<i>Ephrin receptor signaling pathway</i>	GO:0048013	0,0020
<i>Cell morphogenesis involved in differentiation</i>	GO:0000904	0,0039
<i>EPH-Ephrin signaling</i>	R-HSA-2682334	0,0046
<i>EPHA-mediated growth cone collapse</i>	R-HSA-3928663	0,0048
<i>Regulation of GTPase activity</i>	GO:0043087	0,0082
<i>ERBB signaling pathway</i>	GO:0038127	0,0087
<i>EPH-ephrin mediated repulsion of cells</i>	R-HSA-3928665	0,0090
<i>Positive regulation of hydrolase activity</i>	GO:0051345	0,0095
<i>Positive regulation of catalytic activity</i>	GO:0043085	0,0096
<i>Fc gamma R-mediated phagocytosis</i>	hsa04666	0,0159
<i>Ras protein signal transduction</i>	GO:0007265	0,0159
<i>Positive regulation of cellular metabolic process</i>	GO:0031325	0,0161
<i>Regulation of small GTPase mediated signal transduction</i>	GO:0051056	0,0182
<i>Cytoskeleton organization</i>	GO:0007010	0,0220
<i>Cell proliferation</i>	GO:0008283	0,0234
<i>Rap1 signaling pathway</i>	hsa04015	0,0279
<i>Transmembrane receptor protein tyrosine kinase signaling pathway</i>	GO:0007169	0,0288
<i>Movement of cell or subcellular component</i>	GO:0006928	0,0290
<i>Regulation of localization</i>	GO:0006928	0,0316
<i>Phagocytosis</i>	GO:0006909	0,0336
<i>Chemotaxis</i>	GO:0006935	0,0343
<i>Intracellular signal transduction</i>	GO:0035556	0,0374
<i>Cell differentiation</i>	GO:0030154	0,0375
<i>Ras signaling pathway</i>	hsa04014	0,0398
<i>PI3K-Akt signaling pathway</i>	hsa04151	0,0408
<i>Regulation of response to stress</i>	GO:0080134	0,0413
<i>Positive regulation of cellular component movement</i>	GO:0051272	0,0424
<i>MAPK signaling pathway</i>	hsa04010	0,0426
<i>Phosphorylation</i>	GO:0016310	0,0426
<i>Cell migration</i>	GO:0016477	0,0431
<i>Cell motility</i>	GO:0048870	0,0435
<i>Localization of cell</i>	GO:0051674	0,0435
<i>Cellular component organization</i>	GO:0016043	0,0467
<i>Fc receptor mediated stimulatory signaling pathway</i>	GO:0002431	0,0469
<i>Regulation of cell migration</i>	GO:0030334	0,0477
<i>Cell adhesion</i>	GO:0007155	0,0492
<i>Enzyme linked receptor protein signaling pathway</i>	GO:0007167	0,0493
<i>Regulation of cell motility</i>	GO:2000145	0,0497

Dentre as vias que apresentaram diferenças de expressão entre pacientes com SE SOB e NSOB encontram-se quatro vias relacionadas a efrinas. De acordo com a literatura os receptores Eph constituem a maior família de receptores de tirosina quinases, os quais são ativados por ligantes de efrina que se encontram ancorados à membrana ou que possuem um domínio transmembrana. Já é bem estabelecido o papel dessas moléculas no desenvolvimento dos organismos multicelulares apresentando funções importantes no desenvolvimento neuronal, padrão vascular e inflamação durante a lesão tecidual (KOU & KANDPAL, 2018). Efrinas e receptores Eph afetam características cruciais para o desenvolvimento e manutenção tumoral como o crescimento, migração e invasão observado em células cancerígenas em cultura, bem como o crescimento tumoral, invasividade, angiogênese e metástase *in vivo* (PASQUALE, 2010). Na última década essas moléculas se tornaram alvos potenciais para terapêutica do câncer devido a relação da sua expressão aberrante com alterações no potencial proliferativo, de migração e invasão de uma série de tumores. O envolvimento de receptores Eph e ligantes de efrina já foram descritos para o câncer de pulmão, mama, próstata, colorretal, glioblastoma e meduloblastoma (KOU e KANDPAL, 2018).

Neste trabalho encontramos um aumento da expressão das vias relacionadas a efrina nos pacientes com SE que tiveram sobrevida menor do que cinco anos. Nossos achados estão de acordo com Jaspeado e colaboradores que demonstraram o envolvimento de EphA2 na promoção da angiogênese em células de SE através de sinalização dependente de ligante (JASPEADO, *et al*, 2013). Além disso, os pesquisadores mostraram que a proteína CAV1 interage com EphA2 em células de SE promovendo o processo de angiogênese dependente de quinase (JASPEADO *et al*, 2013). Dentre os receptores de efrina, o EphA2 desempenha um papel importante na oncogênese, metástase e resistência ao tratamento. Além disso, o EphA2 já foi implicado em mimetismo vascular, um importante processo que contribui para a malignidade do SE e pior prognóstico (VAN DER SCHAFT, *et al*, 2005). Recentemente, um grupo de pesquisadores mostrou que o EphA2 é fosforilado na S897 na maioria das linhagens celulares de SE e que há uma relação entre EphA2 e CAV1, uma vez que a inibição de CAV1 resulta em uma perda significativa da fosforilação de S897 (MONCLÚS, *et al*, 2018). Além disso, usando mutagênese dirigida através de experimentos de ganho e perda de função, os autores demonstraram que EphA2 é essencial para as propriedades agressivas do SE de uma maneira independente de quinase (MONCLÚS, *et al*, 2018). Esses achados da literatura,

juntamente com os achados deste trabalho mostram a importância das vias de sinalização de efrina em SE.

É importante salientar que neste trabalho, além das vias de sinalização de efrina, também foi identificado um aumento da expressão das vias relacionadas com proliferação e migração celular nos pacientes com SE (coorte COG) com sobrevida menor do que 5 anos. Mostrando que esses processos estão aumentados nestes pacientes, já no momento do diagnóstico, caracterizando um perfil de doença mais agressiva. Além disso, a sinalização alterada das efrinas poderia estar contribuindo para o aumento da proliferação e migração celular visualizada neste trabalho. De fato a literatura aponta que Efrinas e receptores Eph afetam vias importantes relacionadas ao crescimento, migração e invasão tumoral (PASQUALE, 2010).

Outra via detectada com aumento de expressão significativo em NSOB foi a via de sinalização ERBB. De acordo com a literatura a família de receptores transmembrana de tirosina quinase ERBB é formada por 4 receptores de fator de crescimento epidérmico: EGFR (ERBB1), HER2 (ERBB2), HER3 (ERBB3), and HER4 (ERBB4) e atuam na regulação de vários processos biológicos como proliferação, diferenciação, mobilidade celular e sobrevivência. Em vários tipos diferentes de tumores, os membros da família ERBB encontram-se na maioria das vezes superexpressos, amplificados ou mutados (COLLIN *et al*, 2015). Além disso, a via de sinalização de ERBB faz parte de uma rede de sinalização complexa, que ativa vias de sinalização, incluindo PI3K/AKT, Ras/Raf/MAPK, JAK/STAT e PKC (MISHRA, *et al*, 2017). Salienta-se que neste trabalho foi detectado um aumento da expressão das vias PI3K/AKT e MAPK nas amostras de tumor primário de pacientes com SE NSOB, mostrando uma relação entre as vias de ERBB, PI3K/AKT e MAPK e uma maior agressividade tumoral.

Outras vias alteradas quando comparadas amostras de SE de SOB vs NSOB foram as vias de fagocitose e regulação da resposta ao estresse. Estudos demonstram que as condições de hipóxia no microambiente tumoral pode levar a um estresse das células do SE, e esse processo acaba desencadeando a síntese de proteínas adaptativas ao estresse (SOMASEKHARAN, *et al*, 2015; GRUNEWALD, *et al*, 2018). Embora não há um modelo determinado de metástase do SE vários estudos sugerem que os principais fatores metastáticos nessa doença, parecem influenciar as vias adaptativas ao estresse. Nesse sentido foi

evidenciado que ERBB4 além de se encontrar superexpresso em linhagens celulares de SE metastático e contribuir para metástase pela ativação das vias PI3K-AKT e FAK, que cooperam para aumentar a sobrevivência de células tumorais descoladas da matriz extracelular (supressão de anoikis); também aumenta a capacidade invasiva do tumor (MENDOZA-NARANJO, *et al*, 2013). Assim de acordo com os dados obtidos neste trabalho que evidencia que as vias de ERBB e regulação da resposta ao estresse estão super expressas nos pacientes NSOB juntamente com os achados da literatura discutidos acima, sugere-se que estas vias estão correlacionadas e atuam em conjunto contribuindo para um fenótipo tumoral mais agressivo do SE.

Tabela 4. Lista de genes mais significativos da coorte COG.

Gene	p-valor	Gene	p-valor	Gene	p-valor	Gene	p-valor
<i>ULK2</i>	0,00014	<i>SPAG9</i>	0,00381	<i>PLXNA2</i>	0,01036	<i>EFNB2</i>	0,02597
<i>COX15</i>	0,00025	<i>THBS3</i>	0,00398	<i>AHI1</i>	0,01078	<i>ADCYAP1R1</i>	0,02620
<i>RGS16</i>	0,00029	<i>SEMA4C</i>	0,00445	<i>SP1</i>	0,01173	<i>PTPN2</i>	0,02769
<i>TRIM5</i>	0,00032	<i>IFNA6</i>	0,00463	<i>RASGRP3</i>	0,01229	<i>MMP9</i>	0,02847
<i>LPL</i>	0,00049	<i>METTL3</i>	0,00466	<i>ENG</i>	0,01252	<i>PRKD1</i>	0,03006
<i>CAT</i>	0,00052	<i>TRIM32</i>	0,00491	<i>PLAUR</i>	0,01309	<i>FYN</i>	0,03009
<i>LAMA3</i>	0,00072	<i>VPS4B</i>	0,00521	<i>CHN1</i>	0,01369	<i>PGF</i>	0,03009
<i>TNKS2</i>	0,00076	<i>INSIG1</i>	0,00550	<i>MECOM</i>	0,01479	<i>SPHK1</i>	0,03208
<i>ELP2</i>	0,00094	<i>CCNG1</i>	0,00551	<i>THBS1</i>	0,01481	<i>FLT4</i>	0,03221
<i>MLH3</i>	0,00100	<i>RBL2</i>	0,00553	<i>RASSF2</i>	0,01490	<i>CDK5R1</i>	0,03255
<i>NBR1</i>	0,00108	<i>TACCI</i>	0,00554	<i>COL1A1</i>	0,01513	<i>RALBP1</i>	0,03265
<i>CNTRL</i>	0,00116	<i>APBA1</i>	0,00556	<i>BBS4</i>	0,01551	<i>KITLG</i>	0,03359
<i>ACAA2</i>	0,00120	<i>RPS6KA2</i>	0,00569	<i>HSPA6</i>	0,01560	<i>SCAI</i>	0,03446
<i>SPP1</i>	0,00131	<i>GRB10</i>	0,00575	<i>THY1</i>	0,01581	<i>BECN1</i>	0,03478
<i>RAPGEF5</i>	0,00134	<i>SETMAR</i>	0,00648	<i>SLIT2</i>	0,01621	<i>OGT</i>	0,03502
<i>TSC1</i>	0,00142	<i>ITGA9</i>	0,00737	<i>SH3KBP1</i>	0,01706	<i>EPHA3</i>	0,03567
<i>PPP3CB</i>	0,00144	<i>CCR1</i>	0,00777	<i>RAPGEF3</i>	0,01734	<i>ITGA2</i>	0,03601
<i>RBM23</i>	0,00170	<i>NPR2</i>	0,00780	<i>SETDB1</i>	0,02020	<i>GNB4</i>	0,03716
<i>VAMP2</i>	0,00215	<i>WWC1</i>	0,00837	<i>DOCK2</i>	0,02049	<i>RHOBTB2</i>	0,03794
<i>PIK3C3</i>	0,00217	<i>PITRM1</i>	0,00880	<i>PSEN1</i>	0,02055	<i>KDR</i>	0,03815
<i>MKS1</i>	0,00234	<i>SIRT1</i>	0,00880	<i>IL1B</i>	0,02165	<i>RASGRF2</i>	0,03827
<i>IFT88</i>	0,00251	<i>COL15A1</i>	0,00894	<i>CYTH1</i>	0,02264	<i>FCER1G</i>	0,03962
<i>ATG14</i>	0,00279	<i>MACF1</i>	0,00920	<i>CACNB2</i>	0,02264	<i>PPARG</i>	0,04108
<i>DCT</i>	0,00284	<i>SS18</i>	0,00970	<i>NCK2</i>	0,02348	<i>NCKAP1L</i>	0,04297
<i>C5</i>	0,00300	<i>PKP4</i>	0,00997	<i>ABII</i>	0,02363	<i>SYK</i>	0,04522
<i>SMAD4</i>	0,00300	<i>IL6</i>	0,01021	<i>DEPDC1B</i>	0,02391	<i>AMPH</i>	0,04649
<i>PEA15</i>	0,00339	<i>NRP2</i>	0,01026	<i>AKT3</i>	0,02500	<i>EFNA1</i>	0,04781
<i>ARAP3</i>	0,00345	<i>RGS6</i>	0,01027	<i>GRAP</i>	0,02575		

4.7 Vinte e três genes tiveram Kaplan-Meier significativo em todas as três coortes.

Os genes selecionados na coorte do COG (tabela 4) foram utilizados para criar curvas de Kaplan-Meier de sobrevida Global que relacionam o nível de expressão (alto ou baixo) com o desfecho (prognóstico bom ou ruim). Essas análises foram feitas para cada um dos 111 genes individualmente. Para avaliar o impacto desses genes em diferentes amostras de SE, os

111 genes foram avaliados na sua coorte de origem (COG), e nas outras duas coortes (EuroEwing e Finlandeses).

Como resultado obteve-se 23 genes com Kaplan-Meier significativo, com a mesma relação de nível de expressão e desfecho para as três coortes (tabela 5). Na figura 9, pode-se visualizar o gráfico das curvas de Sobrevida Global de Kaplan-Meier dos 23 genes para as três coortes. Dos 23 genes, 17 apresentaram bom prognóstico quando a expressão era alta e seis quando a expressão era baixa (Tabela 5 e Figura 10).

Vamos abordar na discussão os genes com Kaplan-Meier mais significativos cuja superexpressão foi associada a bom prognóstico (*C5*, *ULK2*, *BECN1* e *ATG14*) e o gene *PLAUR* cuja superexpressão foi associada a mau prognóstico.

O gene *C5* (complemento 5) apresentou um bom prognóstico quando superexpresso em três coortes independentes. O *C5* faz parte da cascata do sistema complemento envolvido no desencadeamento de respostas inflamatórias agudas. Uma vez que esse sistema é ativado, *C5* é clivado em *C5a* e *C5b*. A anafilatoxina *C5a*, recruta principalmente neutrófilos para a área de inflamação e dano tecidual. O *C5b* forma um complexo de ataque a membrana (CAM) juntamente com *C6* e *C7*, *C8* e *C9*, em última instância, induz à lise celular (AFSHAR-KHARGHAN, 2017). Embora estudos apontam que a ativação do sistema complemento no microambiente tumoral pode contribuir para um aumento no crescimento tumoral e metástase (MARKIEWSKI, *et al*, 2008; AFSHAR-KHARGHAN, 2017), os achados em SE mostram o contrário. Savola e colaboradores (2011), analisando dados de expressão de 44 amostras de SE (finlandeses), demonstraram que a alta expressão de *C5* estava correlacionada com ausência de eventos e maior sobrevida global. Além disso os pesquisadores identificaram a expressão do *C5* e do seu receptor *C5aR1* em nível da proteína por imuno-histoquímica em um *tissue-array* de amostras de SE independentes e sugerem que um loop autócrino pode estar ocorrendo, uma vez que o *C5* é um poderoso fator pró-inflamatório e desencadeia a liberação de DAMPs (padrões moleculares associados ao dano) das células-alvo (SAVOLA, *et al*, 2011). Assim, há uma necessidade de se avaliar o impacto da ativação *C5* nas células de SE e as de vias de sinalização a jusante que são acionadas.

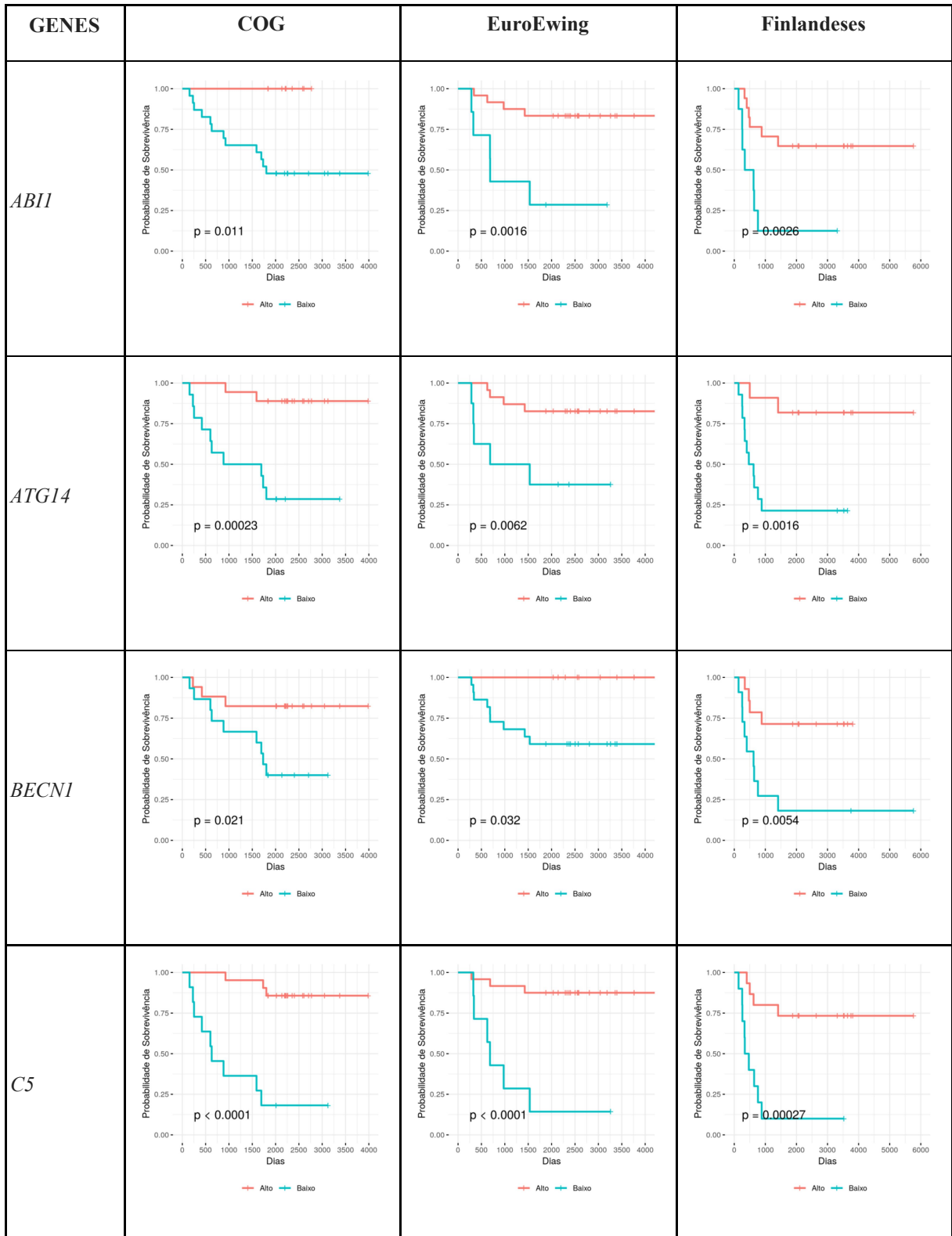
A autofagia é um processo de degradação intracelular que ocorre sob várias condições estressantes, incluindo danos à organela, a presença de proteínas anormais e a privação de nutrientes (MIZUSHIMA, 2007). O mecanismo de autofagia inicia a formação de autofagossomos que capturam os alvos e depois se fundem com os lisossomos para reciclar esses componentes (YUN & LEE, 2018). No câncer, a autofagia desempenha papéis opostos, pode atuar tanto como supressor tumoral, quanto promotor tumoral, dependendo do tipo de tumor (LORIN, *et al*, 2013; WHITE, 2012). Em nível basal a autofagia funciona como um mecanismo de supressão tumoral através da redução de componentes celulares danificados e ajuda na manutenção da homeostase celular (TANG, *et al*, 2010). Por outro lado quando os tumores são expostos a condições extremamente estressantes, incluindo hipóxia e privação de nutrientes, autofagia ajuda as células a superar esses estresses. A autofagia é ativada na parte central dos tumores sólidos, onde as células existem sob condições hipóxicas. Encontramos três genes relacionados diretamente ao processo de autofagia cujo o bom prognóstico estava relacionado à alta expressão gênica. O gene *ULK2* participa do complexo ULK1/2 que em condições nutricionais normais é reprimido pelo complexo mTOR, entretanto quando deixa de ser suprimido, o complexo ULK1/2 migra para um domínio específico do retículo endoplasmático (RE) onde ativa o complexo PI3K que é o responsável pela indução da formação do autofagossomo (MIZUSHIMA & KOMATSU, 2011). Os genes *BECN1* e *ATG14* fazem parte do complexo PI3K e também relacionaram o bom prognóstico com alta expressão gênica. Já foi demonstrado que o gene *BECN1* recruta muitas proteínas envolvidas na maturação e alongamento do autofagossomo (MAIURI, *et al*, 2010). Em um trabalho de revisão, Bortnik e Gorski (2017) elegeram uma lista com os marcadores candidatos da autofagia em câncer. Nesta lista encontramos o *BECN1* e o *ULK1* que é um homólogo muito próximo do *ULK2* com alta homologia no domínio da quinase, assim como no domínio C-terminal, além de possuírem redundância funcional (MCALPINE, *et al*, 2013; BORTNIK & GORSKI, 2017). Com relação a autofagia em SE, recentemente Zhang e colaboradores (2017) verificaram que a translocação EWS-FLI1 regula positivamente a autofagia em células de SE. Além disso os autores verificaram que a expressão de ATG4B foi regulada positivamente pela superexpressão de EWS-FLI1, e o silenciamento de ATG4B inibe a autofagia em células SE. Também verificaram que a apoptose foi inibida em células SE que superexpressam ATG4B e a autofagia potencializada por ATG4B era necessária para a sobrevivência de células SE. Embora alguns genes encontrados na nossa análise relacionado a

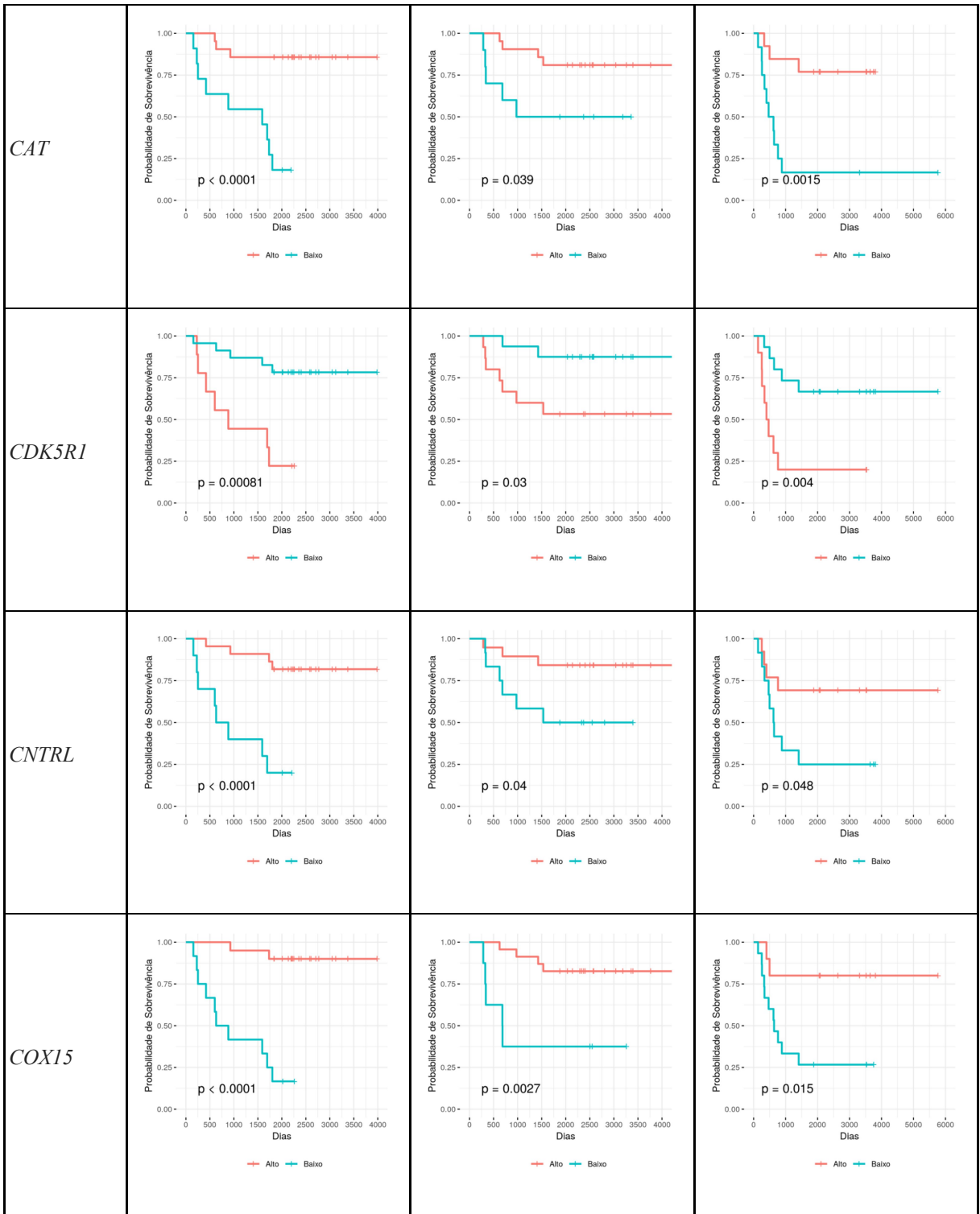
autofagia tiveram bom prognóstico quando superexpressos, precisamos avaliar a via de autofagia nestas entre SOB vs NSOB e tentar entender melhor a relação entre autofagia e SE.

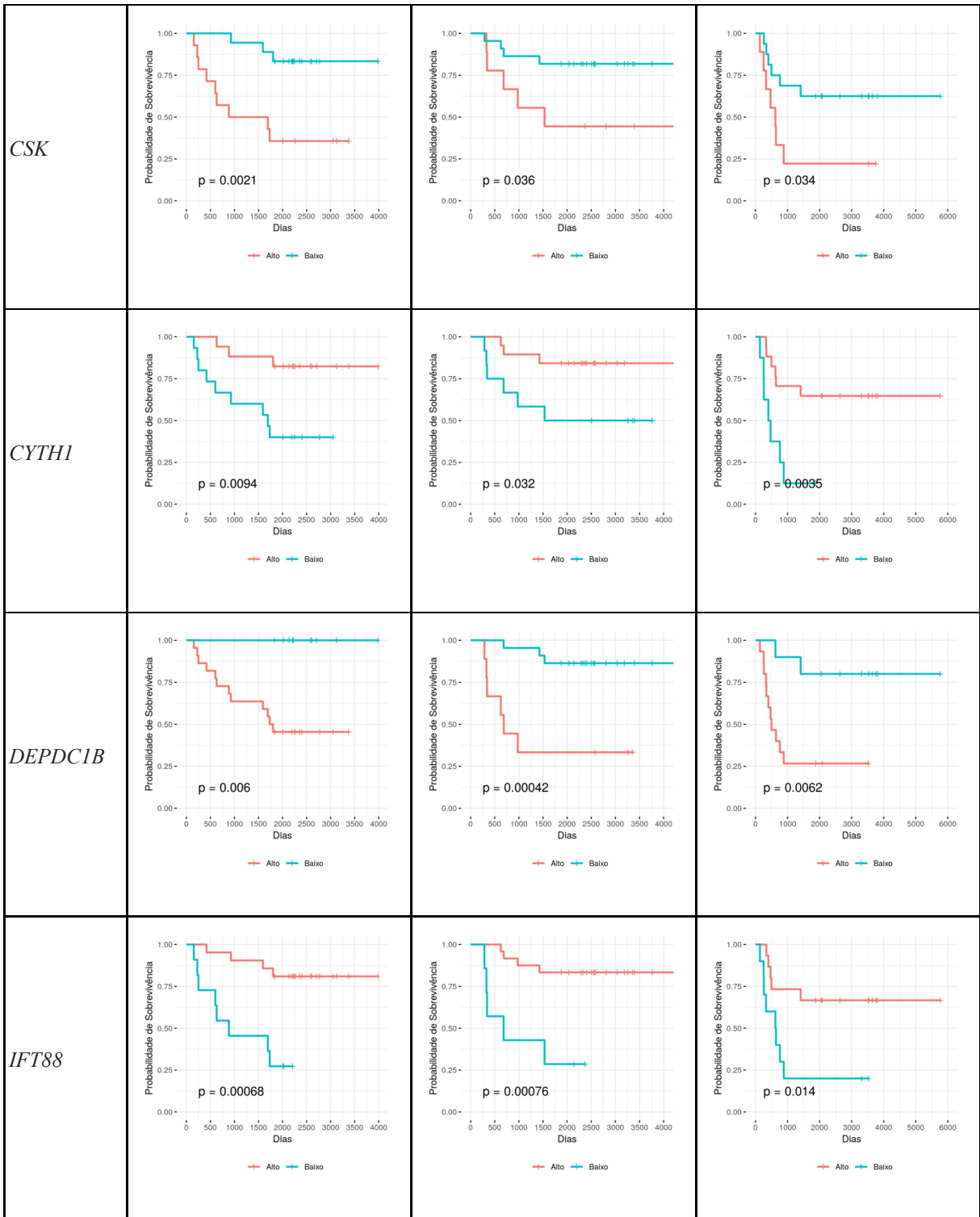
Um dos genes cuja expressão elevada está associado à um mau prognóstico na nossa análise foi o receptor do ativador do plasminogênio tipo uroquinase (*PLAUR*). Esse gene também conhecido como *UPAR*, está envolvido em processos celulares normais e patológicos relacionados com a ativação do plasminogênio na superfície celular e degradação da matriz extracelular. Mais especificamente no caso do tumor tem sido associado com invasão migração e metástase (LIU, *et al*, 2014). A ligação do PLAUR (ativador do plasminogênio tipo uroquinase também conhecido como UPA) ao seu receptor, ativa a conversão do plasminogênio em plasmina, facilitando a degradação da matriz extracelular e a invasão do câncer. De fato, a expressão de *PLAUR* está aumentada em vários tipos de câncer e está correlacionado com invasão precoce e metástase (MAZA, *et al*, 2008). No caso de tumores pediátricos, o RNAm de PLAUR foi identificado em linhagens celulares e amostras de tumores primários de rhabdomyosarcoma (RMS) sendo também identificado em linhagens de SE e algumas amostras de tumor primário de paciente de SE (SANTOS, *et al*, 2007). Recentemente Pilbeam e colaboradores (2018) utilizaram uma toxina bi-específica (EGFATFKDEL) direcionada contra EGFR e PLAUR e testaram a sua sensibilidade na indução de morte celular em duas linhagens de sarcomas; RH30 (linhagem de RMS, expressa EGFR e PLAUR) e TC-71 (linhagem de SE, expressa somente PLAUR), usando testes *in vitro* e modelos murino *in vivo*. Os autores verificaram que os sarcomas foram sensíveis à toxina nos testes *in vitro*, e o crescimento tumoral foi significativamente atenuado nos animais tratados com a toxina, tanto nos ratos RH30 como em TC-71. Além disso os autores verificaram que a resposta antitumoral foi melhor na presença dos dois receptores, linhagem RH30.

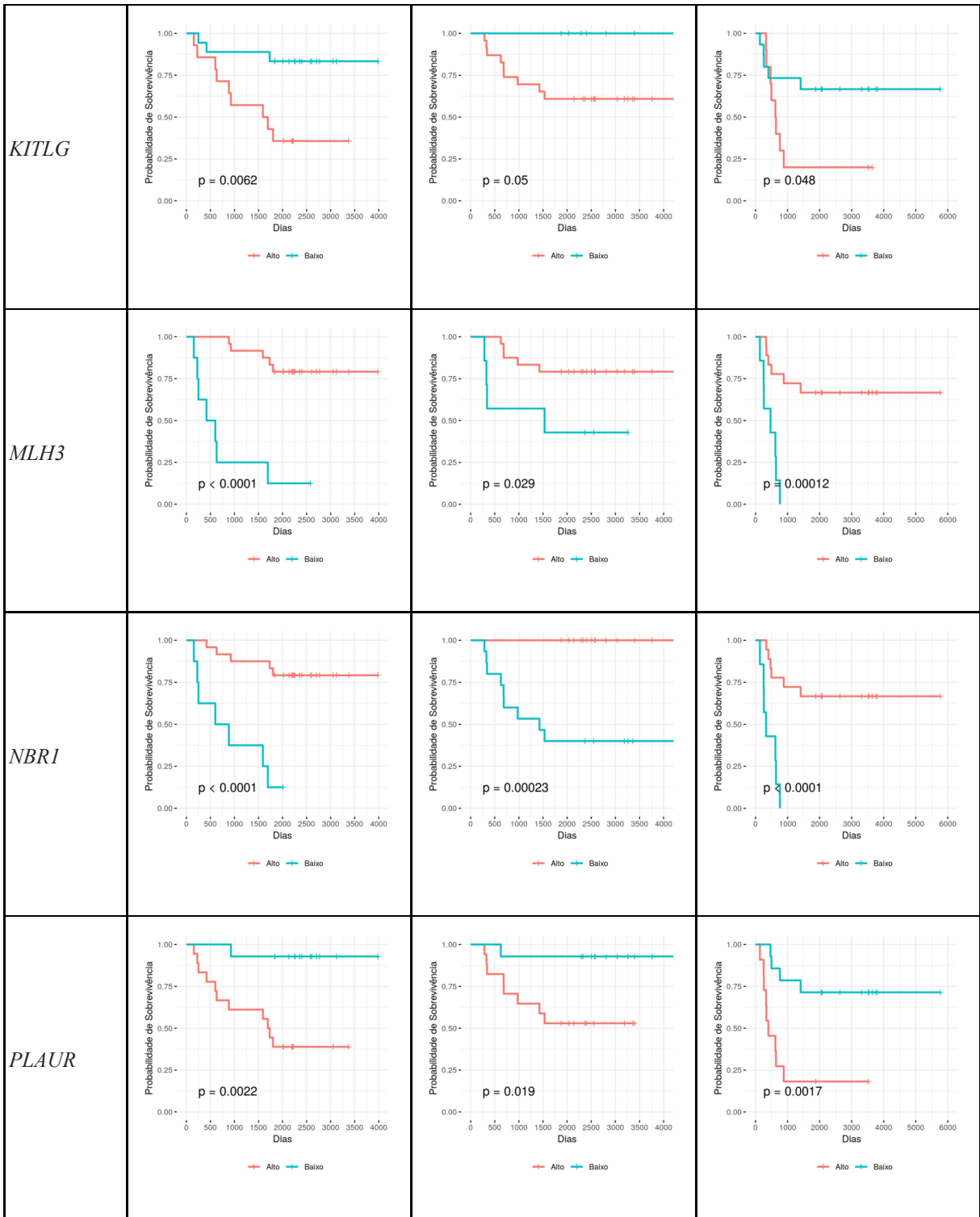
Tabela 5. Relações entre nível de expressão e bom prognóstico dos 23 genes com Kaplan-Meier significativo nas três coortes.

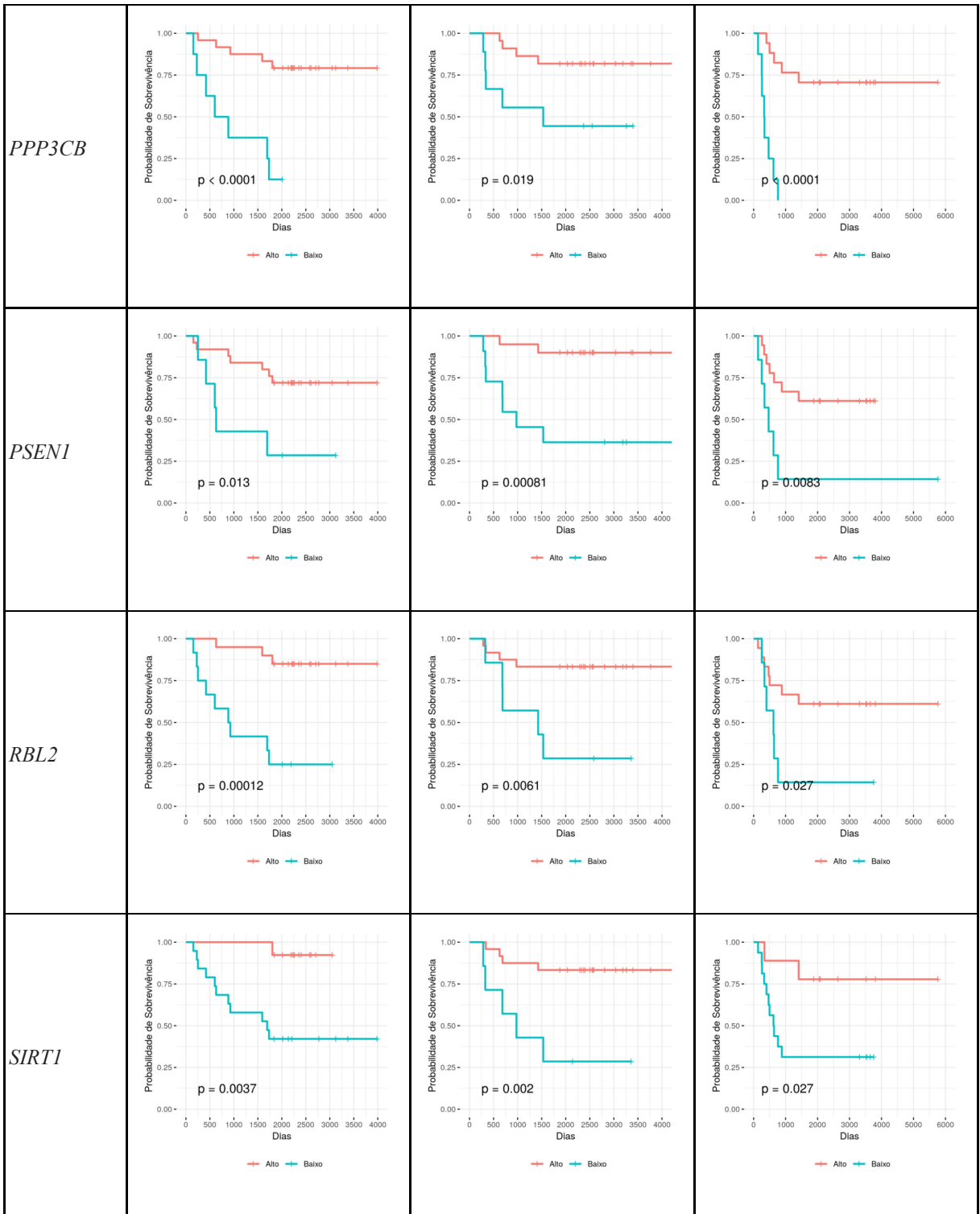
Gene	Nome completo	COG	EuroEwing	Finlandeses
		p-valor	p-valor	p-valor
Expressão alta				
<i>ULK2</i>	<i>Unc-51 Like Autophagy Activating Kinase 2</i>	< 0,0001	0,0008	0,0009
<i>TNKS2</i>	<i>Tankyrase 2</i>	< 0,0001	0,0170	0,0009
<i>SIRT1</i>	<i>Sirtuin 1</i>	0,0037	0,0020	0,0270
<i>PSEN1</i>	<i>Presenilin 1</i>	0,0130	0,0008	0,0083
<i>PPP3CB</i>	<i>Protein Phosphatase 3 Catalytic Subunit Beta</i>	< 0,0001	0,0190	< 0,0001
<i>NBR1</i>	<i>NBR1 Autophagy Cargo Receptor</i>	< 0,001	0,0002	< 0,0001
<i>MLH3</i>	<i>MutL Homolog 3</i>	< 0,0001	0,0290	0,0001
<i>IFT88</i>	<i>Intraflagellar Transport 88</i>	0,0007	0,0008	0,0140
<i>CYTH1</i>	<i>Cytohesin 1</i>	0,0094	0,0320	0,0035
<i>COX15</i>	<i>Cytochrome C Oxidase Assembly Homolog COX15</i>	< 0,0001	0,0027	0,0015
<i>CNTRL</i>	<i>Centriolin</i>	< 0,0001	0,0400	0,0480
<i>RBL2</i>	<i>RB Transcriptional Corepressor Like 2</i>	0,0001	0,0061	0,0270
<i>C5</i>	<i>Complement C5</i>	< 0,0001	< 0,0001	0,0003
<i>BECN1</i>	<i>Beclin 1</i>	0,0210	0,0320	0,0054
<i>ATG14</i>	<i>Autophagy Related 14</i>	0,0002	0,0062	0,0016
<i>ABII</i>	<i>Abl Interactor 1</i>	0,0110	0,0016	0,0026
<i>CAT</i>	<i>Catalase</i>	< 0,0001	0,0390	0,0015
Expressão baixa				
<i>SYK</i>	<i>Spleen Associated Tyrosine Kinase</i>	0,0009	0,0092	0,0250
<i>PLAUR</i>	<i>Plasminogen Activator, Urokinase Receptor</i>	0,0022	0,0190	0,0017
<i>KITLG</i>	<i>KIT Ligand</i>	0,0062	0,0500	0,0480
<i>DEPDC1B</i>	<i>DEP Domain Containing 1B</i>	0,0060	0,0004	0,0062
<i>CSK</i>	<i>C-Terminal Src Kinase</i>	0,0021	0,0360	0,0340
<i>CDK5R1</i>	<i>Cyclin Dependent Kinase 5 Regulatory Subunit 1</i>	0,0008	0,0300	0,0040











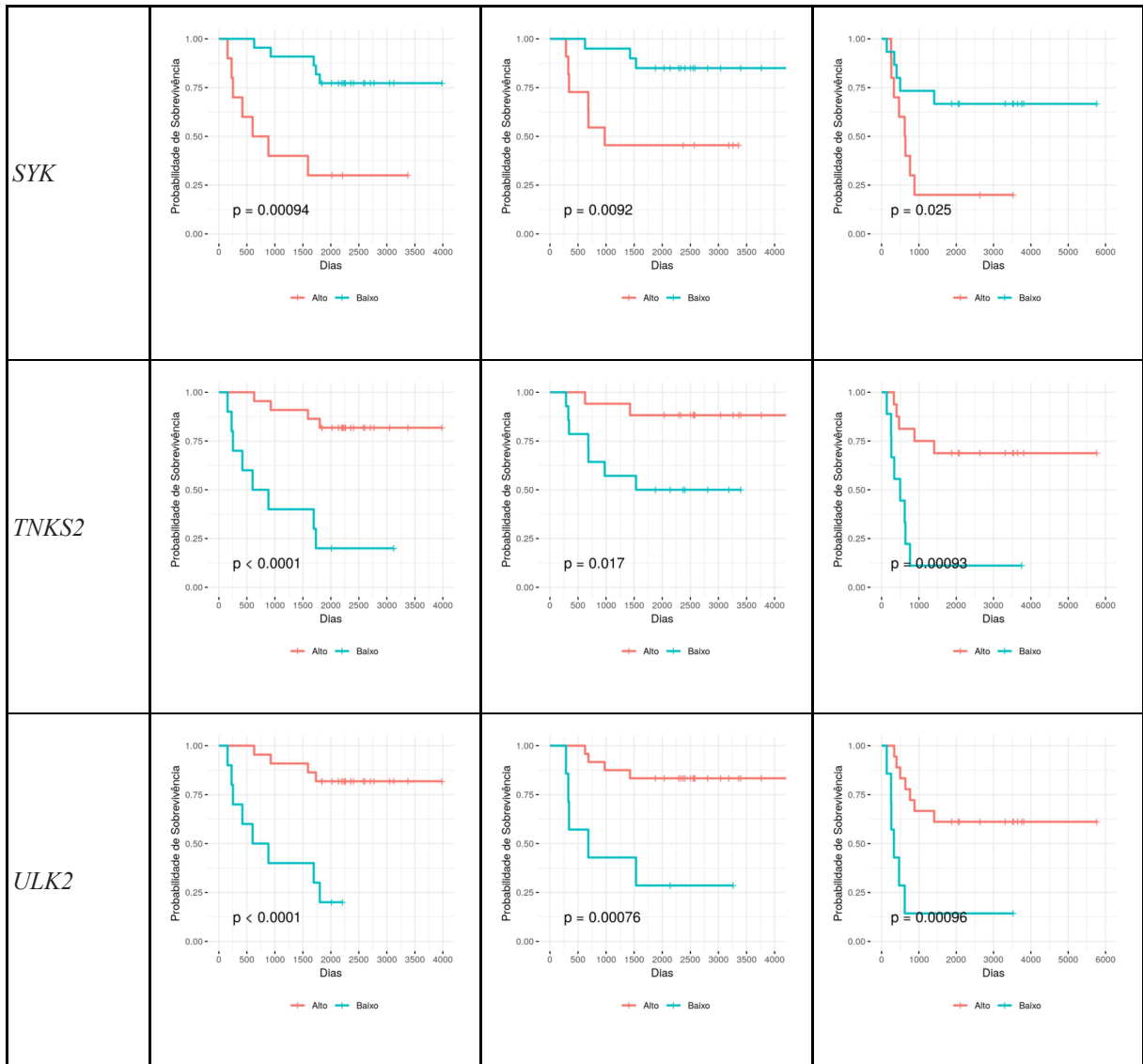


Figura 9. Curvas de Kaplan-Meier para os 23 genes estimadas com dados de sobrevivência e expressão gênica obtidos em cada uma das três coortes (COG, EuroEwing e Finlãndese), totalizando 69 diagramas. Todos os 23 genes apresentaram resultados significantes ($P < 0,05$) e concordantes nas três coortes analisadas.

5 PERSPECTIVAS

Com este estudo, nós conseguimos eleger 23 genes prognósticos validados em duas coortes independentes. Este é o início de um esforço que visa contribuir principalmente para um tratamento mais eficaz dos pacientes acometidos com sarcoma de Ewing. Como complemento ao presente trabalho, iremos utilizar os dados da coorte Finlandeses para selecionar os genes diferencialmente expressos entre SOB e NSOB, da mesma maneira que fizemos com os dados da coorte COG. Após encontrarmos os genes prognósticos nos Finlandeses, iremos validar na coorte EuroEwing e COG. Desta forma poderemos ampliar nossa lista de genes prognósticos, já que algumas regiões de GGDE do Finlandeses não são equivalentes às regiões selecionadas no COG.

Nossos achados foram obtidos através de análises de bioinformática com dados públicos. A nossa perspectiva é validar esses resultados em futuros trabalhos de biologia celular realizados pelo grupo de pesquisa do Instituto do Câncer Infantil, o Laboratório de Câncer e Neurobiologia do HCPA e seus parceiros. Tendo alguns genes devidamente validados como prognósticos, o tratamento dos pacientes acometidos com a doença pode ser modulado conforme a inferência da agressividade do tumor e, conseqüentemente, ser mais eficaz. Teremos a oportunidade de avaliar genes candidatos na população de pacientes sul americanos, onde o Instituto do Câncer Infantil, ao longo dos últimos 15 anos foi responsável pelos protocolos clínicos de tratamento (EWING I e EWING II). Portanto, próximos passos deste trabalho podem impactar diretamente na elaboração de protocolos clínicos futuros desenvolvidos pelo ICI para o tratamento do SE.

6 BIBLIOGRAFIA

- Index of /pub/biosoftwares/transcriptogramer/dictionary/Homosapiens. Disponível em: <<http://lief.if.ufrgs.br/pub/biosoftwares/transcriptogramer/dictionary/Homosapiens/>>.
- KEGG: Kyoto Encyclopedia of Genes and Genomes. Disponível em: <<https://www.genome.jp/kegg/>>.
- AFSHAR-KHARGHAN, V. The role of the complement system in cancer. **Journal of Clinical Investigation**, vol. 127, no. 3, p. 780–789, 2017.
- APPERT-COLLIN, A.; HUBERT, P.; CRÉMEL, G.; *et al.* Role of ErbB Receptors in Cancer Cell Migration and Invasion. **Frontiers in Pharmacology**, vol. 6, 2015.
- ASGHARZADEH, S.; PIQUE-REGI, R.; SPOSTO, R. *et al.* Prognostic Significance of Gene Expression Profiles of Metastatic Neuroblastomas Lacking MYCN Gene Amplification. **JNCI: Journal of the National Cancer Institute**, vol. 98, no. 17, p. 1193–1203, 2006.
- BACCI, G.; FORNI, C.; LONGHI, A.; *et al.* Long-term outcome for patients with non-metastatic Ewing's sarcoma treated with adjuvant and neoadjuvant chemotherapies. 402 patients treated at Rizzoli between 1972 and 1992. **European Journal of Cancer**, vol. 40, no. 1, p. 73–83, 2004.
- BENNANI-BAITI, I. M.; COOPER, A.; LAWLOR, E. R.; *et al.* Intercohort Gene Expression Co-Analysis Reveals Chemokine Receptors as Prognostic Indicators in Ewing's Sarcoma. **Clinical Cancer Research**, vol. 16, no. 14, p. 3769–3778, 2010.
- BORTNIK, S.; and GORSKI, S. M. Clinical Applications of Autophagy Proteins in Cancer: From Potential Targets to Biomarkers. **International Journal of Molecular Sciences**, vol. 18, no. 7, p. 1496, 2017.
- Cancer Facts & Figures 2018**. American Cancer Society. Disponível em: <<https://www.cancer.org/research/cancer-facts-statistics/all-cancer-facts-figures/cancer-facts-figures-2018.html>>.
- CARBON, S.; IRELAND, A.; MUNGALL, C. J.; *et al.* AmiGO: online access to ontology and annotation data. **Bioinformatics**, vol. 25, no. 2, p. 288–289, 2008.
- CARLSON, J.J. and ROTH, J.A.. The impact of the Oncotype Dx breast cancer assay in clinical practice: a systematic review and meta-analysis. **Breast Cancer Research and Treatment**, vol. 141, no. 1, p. 13–22, 2013.
- CARVALHO, B.S. and IRIZARRY, R.A. A framework for oligonucleotide microarray preprocessing. **Bioinformatics**, vol. 26, no. 19, p. 2363–2367, 2010.
- CLEAVER, A.L; BEESLEY, A.H; FIRTH, M.J; *et al.* Gene-based outcome prediction in multiple cohorts of pediatric T-cell acute lymphoblastic leukemia: a Children's Oncology Group study. **Molecular Cancer**, vol. 9, no. 1, p. 105, 2010.
- COTTERILL, S.j.; AHRENS, S.; PAULUSSEN, M.; *et al.* Prognostic Factors in Ewing's Tumor of Bone: Analysis of 975 Patients From the European Intergroup Cooperative

- Ewing's Sarcoma Study Group. **Journal of Clinical Oncology**, vol. 18, no. 17, p. 3108–3114, 2000.
- DAVICIONI, E.; ANDERSON, J.R.; BUCKLEY, J.D.; *et al.* Gene Expression Profiling for Survival Prediction in Pediatric Rhabdomyosarcomas: A Report From the Children's Oncology Group. **Journal of Clinical Oncology**, vol. 28, no. 7, p. 1240–1246, 2010.
- DOYLE, L.A. Sarcoma classification: An update based on the 2013 World Health Organization Classification of Tumors of Soft Tissue and Bone. **Cancer**, vol. 120, no. 12, p. 1763–1774, 2014.
- EDGAR, R. Gene Expression Omnibus: NCBI gene expression and hybridization array data repository. **Nucleic Acids Research**, vol. 30, no. 1, p. 207–210, 2002.
- ERHO, N.; CRISAN, A.; VERGARA, I.A.; *et al.* Discovery and Validation of a Prostate Cancer Genomic Classifier that Predicts Early Metastasis Following Radical Prostatectomy. **PLoS ONE**, vol. 8, no. 6, 2013.
- ESIASHVILI, N.; GOODMAN, M. and MARCUS, R.B. Changes in Incidence and Survival of Ewing Sarcoma Patients Over the Past 3 Decades. **Journal of Pediatric Hematology/Oncology**, vol. 30, no. 6, p. 425–430, 2008.
- EWING, J. Diffuse Endothelioma of Bone. **CA: A Cancer Journal for Clinicians**, vol. 22, no. 2, p. 95–98, 1972.
- FABREGAT, A.; JUPE, S. MATTHEWS, L.; *et al.* The Reactome Pathway Knowledgebase. **Nucleic Acids Research**, vol. 46, no. D1, 2017.
- GASPAR, N.; HAWKINS, D.S.; DIRKSEN, Uta; *et al.* Ewing Sarcoma: Current Management and Future Approaches Through Collaboration. **Journal of Clinical Oncology**, vol. 33, no. 27, p. 3036–3046, 2015.
- GAUTIER, L.; COPE, L.; BOLSTAD, B. M.; *et al.* affy-analysis of Affymetrix GeneChip data at the probe level. **Bioinformatics**, vol. 20, no. 3, p. 307–315, 2004.
- GRIER, H.E.; KRAILO, M.D.; TARBELL, N.J.; *et al.* Addition of Ifosfamide and Etoposide to Standard Chemotherapy for Ewing's Sarcoma and Primitive Neuroectodermal Tumor of Bone. **New England Journal of Medicine**, vol. 348, no. 8, p. 694–701, 2003.
- GRÜNEWALD, T.G.P.; CIDRE-ARANAZ, F.; SURDEZ, D.; *et al.* **Ewing sarcoma**. Nature Reviews Disease Primers, vol. 4, no. 5, p. 1-22, 2018
- HUANG, Da W.; SHERMAN, B.T. and LEMPICKI, R.A. Bioinformatics enrichment tools: paths toward the comprehensive functional analysis of large gene lists. **Nucleic Acids Research**, vol. 37, no. 1, p. 1–13, 2008.
- HUANG, Da W.; SHERMAN, B.T. and LEMPICKI, R.A. Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. **Nature Protocols**, vol. 4, no. 1, p. 44–57, 2008.
- Incid.** Minist. Disponível em: <<http://www1.inca.gov.br/wcm/incidencia/2017/>>.
- Incidência, mortalidade e morbidade hospitalar por câncer em crianças, adolescentes e adultos jovens no Brasil: informações dos registros de câncer e do sistema de**

mortalidade. [s.l.]: INCA, 2016.

- KANEHISA, M. KEGG: Kyoto Encyclopedia of Genes and Genomes. **Nucleic Acids Research**, vol. 28, no. 1, p. 27–30, 2000.
- KANG, H.; CHEN, I.-M.; WILSON, C. S.; *et al.* Gene expression classifiers for relapse-free survival and minimal residual disease improve risk classification and outcome prediction in pediatric B-precursor acute lymphoblastic leukemia. **Blood**, vol. 115, no. 7, p. 1394–1405, 2009.
- KOLB, E. A.; KUSHNER, B. H.; GORLICK, R.; *et al.* Long-Term Event-Free Survival After Intensive Chemotherapy for Ewing’s Family of Tumors in Children and Young Adults. **Journal of Clinical Oncology**, vol. 21, no. 18, p. 3423–3430, 2003.
- KOU, C.J and KANDPAL, R.P. Differential Expression Patterns of Eph Receptors and Ephrin Ligands in Human Cancers. **BioMed Research International**, vol. 2018, p. 1–23, 2018.
- LIU, X.; LAN, Y. ZHANG, D.; *et al.* **SPRY1 promotes the degradation of uPAR and inhibits uPAR-mediated cell adhesion and proliferation.** American journal of cancer research. Disponível em: <<https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4266704/>>.
- LORIN, S.; HAMAĬ, A. MEHRPOUR, M.; *et al.* Autophagy regulation and its role in cancer. **Seminars in Cancer Biology**, vol. 23, no. 5, p. 361–379, 2013.
- LORIN, S.; HAMAĬ, A.; MEHRPOUR, M.; *et al.* Autophagy regulation and its role in cancer. **Seminars in Cancer Biology**, vol. 23, no. 5, p. 361–379, 2013.
- LU, Q.; ZHANG, Y.; MA, L.; *et al.* EWS-FLI1 positively regulates autophagy by increasing ATG4B expression in Ewing sarcoma cells. **International Journal of Molecular Medicine**, vol. 40, no. 4, p. 1217–1225, 2017.
- MAIURI, Maria Chiara; CRIOLLO, Alfredo and KROEMER, Guido. Crosstalk between apoptosis and autophagy within the Beclin 1 interactome. **The EMBO Journal**, vol. 29, no. 3, p. 515–516, 2010.
- MARKIEWSKI, M. M; DEANGELIS, R.A; BENENCIA, F.; *et al.* Modulation of the antitumor immune response by complement. **Nature Immunology**, vol. 9, no. 11, p. 1225–1235, 2008.
- MAZAR, A. P. Urokinase Plasminogen Activator Receptor Choreographs Multiple Ligand Interactions: Implications for Tumor Progression and Therapy. **Clinical Cancer Research**, vol. 14, no. 18, p. 5649–5655, 2008.
- MCALPINE, F.; WILLIAMSON, L.E.; TOOZE, S.A.; *et al.* Regulation of nutrient-sensitive autophagy by uncoordinated 51-like kinases 1 and 2. **Autophagy**, vol. 9, no. 3, p. 361–373, 2013.
- MENDOZA-NARANJO, A.; EL-NAGGAR, A.; WAI, D. H.; *et al.* ERBB4 confers metastatic capacity in Ewing sarcoma. **EMBO Molecular Medicine**, vol. 5, no. 7, p. 1087–1102, 2013.
- MISHRA, R.; HANKER, A.B. and GARRETT, J.T. Genomic alterations of ERBB receptors in cancer: clinical implications. **Oncotarget**, vol. 8, no. 69, 2017.

- MISSIAGLIA, E.; WILLIAMSON, D.; CHISHOLM, J. *et al.* PAX3/FOXO1 Fusion Gene Status Is the Key Prognostic Molecular Marker in Rhabdomyosarcoma and Significantly Improves Current Risk Stratification. **Journal of Clinical Oncology**, vol. 30, no. 14, p. 1670–1677, 2012.
- MIZUSHIMA, N. Autophagy: process and function. **Genes & Development**, vol. 21, no. 22, p. 2861–2873, 2007.
- MIZUSHIMA, N. and KOMATSU, M. Autophagy: Renovation of Cells and Tissues. **Cell**, vol. 147, no. 4, p. 728–741, 2011.
- OBERTHURER, A.; HERO, B.; BERTHOLD, F.; *et al.* Prognostic Impact of Gene Expression–Based Classification for Neuroblastoma. **Journal of Clinical Oncology**, vol. 28, no. 21, p. 3506–3515, 2010.
- OHALI, A.; AVIGAD, S.; ZAIZOV, R.; *et al.* Prediction of high risk Ewing's sarcoma by gene expression profiling. **Oncogene**, vol. 23, no. 55, p. 8997–9006, 2004.
- PAPPO, A. S. and DIRKSEN, U. Rhabdomyosarcoma, Ewing Sarcoma, and Other Round Cell Sarcomas. **Journal of Clinical Oncology**, vol. 36, no. 2, p. 168–179, 2018.
- PASQUALE, E.B. Eph receptors and ephrins in cancer: bidirectional signalling and beyond. **Nature Reviews Cancer**, vol. 10, no. 3, p. 165–180, 2010.
- PATEL, M.; SIMON, J. M.; IGLESIA, M. D.; *et al.* Tumor-specific retargeting of an oncogenic transcription factor chimera results in dysregulation of chromatin and transcription. **Genome Research**, vol. 22, no. 2, p. 259–270, 2011.
- PILBEAM, K.; WANG, H.; TARAS, E.; *et al.* Targeting pediatric sarcoma with a bispecific ligand immunotoxin targeting urokinase and epidermal growth factor receptors. **Oncotarget**, vol. 9, no. 15, 2017.
- RIGGI, N.; KNOECHEL, B.; GILLESPIE, S.M.; *et al.* EWS-FLI1 Utilizes Divergent Chromatin Remodeling Mechanisms to Directly Activate or Repress Enhancer Elements in Ewing Sarcoma. **Cancer Cell**, vol. 26, no. 5, p. 668–681, 2014.
- SANKAR, Savita and LESSNICK, Stephen L. Promiscuous partnerships in Ewing's sarcoma. **Cancer Genetics**, vol. 204, no. 7, p. 351–365, 2011.
- SANTOS, N.D.; HABIBI, G.; WANG, M.; *et al.* Urokinase-type Plasminogen Activator (uPA) is Inhibited with QLT0267 a Small Molecule Targeting Integrin-linked Kinase (ILK). **Translational oncogenomics**. Vol. 23; no. 2, p. 85-97, 2007.
- SAVOLA, S.; KLAMI, A.; TRIPATHI, A.; *et al.* Combined use of expression and CGH arrays pinpoints novel candidate genes in Ewing sarcoma family of tumors. **BMC Cancer**, vol. 9, no. 1, 2009.
- SAVOLA, S.; KLAMI, A.; MYLLYKANGAS, S.; *et al.* High Expression of Complement Component 5 (C5) at Tumor Site Associates with Superior Survival in Ewing's Sarcoma Family of Tumour Patients. **ISRN Oncology**, vol. 2011, p. 1–10, 2011.
- SCHAEFER, K-L.; EISENACHER, M.; BRAUN, Y.; *et al.* Microarray analysis of Ewing's sarcoma family of tumours reveals characteristic gene expression signatures associated with metastasis and resistance to chemotherapy. **European Journal of Cancer**, vol. 44,

no. 5, p. 699–709, 2008.

- SCHAFT, D.W.j. Van Der; HILLEN, F.; PAUWELS, P.; *et al.* Tumor Cell Plasticity in Ewing Sarcoma, an Alternative Circulatory System Stimulated by Hypoxia. **Cancer Research**, vol. 65, no. 24, p. 11520–11528, 2005.
- SCOTLANDI, K.; REMONDINI, D.; CASTELLANI, G.; *et al.* Overcoming Resistance to Conventional Drugs in Ewing Sarcoma and Identification of Molecular Predictors of Outcome. **Journal of Clinical Oncology**, vol. 27, no. 13, p. 2209–2216, 2009.
- SILVA, S.R. Da; PERRONE, G.C; DINIS, J.M; *et al.* Reproducibility enhancement and differential expression of non predefined functional gene sets in human genome. **BMC Genomics**, vol. 15, no. 1, p. 1181, 2014.
- SOMASEKHARAN, S.P.; EL-NAGGAR, A.; LEPRIVIER, G.; *et al.* YB-1 regulates stress granule formation and tumor progression by translationally activating G3BP1. **The Journal of Cell Biology**, vol. 208, no. 7, p. 913–929, 2015.
- STRICKER, T.P.; MADRID, A.M.La; CHLENSKI, A.; *et al.* Validation of a prognostic multi-gene signature in high-risk neuroblastoma using the high throughput digital NanoString nCounter™ system. **Molecular Oncology**, vol. 8, no. 3, p. 669–678, 2014.
- SÁINZ-JASPEADO, M.; HUERTAS-MARTINEZ, J.; LAGARES-TENA, L.; *et al.* EphA2-Induced Angiogenesis in Ewing Sarcoma Cells Works through bFGF Production and Is Dependent on Caveolin-1. **PLoS ONE**, vol. 8, no. 8, 2013.
- TANG, D.; KANG, R.; LIVESEY, K.M.; *et al.* Endogenous HMGB1 regulates autophagy. **The Journal of Experimental Medicine**, vol. 207, no. 10, p. i27–i27, 2010.
- TOOMEY, E C; SCHIFFMAN, J D and LESSNICK, S L. Recent advances in the molecular pathogenesis of Ewing's sarcoma. **Oncogene**, vol. 29, no. 32, p. 4504–4516, 2010.
- TRINQUAND, A.; TANGUY-SCHMIDT, A.; ABDELALI, R.B.; *et al.* Toward a NOTCH1/FBXW7/RAS/PTEN–Based Oncogenetic Risk Classification of Adult T-Cell Acute Lymphoblastic Leukemia: A Group for Research in Adult Acute Lymphoblastic Leukemia Study. **Journal of Clinical Oncology**, vol. 31, no. 34, p. 4333–4342, 2013.
- VOLCHENBOUM, S L.; ANDRADE, J; HUANG, L; *et al.* Gene expression profiling of Ewing sarcoma tumours reveals the prognostic importance of tumour-stromal interactions: a report from the Children's Oncology Group. **The Journal of Pathology: Clinical Research**, vol. 1, no. 2, p. 83–94, 2015.
- WATSON, S; PERRIN, V; GUILLEMOT, D; *et al.* Transcriptomic definition of molecular subgroups of small round cell sarcomas. **The Journal of Pathology**, vol. 245, no. 1, p. 29–40, 2018.
- WHITE, E. Deconvoluting the context-dependent role for autophagy in cancer. **Nature Reviews Cancer**, vol. 12, no. 6, p. 401–410, 2012.
- WILSON, R.A.; TENG, L.; BACHMEYER, K.M.; *et al.* A Novel Algorithm for Simplification of Complex Gene Classifiers in Cancer. **Cancer Research**, vol. 73, no. 18, p. 5625–5632, 2013.
- YIN, X; SUN, J; ZHANG, H; *et al.* Comprehensive analysis of multi Ewing sarcoma

microarray datasets identifies several prognosis biomarkers. **Molecular Medicine Reports**, 2018.

YUN, C W and LEE, S H. The Roles of Autophagy in Cancer. **International journal of molecular sciences**, *19(11)*, 3466., 2018