

**UNIVERSIDADE DE SÃO PAULO**  
**INSTITUTO DE QUÍMICA – DEPARTAMENTO DE BIOQUÍMICA**  
**PROGRAMA INTERUNIDADES DE PÓS-GRADUAÇÃO EM**  
**BIOINFORMÁTICA**

**JHONATAS SIRINO MONTEIRO**

**Investigações sobre os mecanismos moleculares da angiogênese com  
base em dados de expressão gênica num modelo murino, e sua aplicação  
no prognóstico do câncer de mama**

O presente trabalho foi realizado com apoio da Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brasil (CAPES) - Código de Financiamento 001.

Durante o desenvolvimento deste trabalho o autor recebeu auxílio financeiro da FAPESP (Processo 2019/05287-2)

**São Paulo**

**2022**

JHONATAS SIRINO MONTEIRO

**Investigações sobre os mecanismos moleculares da angiogênese com base em dados de expressão gênica num modelo murino, e sua aplicação no prognóstico do câncer de mama**

**Versão Corrigida**

Tese de doutorado apresentada ao Programa Interunidades de Pós-Graduação em Bioinformática da Universidade de São Paulo para obtenção do título de Doutor em Ciências.

Área de concentração: Bioinformática  
Orientador: Prof. Dr. João Carlos Setubal  
Co-orientador: Prof. Dr. Ricardo José Giordano

**São Paulo**

**2022**

Autorizo a reprodução e divulgação total ou parcial deste trabalho, por qualquer meio convencional ou eletrônico, para fins de estudo e pesquisa, desde que citada a fonte.

Ficha catalográfica elaborada com dados inseridos pelo(a) autor(a)  
Biblioteca Carlos Benjamin de Lyra  
Instituto de Matemática e Estatística  
Universidade de São Paulo

---

Monteiro, Jhonatas Sirino  
Investigações sobre os mecanismos moleculares da angiogênese com base em dados de expressão gênica num modelo murino, e sua aplicação no prognóstico do câncer de mama / Jhonatas Sirino Monteiro; orientador, João Carlos Setubal; coorientador, Ricardo José Giordano. - São Paulo, 2022.  
132 p.: il.

Tese (Doutorado) - Programa Interunidades de Pós-Graduação em Bioinformática / Instituto de Matemática e Estatística / Universidade de São Paulo.  
Bibliografia  
Versão corrigida

1. ANGIOGÊNESE PATOLÓGICA ver NEOVASCULARIZAÇÃO PATOLÓGICA. 2. RNA. 3. REGULAÇÃO GÊNICA. 4. APRENDIZADO DE MÁQUINA ver APRENDIZADO COMPUTACIONAL. I. Setubal, João Carlos. II. Título.

---

Bibliotecárias do Serviço de Informação e Biblioteca Carlos Benjamin de Lyra do IME-USP, responsáveis pela estrutura de catalogação da publicação de acordo com a AACR2: Maria Lúcia Ribeiro CRB-8/2766; Stela do Nascimento Madruga CRB 8/7534.

*Dedico a minha esposa Vanessa, a minha avó Genecy, ao meu avô Vieira, a minha mãe Giselle, ao meu sogro Renato e a minha sogra Cleusa*

## AGRADECIMENTOS

Gostaria de agradecer primeiramente a minha esposa Vanessa que esteve ao meu lado durante toda minha trajetória, oferecendo-me conselhos, apoio, motivação (e alguns puxões de orelha). Sem você eu não teria chegado tão longe. Obrigado meu amor.

Quero agradecer também a toda minha família (minha avó Genecy, meu avô Vieira, minha mãe Giselle) que mesmo tão distantes acompanharam de perto todo meu percurso, e estavam sempre dispostos a me oferecer apoio e carinho. Vocês são os pilares da minha vida. Obrigado por tudo.

Agradeço também a minha outra família que me acolheu como filho, que é meu sogro Renato e minha sogra Cleusa que estiveram sempre me oferecendo suporte e carinho. Vocês também representam uma parte muito importante da minha vida.

Aos amigos do “Setulab” que estiveram comigo durante todos esses anos, em especial: Carlos, Suzana, Fernando, Guilherme, Bruno, Lucas, Fábio, Ana, Raquel, Deyvid, Lívia, Ondina, Gianluca, Thibério e Rodrigo, quero agradecer demais pelas risadas, pelos conhecimentos compartilhados e pelos conselhos. Obrigado!

Agradeço ao pesquisador Dr. Emmanuel Dias Neto e ao pesquisador Daniel Gonçalves por terem nos auxiliado na obtenção dos dados do METABRIC.

Agradeço também os meus orientadores João Setubal e Ricardo Giordano que me deram a oportunidade de poder realizar este trabalho. Agradeço também a todo o conhecimento que vocês compartilharam comigo, os quais carregarei comigo para além da minha carreira.

Agradeço a CAPES e a FAPESP (Processo 2019/05287-2) por concederem bolsas de pesquisa, as quais foram fundamentais para execução do presente estudo. Agradeço também ao Programa Interunidades de Pós-Graduação em Bioinformática e a Marlene pelo suporte técnico e administrativo.

*“Por vezes sentimos que aquilo que fazemos não é senão uma gota de água no mar. Mas o mar seria menor se lhe faltasse uma gota”.*

*Madre Teresa de Calcutá*

## RESUMO

A angiogênese é o processo de formação de novos vasos sanguíneos a partir de vasos pré-existentes. A angiogênese fisiológica é bastante restrita durante a fase adulta e ela ocorre sob delicado controle da expressão gênica. Doenças dependentes de angiogênese, como retinopatias e câncer, induzem fatores capazes de alterar a expressão desses genes, promovendo um desenvolvimento desordenado de novos vasos sanguíneos, sendo esse processo caracterizado como angiogênese patológica. Sabe-se que o fator de crescimento endotelial vascular (VEGF) é um dos principais componentes que orquestram a angiogênese e, portanto, esse fator tornou-se um alvo terapêutico para o tratamento de doenças dependentes da angiogênese. Essas terapias são eficazes, mas uma fração importante dos pacientes é refratária ou torna-se resistente à terapia anti-VEGF. Outra dificuldade são os efeitos adversos, uma vez que o VEGF também é importante para a sobrevivência das células endoteliais. Por isso, terapias alternativas mais eficientes e com menos efeitos colaterais são necessárias. Nas últimas décadas, com os avanços das tecnologias genômicas, constatou-se que os RNAs não-codificantes (ncRNAs) têm papel central na regulação gênica, incluindo a angiogênese.

Com isso em mente, o presente trabalho buscou identificar ncRNAs importantes na angiogênese para produzir uma rede integrada de mRNAs e ncRNAs da angiogênese patológica. Foram produzidas duas redes de regulação gênica, onde cada uma representa uma fase da neovascularização: (1) a resposta inicial à hipóxia e (2) a consequente proliferação, migração e diferenciação das células endoteliais durante o processo de neovascularização.

A medicina de precisão é uma área que busca desenvolver terapias direcionadas para cada paciente. Por exemplo, qual paciente responde ou não a determinado medicamento. Utilizando os genes identificados nas duas redes, desenvolvemos 3 assinaturas gênicas de angiogênese com valor prognóstico para pacientes com câncer de mama, outra doença dependente da angiogênese. Paralelamente, foram identificadas 60 variantes diferencialmente expressas e, portanto, possíveis alvos terapêuticos seletivos da angiogênese patológica. Coletivamente, os resultados desses estudos demonstram o potencial das redes gênicas produzidas para o desenvolvimento de novas terapias ou métodos diagnósticos para as doenças dependentes da angiogênese.

**Palavras chaves:** Angiogênese, Retinopatia induzida por oxigênio (OIR), RNA não-codificante (ncRNA), Rede de regulação gênica

## ABSTRACT

Angiogenesis is the process of forming new blood vessels from pre-existing vessels. Physiological angiogenesis is quite restricted during adulthood, and it occurs under the delicate control of gene expression. Angiogenesis-dependent diseases, such as retinopathies and cancer, induce factors capable of altering the expression of these genes, promoting a disordered development of new blood vessels, this process being characterized as pathological angiogenesis. Vascular endothelial growth factor (VEGF) is known to be one of the main components that orchestrate angiogenesis and therefore this factor has become a therapeutic target for the treatment of angiogenesis-dependent diseases. These therapies are effective, but a significant fraction of patients is refractory or resistant to anti-VEGF therapy. Another difficulty is the adverse effects since VEGF is also important for the survival of endothelial cells. Therefore, more efficient alternative therapies with fewer side effects are needed. In recent decades, with advances in genomic technologies, it was found that non-coding RNAs (ncRNAs) play a central role in gene regulation, including angiogenesis.

The present work sought to identify ncRNAs important in angiogenesis to produce an integrated network of pathological angiogenesis mRNAs and ncRNAs. Two gene regulatory networks were produced, each representing a phase of neovascularization: (1) the initial response to hypoxia and (2) the consequent proliferation, migration, and differentiation of endothelial cells during the process of neovascularization.

Precision medicine is an area that seeks to develop targeted therapies for each patient. For example, which patient responds or does not respond to a particular drug. Using the genes identified in the two networks, we developed 3 angiogenesis gene signatures with prognostic value for patients with breast cancer, another angiogenesis-dependent disease. In parallel, 60 differentially expressed isoforms were identified and, therefore, possible selective therapeutic targets of pathological angiogenesis. Collectively, the results of these studies demonstrate the potential of the gene networks produced for the development of new therapies or diagnostic methods for angiogenesis-dependent diseases.

**Keywords:** Angiogenesis, oxygen-induced retinopathy (OIR), non-coding RNA (ncRNA), gene regulatory network (GRN)



## LISTA DE ABREVIATURAS

**AS:** *splicing* alternativo

**BCV:** coeficiente de variação biológica

**BSJ:** *back-spliced junction*

**circRNA:** RNA circular

**CPM:** *counts per million*

**DE:** diferencialmente expresso(a)

**ECM:** Matriz extracelular

**LFC:** Log<sub>2</sub> Fold Change

**lncRNA:** RNA longo não-codificante

**mRNA:** RNA mensageiro

**miRNA:** microRNA

**ncRNA:** RNA não-codificante

**OIR:** Retinopatia induzida por oxigênio

**Poli-A+:** Bibliotecas enriquecidas com RNAs que possuem cauda poli-A

**Ribo-:** Bibliotecas com RNA Total que sofreram tratamento para depleção de RNA ribossômico (rRNA)

**UTR:** Região não traduzida do mRNA

## Lista de Figuras

- FIGURA 1: REPRESENTAÇÃO DO MODELO DE RETINOPATIA INDUZIDA POR OXIGÊNIO (OIR). OS TEMPOS P0, P7, P12 E P17 REPRESENTAM OS DIAS PÓS-NASCIMENTO DO CAMUNDONGO. APÓS O 7<sup>o</sup> DIA DE NASCIMENTO DO CAMUNDONGO, ELE É LEVADO PARA A CÂMARA ONDE FICA EXPOSTO A UM AMBIENTE COM ALTA PRESSÃO DE O<sub>2</sub>. APÓS 5 DIAS, O CAMUNDONGO RETORNA AO AMBIENTE COM CONDIÇÕES DE NORMÓXIA. OS DIAS P12, P15 E P17 SÃO OS DIAS EM QUE FORAM COLETADAS AS RETINAS DO CAMUNDONGO PARA EXTRAÇÃO DE RNA TOTAL. FIGURA RETIRADA DE GUARISCHI-SOUSA ET AL. (2019) - FIGURA 1A. .... 22
- FIGURA 2: COMPARAÇÃO DO ALINHAMENTO DE UM MESMO CONJUNTO DE READS CONTRA O GENOMA DE REFERÊNCIA (C57BL/6J) E O GENOMA MODIFICADO (C57BL/6N) PELA FERRAMENTA RF2M. O ALINHAMENTO FOI REALIZADO COM A FERRAMENTA STAR NO MODO 2-PASS. .... 26
- FIGURA 3: VARIANTES DE SPLICING DO GENE VEGF-A IDENTIFICADAS EM HUMANOS. AS VARIANTES VEGF<sub>xxx</sub>A, QUE POSSUEM O ÉXON 8A SÃO PRÓ-ANGIOGÊNICAS, ENQUANTO AS VARIANTES VEGF<sub>xxx</sub>B, QUE CODIFICAM O ÉXON 8B, SÃO ANTI-ANGIOGÊNICAS, APESAR DE EXISTIR ESTUDOS QUE CONTESTAM SUA EXISTÊNCIA. E POR FIM AS VARIANTES VEGF-AX QUE CODIFICAM OS ÉXONS 8A E 8B, E QUE PODEM ASSUMIR FUNÇÕES PRÓ-ANGIOGÊNICAS OU ANTI-ANGIOGÊNICAS. FIGURA RETIRADA DE PEACH ET AL. (2018) - FIGURA 1. .... 28
- FIGURA 4: HEATMAP DO PERFIL DE EXPRESSÃO DAS 60 VARIANTES DES. A LETRA P REPRESENTA AS AMOSTRAS DO GRUPO CONTROLE, E A LETRA R REPRESENTA AS AMOSTRAS DO GRUPO RETINOPÁTICO. AQUI NOTAMOS QUE AS AMOSTRAS R12\_1 E R12\_2 POSSUEM UM PERFIL SIMILAR ÀS AMOSTRAS DO GRUPO CONTROLE (P12\_1 E P12\_2). COMO AS AMOSTRAS R12 SÃO COLHIDAS LOGO QUANDO OS CAMUNDONGOS RETORNAM AO AMBIENTE COM CONDIÇÕES DE NORMÓXIA, É PROVÁVEL QUE NÃO HOUVE TEMPO O SUFICIENTE PARA OS GENES DE RESPOSTA A HIPÓXIA SEREM ATIVADOS, UMA VEZ QUE O PICO DE HIPÓXIA SOMENTE OCORRE ALGUMAS HORAS DEPOIS (AMOSTRAS R12\_5\_1 E R12\_5\_2). A GRANDE MAIORIA DESSAS VARIANTES COMEÇAM AUMENTAR SUA EXPRESSÃO NO DIA 15 (R15\_1 E R15\_2), E ALGUMAS DELAS AUMENTAM SUA EXPRESSÃO AINDA MAIS NO DIA 17 (R17\_1 E R17\_2). JÁ OUTRAS VARIANTES, COMO AS QUE PERTENCEM AOS GENES VEGFA, SMTLN2, NDUFA4L2, BNIP3 E ADGRG6, POSSUEM UMA MAIOR EXPRESSÃO NO DIA 12\_5 E DEPOIS COMEÇAM A DIMINUIR AO LONGO DO TEMPO, O QUE INDICAM QUE ELAS ESTÃO ENVOLVIDAS COM A FASE DE RESPOSTA À HIPÓXIA DAS CÉLULAS DA RETINA. .... 33
- FIGURA 5: COMPARAÇÃO DAS VARIANTES DO GENE TIMP1. A 5' UTR DA VARIANTE DE (ENSMUST00000009530) É MENOR QUE A DA VARIANTE ENSMUST00000115342. IMAGEM RETIRADA DO ENSEMBL DATABASE..... 35
- FIGURA 6: COMPARAÇÃO DAS VARIANTES DO GENE PTGFR. AMBAS UTRS DA VARIANTE DE (ENSMUST00000029670) É MAIOR QUE A DA VARIANTE ENSMUST00000106126. IMAGEM RETIRADA DO ENSEMBL DATABASE..... 35
- FIGURA 7: COMPARAÇÃO DAS VARIANTES DO GENE MAFF. A 5' UTR DA VARIANTE DE (ENSMUST00000096350) É DIFERENTE QUANDO COMPARADA A DAS VARIANTES ENSMUST00000163691 E ENSMUST00000229130. IMAGEM RETIRADA DO ENSEMBL DATABASE..... 35
- FIGURA 8: COMPARAÇÃO ENTRE A VARIANTE ALTERNATIVA ADGRG6-201 (ENSMUST00000041168), DE NO NOSSO MODELO, E A VARIANTE

- PRINCIPAL ADGRG6-205 (ENSMUST00000238680) DO GENE ADGRG6. A SETA APONTA PARA O ÉXON 24 QUE É EXCLUSIVO DA VARIANTE ADGRG6-201, QUE COMPREENDE UMA REGIÃO DE DESORDEM INTRÍNSECA (IDR). IMAGEM RETIRADA DO ENSEMBL DATABASE..... 36
- FIGURA 9: VARIANTES DO GENE ENG DE CAMUNDONGO. A SETA APONTA PARA O ÉXON 14 DA VARIANTE S-ENG (ENG-202) QUE CORRESPONDE AOS ÉXONS 14 E 15, MAIS O ÍNTRON QUE FICA RETIDO, DA VARIANTE FL-ENG (ENG-201). IMAGEM RETIRADA DO ENSEMBL DATABASE..... 36
- FIGURA 10: VARIANTES DO GENE PIWIL4 DO CAMUNDONGO. A SETA APONTA PARA O ÉXON 11 QUE É EXCLUSIVO DA VARIANTE PRINCIPAL PIWIL4-203. IMAGEM RETIRADA DO ENSEMBL DATABASE. .... 37
- FIGURA 11: VARIANTES DO GENE NES DE CAMUNDONGO. A SETA APONTA PARA O ÉXON 4 DA VARIANTE PRINCIPAL NES-201 QUE CORRESPONDE A JUNÇÃO DOS ÉXONS 4 E 5 DA VARIANTE ALTERNATIVA NES-203. IMAGEM RETIRADA DO ENSEMBL DATABASE..... 37
- FIGURA 12: VARIANTES DO GENE COL2A1 DE CAMUNDONGO. A SETA APONTA PARA ÉXON 2 QUE NÃO SE ENCONTRA NA VARIANTE ALTERNATIVA COL2A1-202, QUE ESTÁ DE, E QUE CONTÉM O DOMÍNIO VWFC. IMAGEM RETIRADA DO ENSEMBL DATABASE..... 38
- FIGURA 13: VARIANTES DO GENE TNFAIP2 DE CAMUNDONGO. A SETA APONTA PARA O ÉXON 6 DA VARIANTE DE TNFAIP2-201, QUE É MENOR QUE O ÉXON PRESENTE NA VARIANTE “IRMÃ” TNFAIP2-202. IMAGEM RETIRADA DO ENSEMBL DATABASE. .... 38
- FIGURA 14: CORRELAÇÃO (PEARSON) ENTRE O PERFIL DE EXPRESSÃO DAS VARIANTES SELECIONADAS COM OS MÉTODOS DE RNA-SEQ E RT-PCR. 39
- FIGURA 15: COMPARAÇÃO DOS LFC OBTIDOS COM RNA-SEQ E RT-PCR PARA AS VARIANTES DOS GENES: A) PIWIL4; B) TNFAIP2; C) COL2A1; D) ENG; E) NES; F) ADGRG6..... 40
- FIGURA 16: RESUMO DO PIPELINE EXECUTADO PARA CONSTRUÇÃO DA REDE DE REGULAÇÃO GÊNICA DA ANGIOGÊNESE PATOLÓGICA..... 45
- FIGURA 17: AS VIAS CANÔNICA E NÃO CANÔNICA DA BIOGÊNESE DE MIRNAS. A VIA CANÔNICA ENVOLVE O PROCESSAMENTO DO PRI-MIRNA PELA ENDONUCLEASE DROSHA/DGCR8, E O PRODUTO DESSA REAÇÃO (PRÉ-MIRNA) É POSTERIORMENTE PROCESSADO NO CITOPLASMA PELA ENDONUCLEASE DICER, GERANDO ASSIM O MIRNA MADURO. NAS VIAS NÃO-CANÔNICAS O MIRNA SERÁ PROCESSADO APENAS POR UMA DAS ENDONUCLEASES CITADAS ANTERIORMENTE. AMBAS AS VIAS RESULTAM NA FORMAÇÃO DO COMPLEXO MIRISC QUE IRÁ ATUAR TANTO NA INIBIÇÃO DA TRADUÇÃO OU NA DEGRADAÇÃO DO MRNA ALVO. FIGURA RETIRADA DE O'BRIEN ET AL. (2018) – FIGURA 1. .... 47
- FIGURA 18: HEATMAP DO PERFIL DE EXPRESSÃO DOS 25 MIRNAS DIFERENCIALMENTE EXPRESSOS COM MAIOR LFC E DOS 25 MIRNAS DIFERENCIALMENTE EXPRESSOS COM MENOR LFC. A LETRA P REPRESENTA AS AMOSTRAS DO GRUPO CONTROLE, E A LETRA R REPRESENTA AS AMOSTRAS DO GRUPO RETINOPÁTICO. É POSSÍVEL OBSERVAR QUE A MAIORIA DOS MIRNAS APRESENTADOS NO HEATMAP TEM SUA EXPRESSÃO REDUZIDA NO DIA 12, QUANDO COMPARAMOS R12 VS P12, E QUE DEPOIS NO DIA 17 ESSE PADRÃO É INVERTIDO. .... 51
- FIGURA 19: VIAS DE BIOGÊNESE DE CIRCNRNAS: A) REGIÕES DE REPETIÇÃO NOS ÍNTRONS FLANQUEADORES OU OS RPBS AUXILIAM NA CIRCULARIZAÇÃO DO RNA APROXIMANDO AS PORÇÕES RECEPTORAS E DOADORAS. B) FORMAÇÃO DO LAÇO (LARIAT) DURANTE SPLICING ALTERNATIVO, PODEM OCORRER O BACKSPLICING DOS ÉXONS INTERNOS OU A CIRCULARIZAÇÃO DO LARIAT, FORMANDO UM CIRCNRNA

INTRÔNICO. FIGURA RETIRADA DE KRISTENSEN ET AL. (2019) – FIGURA 1. .....	55
FIGURA 20: REPRESENTAÇÃO DE UM BJSJ READ DE UM CIRC RNA EXÔNICO. A PORÇÃO 3' DO JUNCTION READ ALINHA COM O ÉXON DA PONTA 5' DO GENE, ENQUANTO A PORÇÃO 5' DO JUNCTION READ ALINHA COM O ÉXON DA PONTA 3' DO GENE. FIGURA BASEADA NA FIGURA 1A PUBLICADA EM GAO ET AL. (2015). .....	57
FIGURA 21: CLASSIFICAÇÃO DOS CIRC RNAs DADA PELO SOFTWARE CIRI2. OS CIRC RNAs EXÔNICOS SÃO AQUELES EM QUE SEUS LIMITES OCORREM DENTRO DA REGIÃO DE UM ÉXON. OS CIRC RNAs INTRÔNICOS SÃO AQUELES EM QUE SEU INÍCIO OU FIM SE ENCONTRA NA REGIÃO DE UM ÍNTRON. OS CIRC RNAs INTERGÊNICOS SÃO AQUELES EM QUE SEUS LIMITES SE ENCONTRAM FORA DOS LIMITES DE UM GENE.....	58
FIGURA 22: RESULTADOS DO BLAST ENTRE OS CIRC RNAs PREDITOS E O BANCO DE DADOS DO CIRCBASE. OS NOMES PIDENT E QCOVER SE REFEREM AO PERCENTUAL DE IDENTIDADE E DE COBERTURA DO ALINHAMENTO, RESPECTIVAMENTE. ....	58
FIGURA 23: HISTOGRAMA DO NÚMERO DE READS BJSJ ALINHADOS NOS CIRC RNAs PREDITOS. A QUANTIFICAÇÃO FOI REALIZADA PELO PRÓPRIO SOFTWARE CIRI2. A GRANDE MAIORIA DOS CIRC RNAs TEM NO MÁXIMO 2 BJSJ READS, IMPOSSIBILITANDO A ANÁLISE DE EXPRESSÃO DIFERENCIAL. .....	59
FIGURA 24: AS 5 CLASSES DE LNC RNAs: SENSO, QUANDO O LNC RNA SE ENCONTRA NA MESMA FITA DO GENE CODIFICANTE DE PROTEÍNA MAIS PRÓXIMO; ANTISENSO, QUANDO SE ENCONTRA NA FITA OPOSTA AO GENE CODIFICANTE DE PROTEÍNA MAIS PRÓXIMO; BIDIRECIONAL, QUANDO SE INICIAM NA REGIÃO PROMOTORA DO GENE CODIFICANTE DE PROTEÍNA; INTRÔNICO, QUANDO SÃO ENCONTRADOS NOS ÍNTRONS DE GENES CODIFICANTES DE PROTEÍNA; E INTERGÊNICO, QUANDO SE ENCONTRAM NA REGIÃO INTERGÊNICA DE DOIS GENES CODIFICANTES DE PROTEÍNA. FIGURA RETIRADA DE KUMAR AND GOYAL (2017) – FIGURA 1. ....	61
FIGURA 25: ANÁLISE DE COMPONENTES PRINCIPAIS (PCA) COM AS AMOSTRAS DAS BIBLIOTECAS RIBO- E POLIA+. É POSSÍVEL VER UMA SEPARAÇÃO CLARA DAS AMOSTRAS RIBO- E POLIA+ COM RELAÇÃO AO COMPONENTE PRINCIPAL 1. É POSSÍVEL OBSERVAR TAMBÉM QUE EXISTE UMA SEGREGAÇÃO ENTRE AS RÉPLICAS BIOLÓGICAS 1 E AS RÉPLICAS BIOLÓGICAS 2 DA BIBLIOTECA RIBO-, O QUE APONTA PARA UM EFEITO DE BATCH INTRODUZIDO PELO SEQUENCIAMENTO DESSAS AMOSTRAS EM LANES DIFERENTES DO ILLUMINA.....	63
FIGURA 26: HEATMAP DO PERFIL DE EXPRESSÃO DE 25 LNC RNAs DE COM MAIOR LFC E 25 LNC RNAs DE COM MENOR LFC. A LETRA P REPRESENTA AS AMOSTRAS DO GRUPO CONTROLE, E A LETRA R REPRESENTA AS AMOSTRAS DO GRUPO RETINOPÁTICO. A GRANDE MAIORIA DOS LNC RNAs COM MAIOR LFC ABSOLUTO SÃO NOVOS CANDIDATOS PREDITOS EM NOSSO PIPELINE. APENAS 3 LNC RNAs PREVIAMENTE DESCRITOS SE ENCONTRAM ENTRE ELES (GM6634-201, GM37799-201 E TMEM250-OS-202) .....	66
FIGURA 27: SELEÇÃO DO VALOR DO SOFT THRESHOLD (B), QUE FOI UTILIZADO PARA GERAR A MATRIZ DE ADJACÊNCIA. A LINHA HORIZONTAL VERMELHA MARCA O COEFICIENTE DE DETERMINAÇÃO ( $R^2$ ) = 0.9 DE UM MODELO DE TOPOLOGIA LIVRE DE ESCALA. O MELHOR VALOR DE B É O MENOR VALOR MAIS PRÓXIMO DA LINHA HORIZONTAL VERMELHA.....	70

- FIGURA 28: TAMANHO, EM NÚMERO DE TRANSCRITOS, DE CADA MÓDULO IDENTIFICADO. NO MÓDULO 0 (COR CINZA) SE ENCONTRAM OS TRANSCRITOS QUE NÃO SE AGRUPARAM EM NENHUM DOS MÓDULOS. .73
- FIGURA 29: HEATMAP DOS VALORES DE CORRELAÇÃO (PEARSON) DOS EIGENGENE DE CADA MÓDULO COM OS GRUPOS DE TRATAMENTO. O EIGENGENE CORRESPONDE A PRIMEIRA COMPONENTE PRINCIPAL DO MÓDULO E REPRESENTA O PERFIL DE EXPRESSÃO DE TODOS OS GENES DO MÓDULO. AS SETAS INDICAM OS MÓDULOS “DARKRED” E “BROWN” QUE FORAM SELECIONADOS PARA AS ETAPAS SEGUINTE DA ANÁLISE. É POSSÍVEL VER QUE O MÓDULO “DARKRED” TEM UMA CORRELAÇÃO POSITIVA E SIGNIFICATIVA (P-VALOR < 0,05) COM AS AMOSTRAS DO GRUPO R12\_5, ENQUANTO O MÓDULO “BROWN” POSSUI UMA CORRELAÇÃO POSITIVA E SIGNIFICATIVA COM AS AMOSTRAS DO GRUPO R17.....75
- FIGURA 30: ÍNDICE DE PRESERVAÇÃO ZSUMMARY: A) ENTRE O CONJUNTO DE REFERÊNCIA (35.331 TRANSCRITOS) E O CONJUNTO 1º QUARTIL (TRANSCRITOS COM VARIÂNCIA MAIOR QUE O PRIMEIRO QUARTIL, UM TOTAL DE 49.618 TRANSCRITOS) E B) ENTRE O CONJUNTO DE REFERÊNCIA E O CONJUNTO 3º QUARTIL (TRANSCRITOS COM VARIÂNCIA MAIOR QUE O TERCEIRO QUARTIL, UM TOTAL DE 17.666 TRANSCRITOS).76
- FIGURA 31: REDE PPI DOS GENES DO MÓDULO “DARKRED” GERADA PELO SOFTWARE STRING.....77
- FIGURA 32: REDE PPI DOS GENES DO MÓDULO “BROWN” GERADA PELO SOFTWARE STRING.....78
- FIGURA 33: REDE DE REGULAÇÃO GÊNICA BASEADA NO MÓDULO “DARKRED”. OS MRNAS ESTÃO REPRESENTADOS PELOS RETÂNGULOS, OS MIRNAS PELOS HEXÁGONOS, OS LNCRNAS PELOS OCTÓGONOS E OS CIRCRNAS PELAS ELIPSES. A COR DO NÓ REPRESENTA A CENTRALIDADE DO RNA NA REDE. QUANTO MAIS QUENTE, MAIOR A CENTRALIDADE DO NÓ, E, PORTANTO, MAIS CAMINHOS (ARESTAS) PASSAM POR ELE. ESTA IMAGEM, ASSIM COMO DAS FIGURA 35FIGURA 36 FORAM GERADAS PELO SOFTWARE CYTOSCAPE.....80
- FIGURA 34: REDE DE REGULAÇÃO GÊNICA BASEADA NO MÓDULO “BROWN”.81
- FIGURA 35: OS 20 RNAs DE MAIOR CENTRALIDADE (BETWEENNESS CENTRALITY) NA GRN DO MÓDULO “BROWN”. OS MRNAS SÃO OS RETÂNGULOS, OS MIRNAS SÃO OS HEXÁGONOS, OS LNCRNAS SÃO OS OCTÓGONOS E OS CIRCRNAS SÃO AS ELIPSES. OS MRNAS ESTÃO REPRESENTADOS PELOS RETÂNGULOS, OS MIRNAS PELOS HEXÁGONOS, OS LNCRNAS PELOS OCTÓGONOS E OS CIRCRNAS PELAS ELIPSES. A COR DO NÓ REPRESENTA A CENTRALIDADE DO RNA NA REDE. QUANTO MAIS QUENTE, MAIOR A CENTRALIDADE DO NÓ, E, PORTANTO, MAIS CAMINHOS (ARESTAS) PASSAM POR ELE.....82
- FIGURA 36: O LNCRNA GM15283 SE ENCONTRA NA FITA OPOSTA AO GENE HIF-1A. IMAGEM RETIRADA DO ENSEMBL (ENMUSG00000087700).....83
- FIGURA 37: FEATURES DA ASSINATURA GÊNICA DO MÓDULO “DARKRED” EM ORDEM DE IMPORTÂNCIA PARA O MODELO. O SUFIXO “\_B” REPRESENTA O VALOR DISCRETO DA EXPRESSÃO DO GENE. A IMPORTÂNCIA DA VARIÁVEL (VIMP) FOI CALCULADA UTILIZANDO O MÉTODO DE BREIMAN-CUTLER VIMP. QUANTO MAIOR O VIMP, MAIOR É A CAPACIDADE PREDITIVA DA VARIÁVEL, ENQUANTO VIMP NEGATIVO, OU IGUAL A ZERO, INDICAM VARIÁVEIS RUIDOSAS (ISHWARAN ET AL., 2021). .....90
- FIGURA 38: CURVA DE KAPLAN-MEIER DA ASSINATURA GÊNICA DO MÓDULO “DARKRED” APLICADA NO CONJUNTO DE VALIDAÇÃO DO METABRIC. O EIXO X REPRESENTA O TEMPO DESDE O DIAGNÓSTICO (EM ANOS), E O EIXO Y REPRESENTA A PROBABILIDADE DE SOBREVIVÊNCIA. A TABELA

- DE RISCO (POSICIONADA ABAIXO DA CURVA DE KAPLAN-MEIER) APRESENTA O NÚMERO DE PACIENTES VIVAS EM CADA UM DOS TEMPOS. AS CATEGORIAS SÃO: RISCO BAIXO, QUANDO PROBABILIDADE DE SOBREVIVÊNCIA (P) > 50% APÓS 15 ANOS DO DIAGNÓSTICO DA DOENÇA; RISCO MÉDIO: QUANDO P > 50% ENTRE 7,5 E 15 ANOS; E RISCO ALTO NOS DEMAIS CASOS. A TESTE DE LOG-RANK MOSTRA QUE A ASSINATURA FOI CAPAZ DE DISTINGUIR AS DIFERENTES CATEGORIAS DE RISCOS DAS PACIENTES. ESTA IMAGEM, ASSIM COMO AS FIGURA 40FIGURA 42FIGURA 43FIGURA 45FIGURA 46FIGURA 48FIGURA 49 FORAM GERADAS PELA FUNÇÃO “GGSURVPLOT” DO PACOTE “SURVMINER”. ..... 91
- FIGURA 39: CURVA DE KAPLAN-MEIER DA ASSINATURA GÊNICA DO MÓDULO “DARKRED” APLICADA PARA CADA UM DOS SUBTIPOS PAM50 DE CÂNCER DE MAMA UTILIZANDO O CONJUNTO DE VALIDAÇÃO DO METABRIC. O EIXO X REPRESENTA O TEMPO DESDE O DIAGNÓSTICO (EM ANOS), E O EIXO Y REPRESENTA A PROBABILIDADE DE SOBREVIVÊNCIA. A TABELA DE RISCO (POSICIONADA ABAIXO DA CURVA DE KAPLAN-MEIER) APRESENTA O NÚMERO DE PACIENTES VIVAS EM CADA UM DOS TEMPOS. AS CATEGORIAS SÃO: RISCO BAIXO, QUANDO PROBABILIDADE DE SOBREVIVÊNCIA (P) > 50% APÓS 15 ANOS DO DIAGNÓSTICO DA DOENÇA; RISCO MÉDIO: QUANDO P > 50% ENTRE 7,5 E 15 ANOS; E RISCO ALTO NOS DEMAIS CASOS. A TESTE DE LOG-RANK MOSTRA QUE A ASSINATURA APENAS TEVE UM BOM DESEMPENHO EM CLASSIFICAR AS PACIENTES DO SUBTIPO LUMINAL (LUMA E LUMB) DO CÂNCER DE MAMA, MAS NÃO FOI EFICAZ NOS SUBTIPOS MAIS AGRESSIVOS (HER2 E BASAL). ..... 92
- FIGURA 40: FEATURES DA ASSINATURA GÊNICA DO MÓDULO “BROWN” EM ORDEM DE IMPORTÂNCIA PARA O MODELO. O SUFIXO “\_B” REPRESENTA O VALOR DISCRETO DA EXPRESSÃO DO GENE. A IMPORTÂNCIA DA VARIÁVEL (VIMP) FOI CALCULADA UTILIZANDO O MÉTODO DE BREIMAN-CUTLER VIMP. QUANTO MAIOR O VIMP, MAIOR É A CAPACIDADE PREDITIVA DA VARIÁVEL, ENQUANTO VIMP NEGATIVO, OU IGUAL A ZERO, INDICAM VARIÁVEIS RUIDOSAS (ISHWARAN ET AL., 2021). ..... 93
- FIGURA 41: CURVA DE KAPLAN-MEIER DA ASSINATURA GÊNICA DO MÓDULO “BROWN” APLICADA NO CONJUNTO DE VALIDAÇÃO DO METABRIC. O EIXO X REPRESENTA O TEMPO DESDE O DIAGNÓSTICO (EM ANOS), E O EIXO Y REPRESENTA A PROBABILIDADE DE SOBREVIVÊNCIA. A TABELA DE RISCO (POSICIONADA ABAIXO DA CURVA DE KAPLAN-MEIER) APRESENTA O NÚMERO DE PACIENTES VIVAS EM CADA UM DOS TEMPOS. AS CATEGORIAS SÃO: RISCO BAIXO, QUANDO PROBABILIDADE DE SOBREVIVÊNCIA (P) > 50% APÓS 15 ANOS DO DIAGNÓSTICO DA DOENÇA; RISCO MÉDIO: QUANDO P > 50% ENTRE 7,5 E 15 ANOS; E RISCO ALTO NOS DEMAIS CASOS. A TESTE DE LOG-RANK MOSTRA QUE A ASSINATURA FOI CAPAZ DE DISTINGUIR AS DIFERENTES CATEGORIAS DE RISCOS DAS PACIENTES. PORÉM, PARA PACIENTES DAS CATEGORIAS DE RISCO MÉDIO E ALTO QUE SOBREVIVERAM 15 OU MAIS, A ASSINATURA NÃO FOI MUITO EFICAZ EM SEPARAR ESSAS CATEGORIAS. ..... 94
- FIGURA 42: CURVA DE KAPLAN-MEIER DA ASSINATURA GÊNICA DO MÓDULO “BROWN” APLICADA PARA CADA UM DOS SUBTIPOS PAM50 DE CÂNCER DE MAMA UTILIZANDO O CONJUNTO DE VALIDAÇÃO DO METABRIC. O EIXO X REPRESENTA O TEMPO DESDE O DIAGNÓSTICO (EM ANOS), E O EIXO Y REPRESENTA A PROBABILIDADE DE SOBREVIVÊNCIA. A TABELA DE RISCO (POSICIONADA ABAIXO DA CURVA DE KAPLAN-MEIER) APRESENTA O NÚMERO DE PACIENTES VIVAS EM CADA UM DOS TEMPOS. AS CATEGORIAS SÃO: RISCO BAIXO, QUANDO PROBABILIDADE DE SOBREVIVÊNCIA (P) > 50% APÓS 15 ANOS DO DIAGNÓSTICO DA DOENÇA; RISCO MÉDIO: QUANDO P > 50% ENTRE 7,5 E 15 ANOS; E RISCO ALTO

- NOS DEMAIS CASOS. A TESTE DE LOG-RANK MOSTRA QUE A ASSINATURA APENAS TEVE UM BOM DESEMPENHO EM CLASSIFICAR AS PACIENTES DO SUBTIPO LUMINAL (LUMA E LUMB) DO CÂNCER DE MAMA, MAS NÃO FOI EFICAZ NOS SUBTIPOS MAIS AGRESSIVOS (HER2 E BASAL). ..... 95
- FIGURA 43: FEATURES DA ASSINATURA GÊNICA DA UNIÃO DOS MÓDULOS “DARKRED” E “BROWN” EM ORDEM DE IMPORTÂNCIA PARA O MODELO. O SUFIXO “\_B” REPRESENTA O VALOR DISCRETO DA EXPRESSÃO DO GENE. A IMPORTÂNCIA DA VARIÁVEL (VIMP) FOI CALCULADA UTILIZANDO O MÉTODO DE BREIMAN-CUTLER VIMP. QUANTO MAIOR O VIMP, MAIOR É A CAPACIDADE PREDITIVA DA VARIÁVEL, ENQUANTO VIMP NEGATIVO, OU IGUAL A ZERO, INDICAM VARIÁVEIS RUIDOSAS (ISHWARAN ET AL., 2021). ..... 96
- FIGURA 44: CURVA DE KAPLAN-MEIER DA ASSINATURA GÊNICA DA UNIÃO DOS MÓDULOS “DARKRED” E “BROWN” APLICADA NO CONJUNTO DE VALIDAÇÃO DO METABRIC. O EIXO X REPRESENTA O TEMPO DESDE O DIAGNÓSTICO (EM ANOS), E O EIXO Y REPRESENTA A PROBABILIDADE DE SOBREVIVÊNCIA. A TABELA DE RISCO (POSICIONADA ABAIXO DA CURVA DE KAPLAN-MEIER) APRESENTA O NÚMERO DE PACIENTES VIVAS EM CADA UM DOS TEMPOS. AS CATEGORIAS SÃO: RISCO BAIXO, QUANDO PROBABILIDADE DE SOBREVIVÊNCIA (P) > 50% APÓS 15 ANOS DO DIAGNÓSTICO DA DOENÇA; RISCO MÉDIO: QUANDO P > 50% ENTRE 7,5 E 15 ANOS; E RISCO ALTO NOS DEMAIS CASOS. A TESTE DE LOG-RANK MOSTRA QUE A ASSINATURA FOI CAPAZ DE DISTINGUIR AS DIFERENTES CATEGORIAS DE RISCOS DAS PACIENTES. PORÉM, PARA PACIENTES DAS CATEGORIAS DE RISCO MÉDIO E ALTO QUE SOBREVIVERAM 15 OU MAIS, A ASSINATURA NÃO FOI MUITO EFICAZ EM SEPARAR ESSAS CATEGORIAS. .... 97
- FIGURA 45: CURVA DE KAPLAN-MEIER DA ASSINATURA GÊNICA DA UNIÃO DOS MÓDULOS “DARKRED” E “BROWN” APLICADA PARA CADA UM DOS SUBTIPOS PAM50 DE CÂNCER DE MAMA UTILIZANDO O CONJUNTO DE VALIDAÇÃO DO METABRIC. O EIXO X REPRESENTA O TEMPO DESDE O DIAGNÓSTICO (EM ANOS), E O EIXO Y REPRESENTA A PROBABILIDADE DE SOBREVIVÊNCIA. A TABELA DE RISCO (POSICIONADA ABAIXO DA CURVA DE KAPLAN-MEIER) APRESENTA O NÚMERO DE PACIENTES VIVAS EM CADA UM DOS TEMPOS. AS CATEGORIAS SÃO: RISCO BAIXO, QUANDO PROBABILIDADE DE SOBREVIVÊNCIA (P) > 50% APÓS 15 ANOS DO DIAGNÓSTICO DA DOENÇA; RISCO MÉDIO: QUANDO P > 50% ENTRE 7,5 E 15 ANOS; E RISCO ALTO NOS DEMAIS CASOS. A TESTE DE LOG-RANK MOSTRA QUE A ASSINATURA APENAS TEVE UM BOM DESEMPENHO EM CLASSIFICAR AS PACIENTES DO SUBTIPO LUMINAL (LUMA E LUMB) DO CÂNCER DE MAMA, MAS NÃO FOI EFICAZ NOS SUBTIPOS MAIS AGRESSIVOS (HER2 E BASAL). ..... 98
- FIGURA 46: FEATURES DA ASSINATURA GÊNICA PROPOSTA POR GUARISCHI-SOUSA ET AL. (2019) EM ORDEM DE IMPORTÂNCIA PARA O MODELO. O SUFIXO “\_B” REPRESENTA O VALOR DISCRETO DA EXPRESSÃO DO GENE, SENDO 0 PARA NÃO-EXPRESSO E 1 PARA EXPRESSO. A IMPORTÂNCIA DA VARIÁVEL (VIMP) FOI CALCULADA UTILIZANDO O MÉTODO DE BREIMAN-CUTLER VIMP. QUANTO MAIOR O VIMP, MAIOR É A CAPACIDADE PREDITIVA DA VARIÁVEL, ENQUANTO VIMP NEGATIVO, OU IGUAL A ZERO, INDICAM VARIÁVEIS RUIDOSAS (ISHWARAN ET AL., 2021). ..... 99
- FIGURA 47: CURVA DE KAPLAN-MEIER DA ASSINATURA GÊNICA PROPOSTA POR GUARISCHI-SOUSA ET AL. (2019) APLICADA NO CONJUNTO DE VALIDAÇÃO DO METABRIC. O EIXO X REPRESENTA O TEMPO DESDE O DIAGNÓSTICO (EM ANOS), E O EIXO Y REPRESENTA A PROBABILIDADE DE

SOBREVIVÊNCIA. A TABELA DE RISCO (POSICIONADA ABAIXO DA CURVA DE KAPLAN-MEIER) APRESENTA O NÚMERO DE PACIENTES VIVAS EM CADA UM DOS TEMPOS. AS CATEGORIAS SÃO: RISCO BAIXO, QUANDO PROBABILIDADE DE SOBREVIVÊNCIA (P) > 50% APÓS 15 ANOS DO DIAGNÓSTICO DA DOENÇA; RISCO MÉDIO: QUANDO P > 50% ENTRE 7,5 E 15 ANOS; E RISCO ALTO NOS DEMAIS CASOS. A TESTE DE LOG-RANK MOSTRA QUE A ASSINATURA FOI CAPAZ DE DISTINGUIR AS DIFERENTES CATEGORIAS DE RISCOS DAS PACIENTES..... 100

FIGURA 48: CURVA DE KAPLAN-MEIER DA ASSINATURA GÊNICA PROPOSTA POR GUARISCHI-SOUSA ET AL. (2019) APLICADA PARA CADA UM DOS SUBTIPOS PAM50 DE CÂNCER DE MAMA UTILIZANDO O CONJUNTO DE VALIDAÇÃO DO METABRIC. O EIXO X REPRESENTA O TEMPO DESDE O DIAGNÓSTICO (EM ANOS), E O EIXO Y REPRESENTA A PROBABILIDADE DE SOBREVIVÊNCIA. A TABELA DE RISCO (POSICIONADA ABAIXO DA CURVA DE KAPLAN-MEIER) APRESENTA O NÚMERO DE PACIENTES VIVAS EM CADA UM DOS TEMPOS. AS CATEGORIAS SÃO: RISCO BAIXO, QUANDO PROBABILIDADE DE SOBREVIVÊNCIA (P) > 50% APÓS 15 ANOS DO DIAGNÓSTICO DA DOENÇA; RISCO MÉDIO: QUANDO P > 50% ENTRE 7,5 E 15 ANOS; E RISCO ALTO NOS DEMAIS CASOS. A TESTE DE LOG-RANK MOSTRA QUE A ASSINATURA TEVE UM BOM DESEMPENHO EM CLASSIFICAR AS PACIENTES DO SUBTIPO LUMINAL (LUMA E LUMB) E BASAL, MAS NÃO FOI EFICAZ PARA O SUBTIPO HER2. .... 101

FIGURA 49: TAXA DE ERRO DE PREDIÇÃO DE CADA UMA DAS ASSINATURAS APRESENTADAS NESSE ESTUDO. TODAS AS ASSINATURAS APRESENTAM TAXAS DE ERRO MUITO SIMILARES, COM UMA DIFERENÇA INFERIOR A 1%. OBSERVA-SE TAMBÉM QUE A TAXA DE ERRO ATINGIU UMA ESTABILIDADE APÓS 375 ÁRVORES GERADAS, O QUE INDICA QUE O AUMENTO DO NÚMERO DE ÁRVORES GERADAS NÃO TRARIA NENHUMA MELHORIA PARA AS ASSINATURAS..... 102



## Lista de Tabelas

TABELA 1: FATORES ANGIOGÊNICOS, SEUS RECEPTORES, E AS VIAS ALTERNATIVAS NÃO DEPENDENTES DE VEGF QUE ESSES FATORES REGULAM. DADOS DESSA TABELA FORAM RETIRADOS DA FIGURA 2 (FIGURA À DIREITA) PUBLICADA EM ITATANI ET AL., 2018. ....	21
TABELA 2: NÚMERO DE READS SEQUENCIADOS PARA AS BIBLIOTECAS POLI-A+ E A PORCENTAGEM DE READS ALINHADAS CONTRA O GENOMA PARA CADA RÉPLICA BIOLÓGICA DE CADA GRUPO/DIA. AS BIBLIOTECAS REPRESENTADAS PELA LETRA P SE REFEREM AS AMOSTRAS DO GRUPO CONTROLE, ENQUANTO AS BIBLIOTECAS REPRESENTADAS PELA LETRA R SE REFEREM AS AMOSTRAS DO GRUPO COM RETINOPATIA. PARA OS TEMPOS 12, 12.5 E 17, OS READS OBTIDOS PARA CADA RÉPLICA TÉCNICA FORAM SOMADOS. NÃO EXISTEM RÉPLICAS TÉCNICAS PARA O DIA 15. O ALINHAMENTO FOI REALIZADO COM O SOFTWARE STAR NO MODO 2-PASS. ....	24
TABELA 3: NÚMERO DE READS SEQUENCIADOS PARA AS BIBLIOTECAS RIBO-E A PORCENTAGEM DE READS ALINHADOS CONTRA O GENOMA PARA CADA RÉPLICA BIOLÓGICA DE CADA GRUPO/DIA. AS BIBLIOTECAS REPRESENTADAS PELA LETRA P SE REFEREM AS AMOSTRAS DO GRUPO CONTROLE, ENQUANTO AS BIBLIOTECAS REPRESENTADAS PELA LETRA R SE REFEREM AS AMOSTRAS DO GRUPO COM RETINOPATIA. O ALINHAMENTO FOI REALIZADO COM O SOFTWARE STAR NO MODO 2-PASS. ....	25
TABELA 4: NÚMERO DE READS SEQUENCIADOS PARA AS BIBLIOTECAS MIRNA E A PORCENTAGEM DE READS ALINHADOS CONTRA O GENOMA PARA CADA AMOSTRA DE CADA GRUPO/DIA. AS BIBLIOTECAS REPRESENTADAS PELA LETRA P SE REFEREM AS AMOSTRAS DO GRUPO CONTROLE, ENQUANTO AS BIBLIOTECAS REPRESENTADAS PELA LETRA R SE REFEREM AS AMOSTRAS DO GRUPO COM RETINOPATIA. O ALINHAMENTO FOI REALIZADO COM O SOFTWARE BOWTIE. ....	25
TABELA 5: SEQUÊNCIAS DOS PRIMERS DAS VARIANTES DES DE INTERESSE PARA VALIDAÇÃO. ....	31
TABELA 6: TOTAL DE VARIANTES COM CANDIDATOS HUMANOS HOMÓLOGOS EM CADA UMA DAS CATEGORIAS: EVIDÊNCIA FORTE, TRANSCRITOS COM UM HIT ÚNICO COM BLASTN; EVIDÊNCIA MODERADA, 2 OU MAIS RESULTADOS CONCORDANTES PROVENIENTES DE DIFERENTES LINHAS DE EVIDÊNCIAS (FERRAMENTAS E/OU LITERATURA); EVIDÊNCIA FRACA, APENAS UM RESULTADO DE UMA ÚNICA LINHA DE EVIDÊNCIA; E SEM EVIDÊNCIA. ....	34
TABELA 7: VARIANTES DES QUE POSSUEM O MESMO PRODUTO PROTEICO DE PELO MENOS UMA DE SUAS VARIANTES “IRMÃS”. ....	35
TABELA 8: TOTAL DE MIRNAS DE EM CADA UM DOS DIAS EM QUE O EXPERIMENTO FOI CONDUZIDO. ....	50
TABELA 9: OS 10 TERMOS GO (ONTOLOGIA “BIOLOGICAL PROCESS”) MAIS ENRIQUECIDOS PARA OS GENES ALVOS DOS MIRNAS DIFERENCIALMENTE EXPRESSOS. A ANÁLISE DE ENRIQUECIMENTO FOI REALIZADA COM A FERRAMENTA WEB DAVID, E O MÉTODO DE CORREÇÃO DO P-VALOR UTILIZADO FOI A DE BENJAMINI-HOCHBERG. ...	52
TABELA 10: AS 10 VIAS DO KEGG MAIS ENRIQUECIDAS PARA OS GENES ALVOS DOS MIRNAS DIFERENCIALMENTE EXPRESSOS. A ANÁLISE DE ENRIQUECIMENTO FOI REALIZADA COM A FERRAMENTA WEB DAVID, E O	

MÉTODO DE CORREÇÃO DO P-VALOR UTILIZADO FOI A DE BENJAMINI-HOCHBERG.....	53
TABELA 11: ALGUNS DOS CIRCNRNAS PREDITOS E SEUS RESPECTIVOS MIRNAS ALVOS.....	59
TABELA 12: RESUMO DOS RESULTADOS OBTIDOS COM O PIPELINE PARA DETECÇÃO DE POTENCIAIS NOVOS LNCNRNAS.....	64
TABELA 13: TOTAL DE LNCNRNAS DE EM CADA UM DOS DIAS EM QUE O EXPERIMENTO FOI CONDUZIDO. ....	65
TABELA 14: OS 5 TERMOS GO (ONTOLOGIA "BIOLOGICAL PROCESS") MAIS ENRIQUECIDOS ENTRE OS GENES DO MÓDULO "DARKRED". A ANÁLISE DE ENRIQUECIMENTO FOI REALIZADA COM A FERRAMENTA WEB DAVID, E O MÉTODO DE CORREÇÃO DO P-VALOR UTILIZADO FOI A DE BENJAMINI-HOCHBERG.....	74
TABELA 15: OS 5 TERMOS GO (ONTOLOGIA "BIOLOGICAL PROCESS") MAIS ENRIQUECIDOS ENTRE OS GENES DO MÓDULO "BROWN". A ANÁLISE DE ENRIQUECIMENTO FOI REALIZADA COM A FERRAMENTA WEB DAVID, E O MÉTODO DE CORREÇÃO DO P-VALOR UTILIZADO FOI A DE BENJAMINI-HOCHBERG.....	74
TABELA 16: OS 10 GENES MAIS SIGNIFICATIVOS (EM ORDEM DECRESCENTE) DO MÓDULO "DARKRED".....	78
TABELA 17: OS 10 GENES MAIS SIGNIFICATIVOS (EM ORDEM DECRESCENTE) DO MÓDULO "BROWN".....	79
TABELA 18: RESUMO COM O TOTAL DE MIRNAS DE (CONHECIDOS E NOVOS) INTERAGINDO COM GENES VALIDADOS DOS MÓDULOS "DARKRED" E "BROWN". ....	79
TABELA 19: RESUMO COM O TOTAL DE CIRCNRNAS PREDITOS PARA OS GENES VALIDADOS NOS MÓDULOS "DARKRED" E "BROWN".....	79
TABELA 20: RESUMO COM O TOTAL DE LNCNRNAS CO-EXPRESSOS NOS MÓDULOS "DARKRED" E "BROWN".....	80
TABELA 21: DADOS CLÍNICOS DAS PACIENTES DO BANCO DE DADOS DO METABRIC NAS COORTES DE TREINAMENTO E VALIDAÇÃO. NA = NÃO DISPONÍVEL; * = MEDIANA (QUARTIL INFERIOR, QUARTIL SUPERIOR). ....	89

## Sumário

1	Introdução e objetivo do estudo .....	20
1.1	Angiogênese.....	20
1.2	Modelo murino de retinopatia induzida por oxigênio (OIR).....	21
1.3	Objetivo .....	23
2	Dados de sequenciamento de RNA e o genoma de referência .....	24
2.1	Conjuntos de dados.....	24
2.2	Obtendo o genoma de referência do <i>Mus musculus</i> da linhagem C57BL/6N	25
3	Análise das variantes de <i>splicing</i> diferencialmente expressas durante angiogênese patológica .....	27
3.1	Métodos.....	29
3.1.1	Quantificação e análise de expressão diferencial.....	29
3.1.2	Busca por homólogos humanos para as variantes de <i>splicing</i> diferencialmente expressas .....	29
3.1.3	Comparação estrutural das variantes diferencialmente expressas .....	30
3.1.4	Validação das variantes de interesse por RT-PCR.....	31
3.2	Resultados.....	31
3.2.1	Expressão diferencial das variantes .....	31
3.2.2	Homólogos humanos para as variantes de <i>splicing</i> diferencialmente expressas	34
3.2.3	Comparação estrutural das variantes diferencialmente expressas .....	34
3.2.4	Validação por RT-PCR.....	38
3.3	Discussão.....	40
4	Análise dos ncRNAs diferencialmente expressos durante angiogênese patológica .....	45
4.1	miRNAs .....	46
4.1.1	Métodos .....	48
4.1.2	Expressão diferencial dos miRNAs .....	49
4.1.3	Predição dos alvos.....	52
4.1.4	Discussão .....	53
4.2	circRNAs .....	54
4.2.1	Métodos .....	57
4.2.2	circRNAs preditos .....	57
4.2.3	Expressão diferencial dos circRNAs.....	58
4.2.4	Predição de sítios de ligação de miRNAs.....	59
4.2.5	Discussão .....	60
4.3	lncRNAs .....	60
4.3.1	Métodos .....	62
4.3.2	Detecção dos lncRNAs .....	64

4.3.3	Expressão diferencial dos lncRNAs.....	64
4.3.4	Discussão .....	67
5	Rede de regulação gênica da angiogênese patológica.....	68
5.1	Métodos.....	69
5.1.1	Identificação dos módulos de coexpressão de genes.....	69
5.1.2	Análise de preservação dos módulos de interesse.....	71
5.1.3	Seleção dos genes mais relevantes para angiogênese.....	71
5.1.4	Construção da rede de regulação gênica para angiogênese.....	72
5.2	Resultados.....	72
5.2.1	Identificação dos módulos de coexpressão de genes.....	72
5.2.2	Análise de preservação dos módulos de interesse.....	76
5.2.3	Seleção dos genes mais relevantes para angiogênese.....	77
5.2.4	Construção da rede de regulação gênica .....	79
5.3	Discussão.....	82
6	Assinatura gênica de angiogênese para prognóstico de câncer de mama ....	87
6.1	Métodos.....	88
6.2	Resultados.....	90
6.2.1	Módulo “Darkred” .....	90
6.2.2	Módulo “Brown” .....	92
6.2.3	Módulos “Darkred” + “Brown” .....	95
6.2.4	Comparação com a assinatura gênica para angiogênese proposta em Guarischi-Sousa et al. (2019) .....	98
6.3	Discussão.....	102
7	Conclusão.....	104
	Referências.....	106
	Apêndice I: Artigos publicados ao longo do doutorado .....	115
	Apêndice II: Lista de variantes diferencialmente expressas (DE) .....	116
	Apêndice III: Lista de candidatos homólogos para as variantes diferencialmente expressas .....	118
	Apêndice IV: Lista de miRNAs diferencialmente expressos (DE) .....	119
	Apêndice V: Lista de lncRNAs diferencialmente expressos (DE) .....	123
	Apêndice VI: Lista de RNAs que pertencem a rede “darkred” .....	129
	Apêndice VII: Lista de RNAs que pertencem a rede “brown” .....	130

# 1 Introdução e objetivo do estudo

## 1.1 Angiogênese

Angiogênese é o nome que se dá ao processo de formação de novos vasos sanguíneos a partir de outros já existentes (BETZ et al., 2016; FOLKMAN, 2007). Em condições normais durante a fase adulta, esse processo é restrito ao ciclo reprodutivo feminino, placentação e cicatrização, sendo realizado sob um fino controle da expressão gênica. Doenças dependentes de angiogênese, como retinopatias (CRAWFORD et al., 2009; HARTNETT; PENN, 2012) e câncer (POTENTE; GERHARDT; CARMELIET, 2011) induzem fatores capazes de alterar a expressão desses genes, promovendo um desenvolvimento desordenado de novos vasos sanguíneos, sendo esse processo caracterizado como angiogênese patológica (SEAMAN et al., 2007).

O fator induzido por hipóxia (*HIF-1*) é um dos principais reguladores da angiogênese. Durante hipóxia, a subunidade alfa do *HIF-1* (*HIF1A*) é levada para o núcleo onde forma um complexo transcricional com a subunidade beta (*HIF1B*) e outras proteínas acessórias, que por sua vez ativa a transcrição de genes que respondem ao estado de hipóxia, como o fator de crescimento endotelial vascular (*VEGF*). *VEGF* e outros fatores se acumulam nas células em hipóxia formando um gradiente extracelular, e se ligam a proteínas com receptores tirosina quinase (ex.: *VEGFRs*), ou receptores de neuropilina, na superfície de células endoteliais, induzindo transfosforilação nesses receptores, que por sua vez ativam diversas cascatas de sinalização intracelular, promovendo a proliferação e diferenciação de células endoteliais (GIORDANO; PASQUALINI; ARAP, 2010).

Sendo o *VEGF* um importante fator para a angiogênese, alguns fármacos foram desenvolvidos com intuito de inibir sua ação e de seus receptores, como bevacizumab, sunitinib, aflibercept, ranibizumab e outros (ITATANI et al., 2018; ZHAO; SINGH, 2018). No entanto, o *VEGF* também atua como fator de sobrevivência dos vasos sanguíneos normais, e a inibição desse fator em tecidos normais pode estar relacionado com os efeitos colaterais observados em uma fração dos pacientes (KAMBA; MCDONALD, 2007). Além disso, também foi observado que alguns pacientes adquirem resistência a essas drogas (ITATANI et al., 2018). Essa resistência pode ser devida a existência de vias alternativas não dependentes da sinalização do *VEGF* (Tabela 1) (ITATANI et al., 2018). Esses fatos demonstram a importância de se entender melhor esse complexo processo biológico.

Tabela 1: Fatores angiogênicos, seus receptores, e as vias alternativas não dependentes de VEGF que esses fatores regulam. Dados dessa tabela foram retirados da Figura 2 (figura à direita) publicada em ITATANI et al., 2018.

<b>Ligantes</b>	<b>Receptores</b>	<b>Vias</b>
Ang2	Tie2	- Ras/Raf/MEK/ERK; - Pi3K/Akt
Bv8	PROKR2	- Ras/Raf/MEK/ERK - Ras/Raf/MEK/ERK;
bFGF	FGFR	- Pi3K/Akt; - JAK/STAT - PLCy/DAG/PKC - NF-kB
IL-1	IL-1RI; IL-1RAcP	- JNK - P38/MAPK
PDGF	PDGFR	- Pi3K/Akt; - PLCt/DAG/PKC
PIGF	VEGFR1; NRP1/2	- Pi3K/MAPK; - P38/MAPK
TGF- $\beta$ 1	TGF- $\beta$ RII; TGF- $\beta$ RI	- Smad

## 1.2 Modelo murino de retinopatia induzida por oxigênio (OIR)

Estudar a formação de novos vasos sanguíneos é um desafio. Modelos celulares são incompletos, pois não reproduzem as intrincadas interações que ocorrem *in vivo* entre endotélio, parênquima e matriz extracelular (ECM). Outro desafio é visualizar as células vasculares e distinguir os vasos sanguíneos dentro de um tecido. Por isso, a retina, sendo um tecido transparente e de fácil acesso é um excelente modelo de estudos. Em particular, o modelo OIR é um dos que melhor reproduzem a angiogênese *in vivo*. O modelo consiste em manter camundongo em câmaras com alta pressão de oxigênio (hiperóxia, 75% de O<sub>2</sub>) do 7º ao 12º dia após seu nascimento. Durante esse período, a alta pressão de O<sub>2</sub> reduz a produção natural de VEGF, inibindo também a neovascularização na retina. No 12º dia, o jovem camundongo retorna as condições de normóxia do biotério, o que paradoxalmente, leva a uma condição de hipóxia no tecido, pois os poucos vasos existentes não conseguem suprir a demanda das células fotorreceptoras, levando a um aumento exacerbado da expressão do VEGF e o consequente crescimento descontrolado de vasos sanguíneos. Esse processo de neovascularização patológica alcança seu pico no 17º dia (Figura 1).

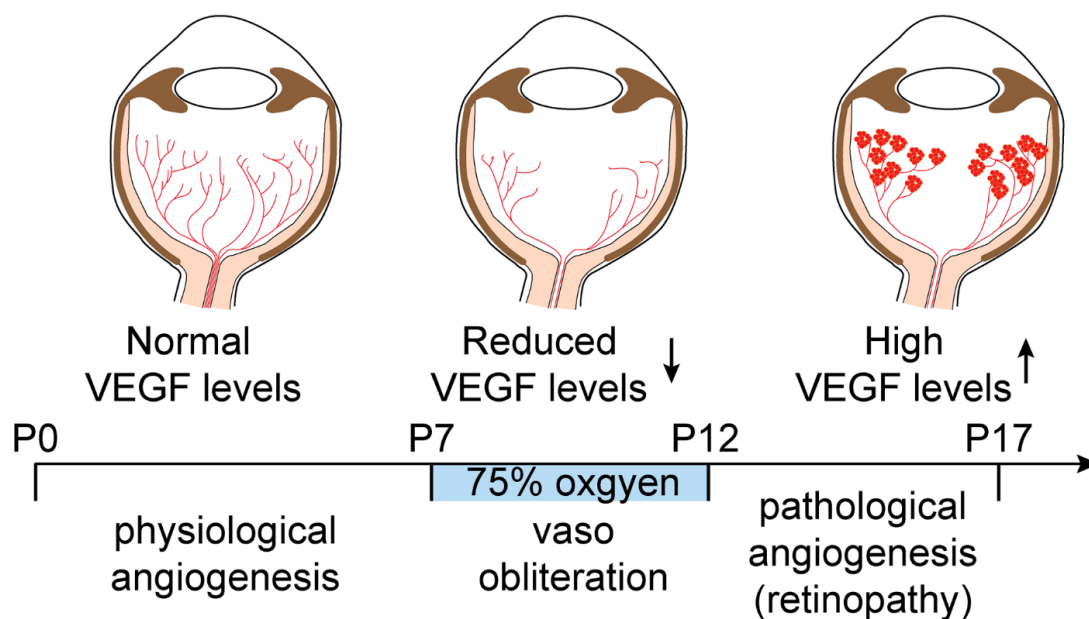


Figura 1: Representação do modelo OIR. Os tempos P0, P7, P12 e P17 representam os dias pós-nascimento do camundongo. Após o 7º dia de nascimento do camundongo, ele é levado para a câmara onde fica exposto a um ambiente com alta pressão de O<sub>2</sub>. Após 5 dias, o camundongo retorna ao ambiente com condições de normóxia. Os dias P12, P15 e P17 são os dias em que foram coletadas as retinas do camundongo para extração de RNA total. Figura retirada de GUARISCHI-SOUSA et al. (2019) - figura 1A.

Nosso grupo gerou dados de RNA-seq a partir de retinas do modelo OIR (GUARISCHI-SOUSA et al., 2019). Ao final das análises, obteve-se uma lista de 153 genes diferencialmente expressos (DE) relacionados com processo de angiogênese, sendo esses possíveis candidatos a alvos terapêuticos no tratamento de alguns tipos de câncer, retinopias e outras doenças promovidas pela angiogênese patológica. Este estudo se restringiu aos genes codificantes DE. No entanto, cada vez mais estudos têm evidenciado o papel de RNAs longos não-codificantes (ncRNAs) na regulação da angiogênese, como os microRNAs (miRNAs) (FASANARO et al., 2008; FISH et al., 2008; LANDSKRONER-EIGER; MONEKE; SESSA, 2013; SUÁREZ; SESSA, 2009; TIWARI; MUKHERJEE; DIXIT, 2018; WANG et al., 2013b), os RNAs circulares (circRNAs) (LI; MA; YU, 2017; LIU et al., 2017a; SHAN et al., 2017) e os RNA longos não codificantes (lncRNAs) (CHUNHAROJRITH et al., 2015; FU et al., 2016; GORDON et al., 2010; KUMAR; GOYAL, 2017; MICHALIK et al., 2014; TRIMARCHI et al., 2014).

Por esse motivo, a presente tese resume os estudos para identificar esses ncRNAs no modelo OIR, e analisar como esses diferentes fatores (RNAs mensageiros [mRNAs] e ncRNAs) interagem entre si, para então, propor 2 redes de regulação gênica para angiogênese patológica. Nós também analisamos as variantes de *splicing* dos 153 genes DE identificados por (GUARISCHI-SOUSA et al., 2019) a fim de

identificar as variantes que contribuíram para os resultados obtidos no primeiro estudo publicado.

### 1.3 Objetivo

O objetivo geral do presente estudo foi obter um novo conhecimento sobre os mecanismos moleculares da angiogênese, em particular sobre os principais genes que atuam nesse processo, sejam os codificadores de proteína (e suas variantes), sejam aqueles não codificantes. Nosso objetivo foi não apenas descobrir novos genes relacionados à angiogênese, mas também alcançar um maior entendimento sobre como esses genes se relacionam durante esse fundamental processo celular. Assim sendo, como parte do objetivo geral e com base no novo conhecimento sobre os genes participantes do processo, obtivemos duas redes de coexpressão gênica que fazem parte da regulação da angiogênese patológica na retina de camundongo.

Esta tese está estruturada da seguinte forma. No capítulo 2 são apresentados os conjuntos de dados que foram analisados no presente estudo, e descreve a metodologia para obter o genoma utilizado como referência nas análises. No capítulo 3, são apresentadas as variantes de *splicing* DE, suas diferenças estruturais em relação as demais variantes, e seus prováveis homólogos humanos. No Capítulo 4 apresentamos os ncRNAs (miRNAs, circRNAs e lncRNAs) DE no modelo OIR, assim como suas prováveis funções nesse contexto. No Capítulo 5 apresentamos uma proposta de rede de coexpressão de mRNAs e ncRNAs. Como forma de validação dos resultados alcançados, no Capítulo 6 apresentamos novas assinaturas gênicas para prognóstico de câncer de mama a partir dos genes presentes nas redes de coexpressão. A tese é finalizada com conclusões no Capítulo 7.



## 2 Dados de sequenciamento de RNA e o genoma de referência

### 2.1 Conjuntos de dados

Analizamos os dados de RNA-Seq, gerados a partir de bibliotecas de enriquecimento de poli-A (Poli-A+), depleção de rRNA (Ribo-) e miRNAs, construídas com base no RNA extraído do modelo OIR. As coletas das retinas dos camundongos foram feitas nos dias 12, 15 e 17, com exceção do grupo tratado para o qual foi feita uma coleta a mais denominada 12.5, que representa o pico de hipóxia na série de tempo.

Para a biblioteca Poli-A+ foi feita, para cada grupo, 2 réplicas biológicas e 6 a 7 réplicas técnicas para cada réplica biológica, com exceção do dia 15 que não foram feitas réplicas técnicas, totalizando 69 bibliotecas (Tabela 2). Para as bibliotecas Ribo- foram feitas, para cada grupo, 2 réplicas biológicas para cada tempo, totalizando 14 bibliotecas (Tabela 3). E, por fim, a biblioteca de miRNA a qual não foi feita nenhuma réplica, totalizando 7 bibliotecas (Tabela 4).

Tabela 2: Número de *reads* sequenciados para a biblioteca Poli-A+ e a porcentagem de *reads* alinhadas contra o genoma para cada réplica biológica de cada grupo/dia. As bibliotecas representadas pela letra P se referem as amostras do grupo controle, enquanto as bibliotecas representadas pela letra R se referem as amostras do grupo com retinopatia. Para os tempos 12, 12.5 e 17, os *reads* obtidos para cada réplica técnica foram somados. Não existem réplicas técnicas para o dia 15. O alinhamento foi realizado com o software STAR no modo 2-pass.

Amostra	# reads	Alinhamento único no genoma	
		#	%
P12	212.787.025	193.320.331	90,85
	149.856.330	136.842.281	91,32
P15	30.016.795	26.806.312	89,30
	13.409.462	12.171.275	90,77
P17	123.309.788	112.560.932	91,28
	129.650.238	120.269.531	92,76
R12	179.837.461	162.604.906	90,42
	117.605.934	105.115.135	89,38
R12.5	130.308.871	119.652.279	91,82
	122.023.496	110.803.132	90,80
R15	30.697.481	27.763.847	90,44
	13.013.137	12.053.606	92,63
R17	194.750.445	176.373.067	90,56
	110.163.809	99.082.266	89,94

Tabela 3: Número de *reads* sequenciados para a biblioteca Ribo- e a porcentagem de *reads* alinhados contra o genoma para cada réplica biológica de cada grupo/dia. As bibliotecas representadas pela letra P se referem as amostras do grupo controle, enquanto as bibliotecas representadas pela letra R se referem as amostras do grupo com retinopatia. O alinhamento foi realizado com o software STAR no modo 2-pass.

Amostra	# reads	Alinhamento único no genoma	
		#	%
P12	34.350.321	30.131.571	87,72
	19.355.217	17.648.945	91,18
P15	29.345.846	25.402.620	86,56
	15.802.720	14.281.860	90,38
P17	28.383.149	24.602.343	86,68
	18.715.059	17.047.108	91,09
R12	33.385.202	28.760.677	86,15
	26.843.976	20.654.460	76,94
R12.5	25.288.362	22.034.360	87,13
	16.326.175	15.163.077	92,88
R15	29.759.239	25.760.842	86,56
	21.503.232	19.936.502	92,71
R17	31.314.208	26.916.383	85,95
	19.029.728	17.026.606	89,47

Tabela 4: Número de *reads* sequenciados para a biblioteca miRNA e a porcentagem de *reads* alinhados contra o genoma para cada amostra de cada grupo/dia. As bibliotecas representadas pela letra P se referem as amostras do grupo controle, enquanto as bibliotecas representadas pela letra R se referem as amostras do grupo com retinopatia. O alinhamento foi realizado com o software Bowtie.

Amostra	# reads	Alinhamento único no genoma	
		#	%
P12	15.858.160	9.193.075	57,97
P15	16.715.938	9.332.424	55,83
P17	14.698.562	10.138.681	68,98
R12	12.246.680	8.178.503	66,78
R12.5	17.691.828	11.988.358	67,76
R15	13.608.053	8.510.907	62,54
R17	14.854.390	9.198.395	61,92

## 2.2 Obtendo o genoma de referência do *Mus musculus* da linhagem C57BL/6N

O genoma disponível em bancos de dados públicos (ex.: ENSEMBL e NCBI) são referentes a linhagem C57BL/6J produzida por *The Jackson Laboratory*, a primeira a ter o genoma sequenciado. Contudo, a linhagem utilizada nesse estudo foi a C57BL/6N da NIH (*National Institutes of Health*), a qual possui algumas variações genéticas comparadas a linhagem 6J (SIMON et al., 2013). Para isso foi desenvolvido um *script* (rf2m) para modificar o genoma de referência (6J) com as mutações da linhagem 6N descritas em (KEANE et al., 2011). O *script* está disponível no GitHub (<https://github.com/lbi-usp/rf2m>).

A Figura 2 mostra uma comparação do alinhamento de um mesmo conjunto de *reads* entre o genoma referência (6J) e o genoma 6N.

C57BL/6N		C57BL/6J	
Started job on	Nov 29 11:00:3	Started job on	Feb 16 15:23:0
Started mapping on	Nov 29 11:07:2	Started mapping on	Feb 16 15:31:1
Finished on	Nov 29 11:11:5	Finished on	Feb 16 15:35:0
Mapping speed, Million of reads per hour	409.27	Mapping speed, Million of reads per hour	480.20
Number of input reads	30012803	Number of input reads	30012803
Average input read length	100	Average input read length	100
UNIQUE READS:		UNIQUE READS:	
Uniquely mapped reads number	26873346	Uniquely mapped reads number	26914421
Uniquely mapped reads %	89.54%	Uniquely mapped reads %	89.68%
Average mapped length	99.60	Average mapped length	99.60
Number of splices: Total	6294301	Number of splices: Total	6175127
Number of splices: Annotated (sjdb)	6294104	Number of splices: Annotated (sjdb)	6163409
Number of splices: GT/AG	6224056	Number of splices: GT/AG	6060489
Number of splices: GC/AG	59083	Number of splices: GC/AG	82928
Number of splices: AT/AC	7796	Number of splices: AT/AC	10878
Number of splices: Non-canonical	3366	Number of splices: Non-canonical	20832
Mismatch rate per base, %	0.49%	Mismatch rate per base, %	0.49%
Deletion rate per base	0.01%	Deletion rate per base	0.01%
Deletion average length	1.53	Deletion average length	1.53
Insertion rate per base	0.01%	Insertion rate per base	0.01%
Insertion average length	1.28	Insertion average length	1.27
MULTI-MAPPING READS:		MULTI-MAPPING READS:	
Number of reads mapped to multiple loci	1774790	Number of reads mapped to multiple loci	1703347
% of reads mapped to multiple loci	5.91%	% of reads mapped to multiple loci	5.68%
Number of reads mapped to too many loci	30833	Number of reads mapped to too many loci	65778
% of reads mapped to too many loci	0.10%	% of reads mapped to too many loci	0.22%
UNMAPPED READS:		UNMAPPED READS:	
% of reads unmapped: too many mismatches	0.00%	% of reads unmapped: too many mismatches	0.00%
% of reads unmapped: too short	4.11%	% of reads unmapped: too short	4.09%
% of reads unmapped: other	0.34%	% of reads unmapped: other	0.34%
CHIMERIC READS:		CHIMERIC READS:	
Number of chimeric reads	0	Number of chimeric reads	0
% of chimeric reads	0.00%	% of chimeric reads	0.00%

Figura 2: Comparação do alinhamento de um mesmo conjunto de *reads* contra o genoma de referência (C57BL/6J) e o genoma modificado (C57BL/6N) pela ferramenta rf2m. O alinhamento foi realizado com a ferramenta STAR no modo 2-pass.

### 3 Análise das variantes de *splicing* diferencialmente expressas durante angiogênese patológica

Sabe-se que a diversidade de proteínas encontradas em uma grande parcela dos eucariontes é devida ao *splicing* alternativo (AS) dos transcritos, que consiste na remoção de íntrons específicos e a junção de uma combinação específica de éxons de um gene, permitindo que um único gene seja capaz de gerar distintas proteínas, que podem desempenhar funções diferentes ou até mesmo opostas (CHEN et al., 2019; EKSI et al., 2013; LAROCHELLE, 2016).

O gene *VEGF-A* humano possui 8 éxons (Figura 3). Os éxon 6, 7 e 8 são os que mais participam do AS, resultando em variantes que codificam proteínas com propriedades distintas. Por exemplo, os *VEGF*<sub>165</sub> e *VEGF*<sub>189</sub> humanos, codificados pelas variantes que contêm os éxons 6 e 7, se ligam com alta afinidade aos proteoglicanos da ECM (PEACH et al., 2018). Já o éxon 8 possui duas regiões de *splicing*, 8a e 8b, cujas variantes codificam proteínas que se diferem apenas nos 6 aminoácidos C terminais (PEACH et al., 2018; VAREY et al., 2008). As variantes que terminam com éxon 8a codificam proteínas do tipo *VEGF*<sub>xxx</sub>*a* (“xxx” representa o número de aminoácidos da proteína gerada) e estão associadas com funções pró-angiogênicas, enquanto as variantes que terminam com éxon 8b, *VEGF*<sub>xxx</sub>*b*, estão associadas com funções anti-angiogênicas, pois essas competem com as variantes *VEGF*<sub>xxx</sub>*a* pelo sítio ativo do receptor *VEGFR2*, inibindo a sinalização da via que promove a proliferação celular e migração das células endoteliais (PEACH et al., 2018; VAREY et al., 2008). Em 2014 uma nova variante foi identificada por Eswarappa et al., (2014), que possui tanto os éxons 8a e 8b, cuja proteína resultante (*VEGF-Ax*) tem sido associada tanto com funções pró-angiogênicas quanto anti-angiogênicas (PEACH et al., 2018). Essa variante é gerada a partir da extensão da tradução do éxon 8a, onde a tradução não termina no *stop* códon canônico, mas sim em um *stop* códon alternativo dentro da região 3' UTR (PEACH et al., 2018).

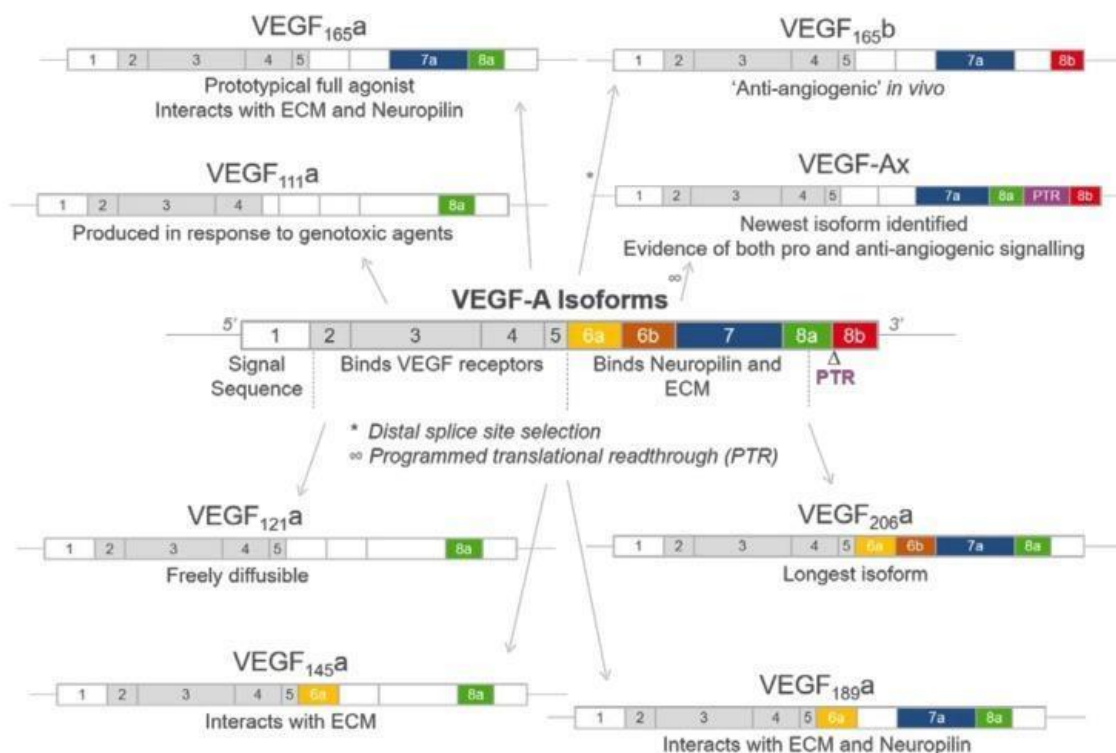


Figura 3: Variantes de *splicing* do gene VEGF-A identificadas em humanos. As variantes VEGF<sub>xxx</sub>a, que possuem o éxon 8a são pró-angiogênicas, enquanto as variantes VEGF<sub>xxx</sub>b, que codificam o éxon 8b, são anti-angiogênicas, apesar de existir estudos que contestam sua existência. E por fim as variantes VEGF-Ax que codificam os éxons 8a e 8b, e que podem assumir funções pró-angiogênicas ou anti-angiogênicas. Figura retirada de Peach et al. (2018) - figura 1.

Com relação as variantes anti-angiogênicas, VEGF<sub>xxx</sub>b, um estudo conduzido em 2017 com dados de RNA-Seq de humanos contestou a existência dessas variantes, alegando que elas se encontram em baixas concentrações ou não existem *in vivo*, e que as técnicas utilizadas nos estudos anteriores podem ter introduzido vieses nos resultados (BRIDGETT; DELLETT; SIMPSON, 2017). Isso demonstra que, mesmo para genes como o VEGF que já vem sendo estudado a algum tempo, ainda existem lacunas que precisam ser preenchidas. A angiogênese é um processo complexo, e o VEGF é apenas um dentre outros diversos fatores que contribuem para esse processo. Isso nos permite questionar: será que, assim como VEGF, existem variantes de outros genes que participam da angiogênese que podem estar desempenhando papéis diferentes da variante canônica? Por exemplo, identificar variantes preponderantes na angiogênese patológica pode auxiliar na identificação de novos alvos terapêuticos para os tratamentos das diversas doenças dependentes de angiogênese. Por esse motivo, decidimos verificar quais variantes de *splicing* dos 153 genes DE identificados por GUARISCHI-SOUSA et al., 2019, estão sendo reguladas no modelo OIR.

As análises de bioinformática referentes a identificação, quantificação, expressão diferencial, busca por homólogos humanos e comparação das estruturas das variantes, que serão descritas ainda neste capítulo, foram desenvolvidas em conjunto com a aluna de Iniciação Científica Débora Guerra Peixe (orientada pelo Prof. Dr. João Carlos Setubal). As validações das variantes por RT-PCR foram realizadas pela aluna de doutorado Lilian Costa Alecrim de Oliveira (orientada pelo Prof. Dr. Ricardo José Giordano)

### 3.1 Métodos

#### 3.1.1 Quantificação e análise de expressão diferencial

O primeiro passo foi selecionar todas as variantes dos 153 genes DE identificados por GUARISCHI-SOUSA et al., 2019. De acordo com o arquivo de anotação do genoma do camundongo, disponibilizado pelo ENSEMBL (CUNNINGHAM et al., 2019), dos 153 genes DE encontrados pelo grupo, apenas 124 genes possuem duas ou mais variantes descritas. Criamos então uma lista com todas as variantes desses 124 genes, resultando em 616 variantes. Quantificamos a expressão de cada uma delas com o software RSEM (LI; DEWEY, 2011) e filtramos as variantes com baixo nível de expressão (*counts per million* (CPM)  $\geq 1$  em pelo menos 2 bibliotecas). Para a análise de expressão diferencial utilizamos o software edgeR (ROBINSON; MCCARTHY; SMYTH, 2010). Utilizamos como critério o *false discovery rate* (FDR)  $\leq 0,05$  para classificar uma variante como DE.

#### 3.1.2 Busca por homólogos humanos para as variantes de *splicing* diferencialmente expressas

A correspondência entre genes humanos e murinos é um passo necessário para que o estudo de doenças em modelo animal possa levar ao melhor entendimento das doenças humanas e, possivelmente, identificar novos alvos farmacológicos para desenvolvimento de novos medicamentos. No nosso caso, essa relação apenas entre os genes, sem considerar suas variantes, não é suficiente. O motivo é que 2 genes podem ser considerados homólogos, mas suas diferentes variantes podem ter papéis bem distintos nos processos celulares dos quais participam. Entendemos que apenas a inferência da homologia das variantes DEs poderá efetivamente nos ajudar a

compreender o papel dessas variantes nas doenças humanas dependentes de angiogênese.

Para determinar a homologia das variantes, realizamos o alinhamento das variantes de camundongo com as variantes humanas, disponíveis no ENSEMBL (GRCh38 cdna e ncna), utilizando o BLASTn (cobertura  $\geq 98\%$  e e-value  $\leq 1 \times 10^{-5}$ ). Outro software utilizado foi o Exalign (PAVESI et al., 2008), onde comparamos a composição e tamanho dos éxons entre as variantes humanas e de camundongo. Também consultamos os resultados publicados de dois estudos (BLANQUART et al., 2016; ZAMBELLI et al., 2010) que propuseram a homologia entre as variantes humanas e de camundongo. Para as variantes codificantes, também fizemos o alinhamento das proteínas de camundongo e humana com o BLASTp (identidade  $\geq 80\%$ , cobertura  $\geq 95\%$ , e e-value  $\leq 1,0 \times 10^{-5}$ ). Nós estabelecemos algumas categorias para os candidatos homólogos:

- evidência forte: transcritos com um *hit* único com BLASTn;
- evidência moderada: 2 ou mais resultados concordantes provenientes de diferentes linhas de evidências (ferramentas e/ou literatura);
- evidência fraca: apenas um resultado de uma única linha de evidência.
- sem evidência

### 3.1.3 Comparação estrutural das variantes diferencialmente expressas

A fim de se entender melhor as implicações das variantes que foram DEs no nosso modelo animal, nós realizamos a comparação da estrutura das variantes, em termos de composição de éxons e proteína gerada. Todas as variantes DEs que identificamos são potenciais candidatas para serem alvos de estudos mais aprofundados para se elucidar sua função na angiogênese.

Para essa análise, consideramos apenas as variantes dos genes que codificam 2 ou mais variantes codificantes. Utilizamos as anotações do genoma de camundongo (GTF mouse ENSEMBL 104) para comparar as variantes e identificar os éxons exclusivos de cada variante, e verificamos se nesses éxons havia algum domínio de proteína utilizando os modelos do PFAM (MISTRY et al., 2021) como referência e o HMMER (FINN; CLEMENTS; EDDY, 2011) para fazer as buscas. Para comparação das proteínas geradas por cada variante, realizamos alinhamentos par-a-par das sequências de aminoácidos de cada proteína com o software EMBOSS Needle (MADEIRA et al., 2019).

### 3.1.4 Validação das variantes de interesse por RT-PCR

Seguindo os critérios citados, selecionamos para os estudos de validação, 6 variantes DEs que pertencem aos seguintes genes: *ADGRG6*, *ENG*, *NES*, *PIWIL4*, *COL2A1* e *TNFAIP2*. A sequência dos *primers* se encontram na Tabela 5.

Tabela 5: Sequências dos primers das variantes DEs de interesse para validação.

<b>Símbolo</b>	<b>Variante</b>	<b>Forward Primer (5'-&gt;3')</b>	<b>Reverse Primer (5'-&gt;3')</b>
<i>ADGRG6</i>	ENSMUST00000041168	CCTTCAACAAAAGTGGATCACTC	CAGATGCCTTCAGCCACAG
<i>COL2A1</i>	ENSMUST00000088355	CCAGGATGCCCGAAAATTAGG	GGTCCTCTGGGTCCTATGATATC
<i>NES</i>	ENSMUST00000160694	CAGGAGTCCCTGAAGTCTCC	CTGGTCCTCTGCTTCTTCAAAC
<i>PIWIL4</i>	ENSMUST00000115644	CCTGGGTTTGCCATTCTG	GATCAGTCATGAAGTCCAGAAC
<i>ENG</i>	ENSMUST00000113272	CTGAAGAAGCGGGGAGTC	GGTTGGTACTGCTGCTCTC
<i>TNFAIP2</i>	ENSMUST00000102745	CGACATTCTCCAGATCATTTTAC	GCTTTATCTGCATCCCCAC

Retinas de filhotes de camundongos submetidos ao modelo OIR e em quatro estágios de desenvolvimento (P12, P12.5, P15 e P17) foram dissecadas em RNAlater (Sigma) e o RNA total foi extraído utilizando o protocolo Qiagen RNeasy Mini Kit (Qiagen). As amostras foram tratadas com DNase (TURBO DNA-free, Thermo Fisher Scientific) seguindo o protocolo do fabricante e a ausência de contaminação de DNA genômico foi confirmada por PCR convencional. Utilizamos 500 ng de RNA total para realizar a transcrição reversa do RNA com o kit SuperScript III (Invitrogen) em um volume total de 20 µL, de acordo com as instruções do fabricante. qPCR foi realizada em um volume total de 12 µL por reação. Colocamos 3 µL de cDNA (diluído 1:10) em uma mistura de reação de 9 µL que continha 3 µL do mix de primers de concentração 1,2 µM e 6 µL de SYBR Green PCR Master Mix. O gene endógeno utilizado foi o *SDHA*.

## 3.2 Resultados

### 3.2.1 Expressão diferencial das variantes

Ao todo foram identificadas 69 variantes DEs. Dentre elas, algumas correspondem ao HLA humano, moléculas de histocompatibilidade envolvidas na apresentação de antígenos para o sistema imune. Como estes são genes específicos para a resposta imune e, até onde sabemos, não estão relacionados com a angiogênese, eles foram



excluídos das análises posteriores. Temos interesse em estudar apenas variantes específicas para angiogênese patológica, pois estas representam potenciais alvos para desenvolvimento de novas drogas. Com isso, obtivemos um total de 60 variantes DEs (Apêndice II: Lista de variantes diferencialmente expressas), as quais correspondem a 47 genes, sendo que 10 desses genes possuem mais de uma variante DE. Dez variantes entre as 60 DEs são possíveis lncRNAs, pois são classificadas como não-codificantes de acordo com o ENSEMBL, e estão descritas como lncRNAs no banco de dados Alliance of Genome Resources (BULT et al., 2019). Usando a classificação do APPRIS (RODRIGUEZ et al., 2018), 7 variantes foram classificadas como *Alternative* ou *Minor*. A Figura 4 apresenta o perfil de expressão gênica das variantes DEs.

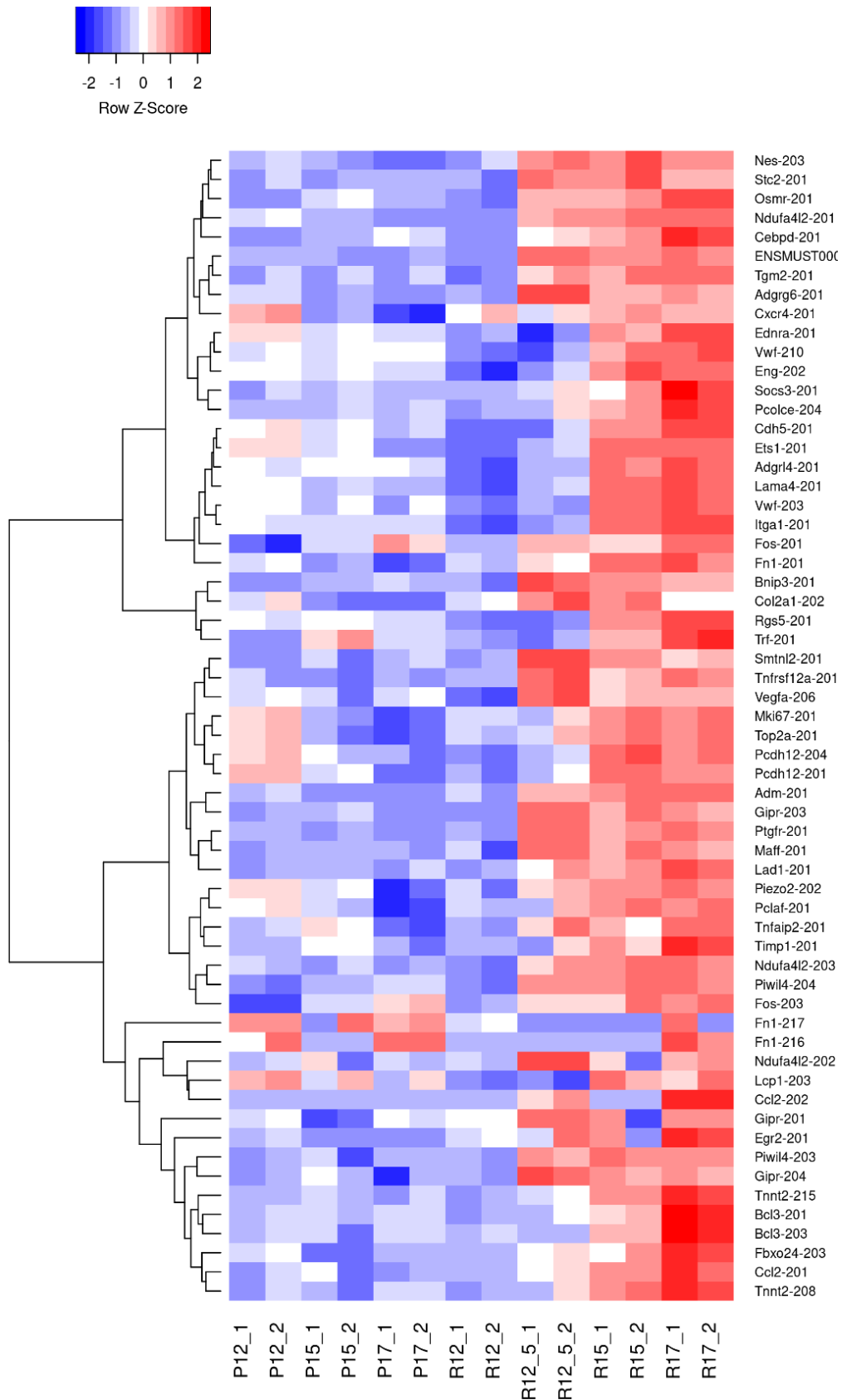


Figura 4: *Heatmap* do perfil de expressão das 60 variantes DEs. A letra P representa as amostras do grupo controle, e a letra R representa as amostras do grupo retinopático. Aqui notamos que as amostras R12\_1 e R12\_2 possuem um perfil similar às amostras do grupo controle (P12\_1 e P12\_2). Como as amostras R12 são colhidas logo quando os camundongos retornam ao ambiente com condições de normóxia, é provável que não houve tempo o suficiente para os genes de resposta a hipóxia serem

ativados, uma vez que o pico de hipóxia somente ocorre algumas horas depois (amostras R12\_5\_1 e R12\_5\_2). A grande maioria dessas variantes começam aumentar sua expressão no dia 15 (R15\_1 e R15\_2), e algumas delas aumentam sua expressão ainda mais no dia 17 (R17\_1 e R17\_2). Já outras variantes, como as que pertencem aos genes VEGFA, SMTLN2, NDUFA4L2, BNIP3 e ADGRG6, possuem uma maior expressão no dia 12\_5 e depois começam a diminuir ao longo do tempo, o que indicam que elas estão envolvidas com a fase de resposta à hipóxia das células da retina.

### 3.2.2 Homólogos humanos para as variantes de *splicing* diferencialmente expressas

Seguindo a metodologia detalhada na seção 3.1.2 , conseguimos apontar candidatos homólogos em humanos para 26 variantes das 60 DE, como mostra a Tabela 6 (lista completa no Apêndice III: Lista de candidatos homólogos para as variantes diferencialmente expressas).

Tabela 6: Total de variantes com candidatos humanos homólogos em cada uma das categorias: evidência forte, transcritos com um hit único com BLASTn; evidência moderada, 2 ou mais resultados concordantes provenientes de diferentes linhas de evidências (ferramentas e/ou literatura); evidência fraca, apenas um resultado de uma única linha de evidência; e sem evidência.

<b>Categoria</b>	<b>Total</b>
Evidência forte	6
Evidência moderada	20
Evidência fraca	11
Sem evidência	23

### 3.2.3 Comparação estrutural das variantes diferencialmente expressas

Seguindo a metodologia apresentada na seção 3.1.3 , apenas 25 variantes DE atendiam aos critérios. Dessas, 3 variantes geravam proteínas idênticas a pelo menos uma variante “irmã” (Tabela 7). Foi observado que para esses casos, a diferença entre as variantes se encontra nas UTRs (Figura 5, Figura 6 e Figura 7).

Tabela 7: Variantes DEs que possuem o mesmo produto proteico de pelo menos uma de suas variantes “irmãs”.

Gene	Variante DE	Variante não-DE
<i>TIMP1</i>	ENSMUST00000009530	ENSMUST00000115342
<i>PTGFR</i>	ENSMUST00000029670	ENSMUST00000106126
<i>MAFF</i>	ENSMUST00000096350	ENSMUST00000163691
<i>MAFF</i>	ENSMUST00000096350	ENSMUST00000229130

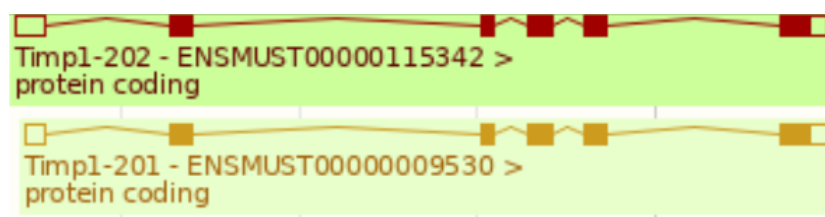


Figura 5: Comparação das variantes do gene *TIMP1*. A 5' UTR da variante DE (ENSMUST00000009530) é menor que a da variante ENSMUST00000115342. Imagem retirada do ENSEMBL database.

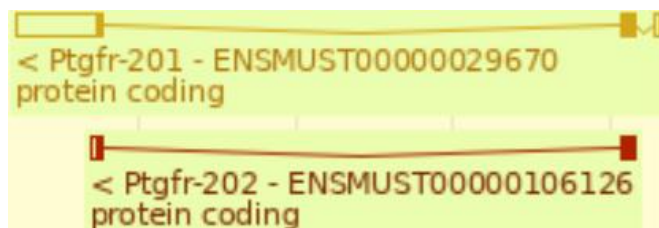


Figura 6: Comparação das variantes do gene *PTGFR*. Ambas UTRs da variante DE (ENSMUST00000029670) é maior que a da variante ENSMUST00000106126. Imagem retirada do ENSEMBL database.

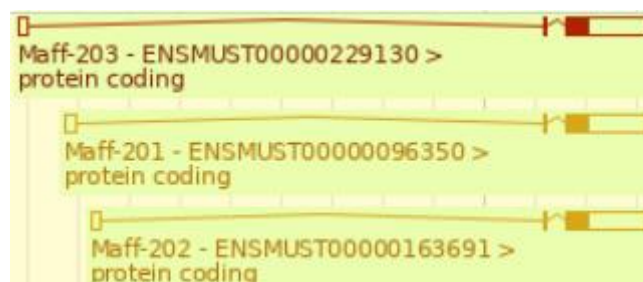


Figura 7: Comparação das variantes do gene *MAFF*. A 5' UTR da variante DE (ENSMUST00000096350) é diferente quando comparada a das variantes ENSMUST00000163691 e ENSMUST00000229130. Imagem retirada do ENSEMBL database.

Todas as demais variantes geram proteínas que diferem em tamanho e/ou composição de aminoácidos de suas variantes “irmãs” e requerem uma maior atenção

para relacionar essas diferenças com o processo de angiogênese. Para esta Tese, nós analisamos em maiores detalhes as variantes que foram selecionadas para validação em RT-PCR, sendo elas pertencentes aos genes *ADGRG6*, *ENG*, *NES*, *PIWIL4*, *COL2A1* e *TNFAIP2*.

A variante *ADGRG6-201* (ENSMUST00000041168), uma variante alternativa do gene *ADGRG6*, foi regulada positivamente no dia 12, durante o pico de hipóxia. Comparando a com a variante principal *ADGRG6-205* (ENSMUST00000238680), eles diferem quanto ao último éxon, no qual a variante *ADGRG6-201* contém a 3' UTR maior, e essa mesma variante possui um éxon exclusivo próximo a ponta 3' (Figura 8). Os domínios presentes em ambas as proteínas são os mesmos, porém na ponta C terminal da proteína da variante alternativa, existe uma região de desordem intrínseca (IDR) que não está presente na variante principal.

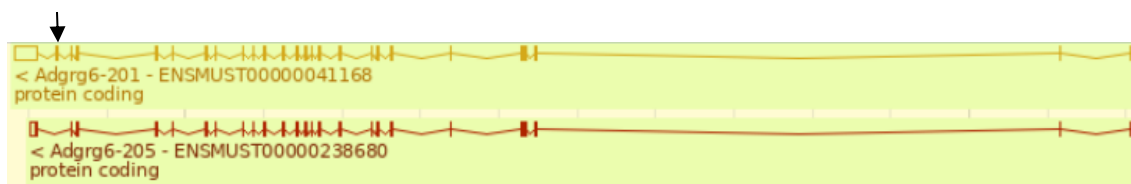


Figura 8: Comparação entre a variante alternativa *ADGRG6-201* (ENSMUST00000041168), DE no nosso modelo, e a variante principal *ADGRG6-205* (ENSMUST00000238680) do gene *ADGRG6*. A seta aponta para o éxon 24 que é exclusivo da variante *ADGRG6-201*, que compreende uma região de desordem intrínseca (IDR). Imagem retirada do ENSEMBL Database.

A variante *ENG-202* (ENSMUST00000113272), conhecida como *S-Eng*, está regulada negativamente no dia 12 durante pico de hipóxia. Comparando os produtos proteicos gerados pela variante *S-Eng* e a variante principal, *FL-Eng* (*ENG-201*), elas diferem apenas na ponta C-terminal, que corresponde a porção citoplasmática da proteína, que é maior na *FL-Eng* e tem uma composição de aminoácidos diferente da *S-Eng*. A *S-Eng* tem 14 éxons enquanto a *FL-Eng* possui 15. No entanto, o 14º éxon da *S-Eng* corresponde aos éxons 14 e 15 da *FL-Eng*, além de reter o íntron entre esses éxons que contém um *stop* códon em frame, resultando em um peptídeo menor (Figura 9).



Figura 9: Variantes do gene *ENG* de camundongo. A seta aponta para o éxon 14 da variante *S-Eng* (*ENG-202*) que corresponde aos éxons 14 e 15, mais o íntron que fica retido, da variante *FL-Eng* (*Eng-201*). Imagem retirada do ENSEMBL Database.

A variante principal, *PIWIL4-203* (ENSMUST00000115644), e a variante alternativa *PIWIL4-204* (ENSMUST00000136399) estavam reguladas positivamente do dia 12 durante o pico de hipóxia, mas somente a variante principal estava regulada positivamente no dia 17. O éxon 11 (ENSMUSE00000380059) é exclusivo da variante principal (Figura 10). Não foi possível encontrar nenhum domínio nessa região com o PFAM, no entanto o CDD (*Conserved Domains Database*) aponta que esse éxon contém uma porção do domínio PIWI (cd04658).

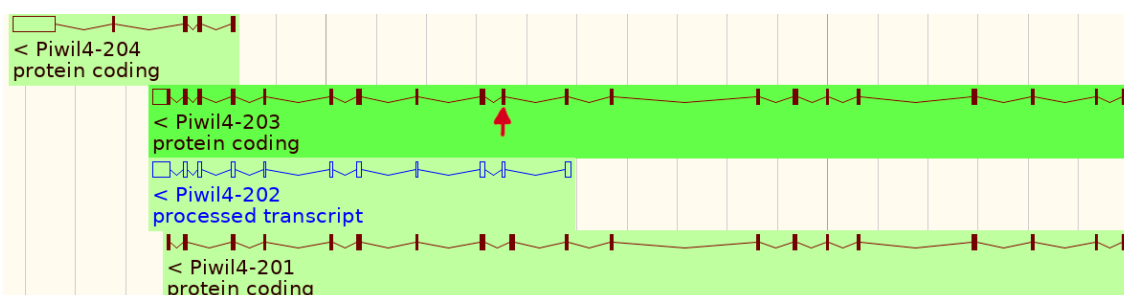


Figura 10: Variantes do gene *PIWIL4* do camundongo. A seta aponta para o éxon 11 que é exclusivo da variante principal *PIWIL4-203*. Imagem retirada do ENSEMBL Database.

A variante alternativa *NES-203* (ENSMUST00000160694) estava regulada positivamente no dia 12 durante o pico de hipóxia. Comparamos a composição de éxons da variante alternativa com a variante principal *NES-201* (ENSMUST00000090973), e observamos que o último éxon (4<sup>o</sup>) do *Nes-201* tem uma região de *splicing* que divide esse éxon em 2 novos éxons que correspondem aos éxons 4 e 5 da variante *NES-203* (Figura 11), resultando em um peptídeo 44 aminoácidos menor que a variante principal. Essa região não contém nenhum domínio funcional conhecido, e não existe um estudo prévio sobre o papel dessa variante alternativa.

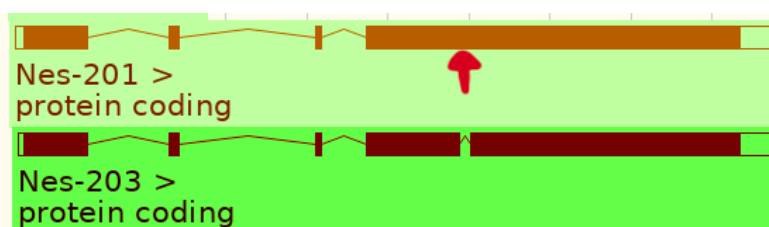


Figura 11: Variantes do gene *NES* de camundongo. A seta aponta para o éxon 4 da variante principal *NES-201* que corresponde a junção dos éxons 4 e 5 da variante alternativa *NES-203*. Imagem retirada do ENSEMBL Database.

Em camundongos, *COL2A1* tem apenas 2 variantes codificantes de proteína. A variante alternativa menor (*COL2A1-202*, ENSMUST00000088355) foi regulada

negativamente no dia 17 no nosso modelo animal. Essa variante não possui o éxon 2 do gene, que está presente na variante principal (*COL2A1-201*, ENSMUST00000023123) (Figura 12), resultando em uma proteína com 68 aminoácidos a menos que o produto proteico da variante principal. Esse éxon codifica parte do domínio VWFC (Von Willebrand factor tipo C).

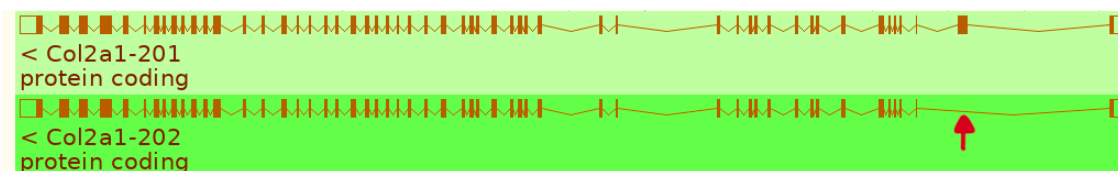


Figura 12: Variantes do gene *COL2A1* de camundongo. A seta aponta para éxon 2 que não se encontra na variante alternativa *COL2A1-202*, que está DE, e que contém o domínio VWFC. Imagem retirada do ENSEMBL Database.

O gene *TNFAIP2* possui 3 variantes codificantes, uma menor (*TNFAIP2-205*, ENSMUST00000129467) e outras duas de tamanho e composição de éxons bem semelhantes, *TNFAIP2-201* (ENSMUST00000102745) e a *TNFAIP2-202* (ENSMUST00000109792). A APPRIS classifica a variante *TNFAIP2-201* como uma variante principal, enquanto a ENSEMBL classifica a variante *TNFAIP2-202* como a variante canônica. A variante *TNFAIP2-201* foi classificada como DE no dia 17. O éxon 6 dessa variante é menor que a *TNFAIP2-202*, gerando uma proteína com 17 aminoácidos a menos (Figura 13). Esse éxon codifica parte do domínio conhecido como complexo de excitose multiproteico, o *Sec6*. Outra diferença reside nas regiões 5' e 3' UTRs que são maiores que da variante *TNFAIP2-202*.

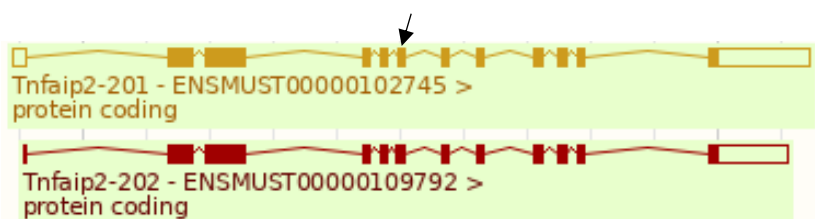


Figura 13: Variantes do gene *TNFAIP2* de camundongo. A seta aponta para o éxon 6 da variante DE *TNFAIP2-201*, que é menor que o éxon presente na variante “irmã” *TNFAIP2-202*. Imagem retirada do ENSEMBL Database.

### 3.2.4 Validação por RT-PCR

De maneira geral, as variantes escolhidas foram validadas com sucesso, pois existe uma correlação significativa entre os  $\log_2$  Fold Change (LFC) obtidos com RNA-Seq e RT-PCR (Figura 14). Mas se olharmos individualmente as variantes (Figura 15), podemos observar que para algumas delas houve uma divergência entre o observado

com RNA-Seq e o RT-PCR em quase todos os tempos analisados. São elas: *ENG-202* (Figura 15d) e *ADGRG6-201* (Figura 15f).

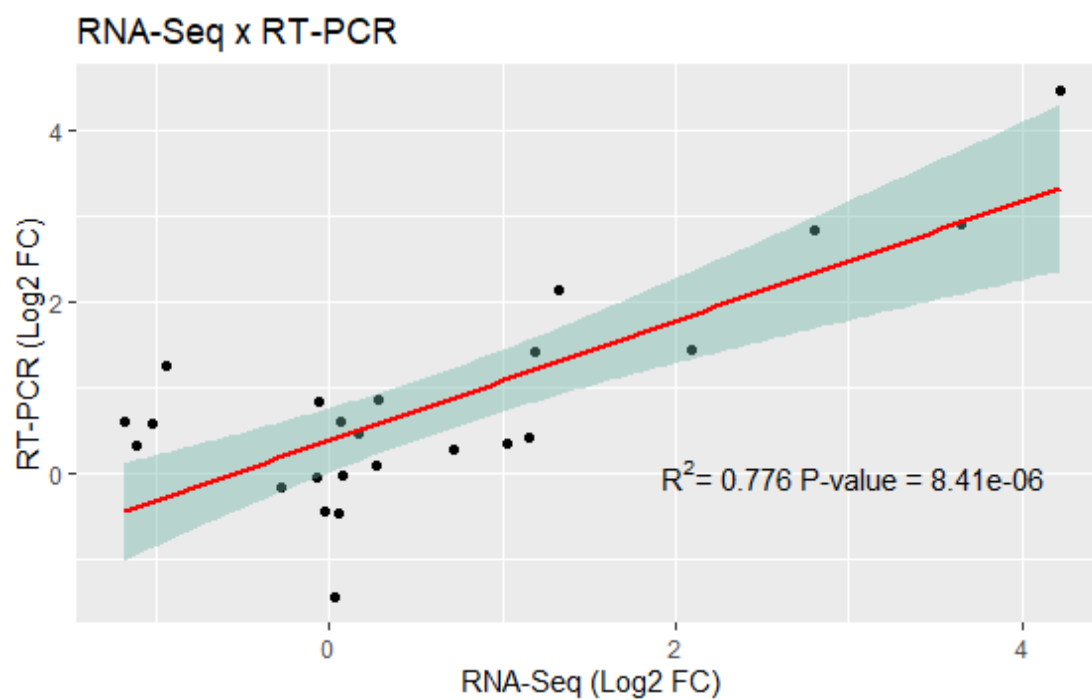


Figura 14: Correlação (*Pearson*) entre o perfil de expressão das variantes selecionadas com os métodos de RNA-Seq e RT-PCR.



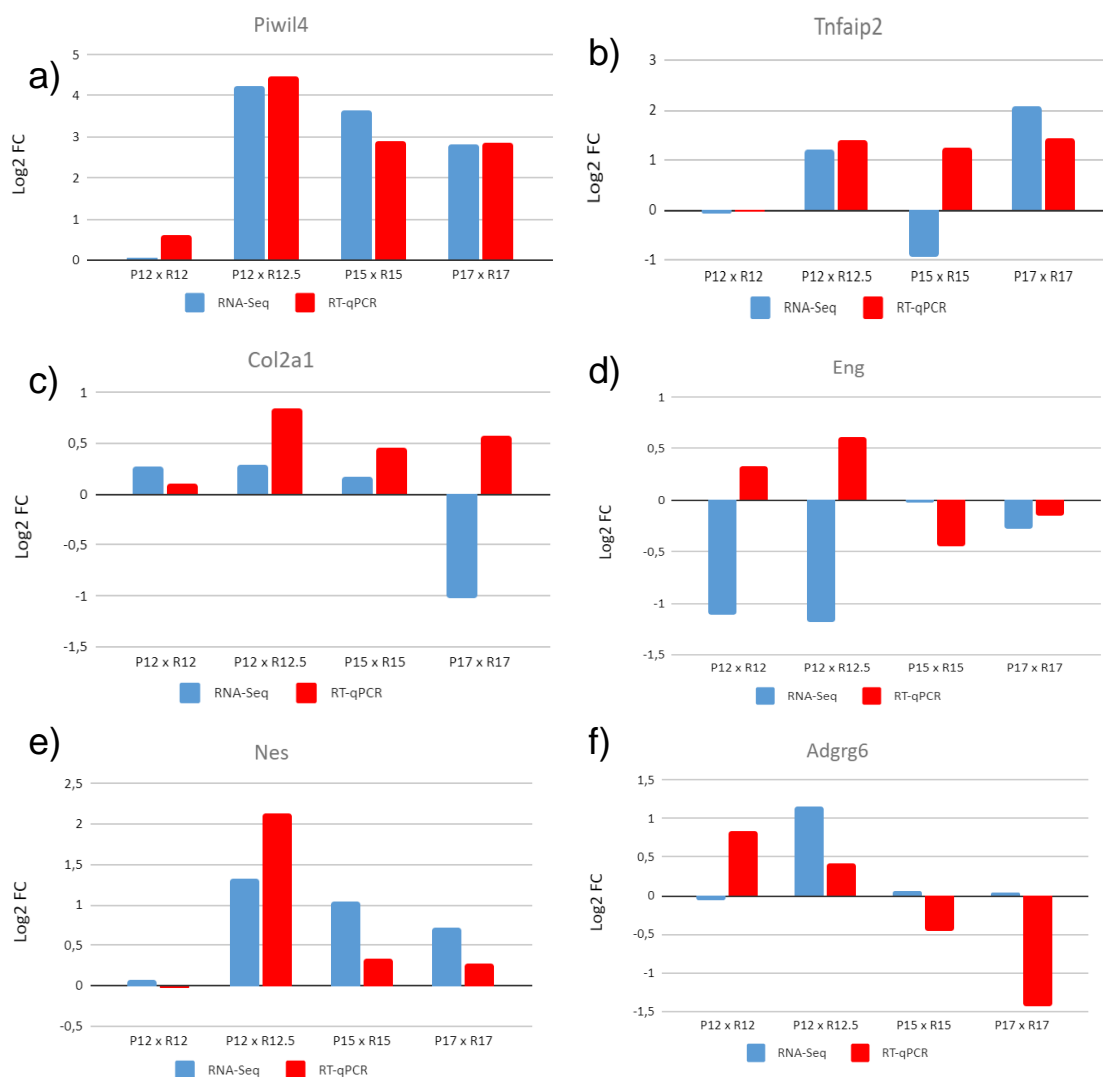


Figura 15: Comparação dos LFC obtidos com RNA-Seq e RT-PCR para as variantes dos genes: a) *PIWIL4*; b) *TNFAIP2*; c) *COL2A1*; d) *ENG*; e) *NES*; f) *ADGRG6*.

### 3.3 Discussão

Analisando as variantes dos genes DE identificados em GUARISCHI-SOUSA et al. (2019), nós identificamos que 60 variantes estão DE, sendo que 10 representam variantes alternativas. Conseguimos apontar para 26 variantes DE suas possíveis variantes homólogas em humanos. Também analisamos a diferença em termos de composições de éxons e o tamanho da proteína codificada entre as variantes identificadas como DE e as demais variantes de cada gene.

Verificamos que três variantes (*TIMP1*, *PTGFR* e *MAFF* - Tabela 7) geram a mesma proteína, mas as regiões não traduzidas (UTR) tem tamanhos diferentes. Sabe-se que essas regiões são importantes para regulação da expressão dos mRNAs (LIN; LI, 2012). Uma investigação interessante para futuros trabalhos é verificar como

essas UTRs podem estar favorecendo a expressão dessas variantes DE em detrimento das demais variantes. A sequência de nucleotídeos da UTR determina como é feita essa regulação (regiões de ligação de miRNA, elementos ricos em AU, *riboswitches*, e elementos de inserção de sequências de selenocisteínas), mas o seu tamanho também pode influenciar nesse controle. RNAs mensageiros de vertebrados que possuem 5' UTR maiores, geralmente são aqueles que possuem uma regulação muito mais controlada, por exemplo genes envolvidos no ciclo de divisão celular (KOZAK, 1987; LIN; LI, 2012), enquanto mRNAs de metazoários com 3' UTR maiores tendem a ter mais regiões de ligação de miRNAs, e, portanto, possuem sua expressão mais controlada (CHEN et al., 2012; WANG et al., 2019).

Selecionamos 6 variantes DE para validação por RT-PCR. O que elas possuem em comum é que se trata de variantes não-canônicas, e, portanto, são interessantes candidatos para estudos mais direcionados.

ADGRG6, também conhecido como GPR126, pertence à família de receptores transmembranares GPCRs (G protein-coupled receptors) (CUI et al., 2014). Esse receptor tem uma larga porção extracelular que contém diversos domínios (ex. fator de crescimento epidérmico, caderina, pentraxina e repetições ricas em leucina) responsável por mediar interações célula-matriz, célula-célula e proteína-proteína (GAD; BALENGA, 2020). Estudos anteriores mostraram que ADGRG6 possui diversos papéis em diferentes tecidos, e um deles é na angiogênese (CUI et al., 2014, 2021). Foi observado que ADGRG6 é capaz de induzir a expressão de STAT5 e GATA2 através da via de sinalização cAMP-PKA, regulando positivamente a expressão de VEGFR2 (CUI et al., 2014) e assim induzindo a angiogênese. O silenciamento de ADGRG6 em retinas de camundongo resultou na redução da angiogênese induzida por hipóxia (CUI et al., 2014). Outros trabalhos também observaram que ADGRG6 é capaz de induzir proliferação celular em câncer colorretal (CUI et al., 2021), e formação de tubo e recrutamento de células endoteliais em câncer de bexiga (WU et al., 2019b).

Endogлина (ENG), também conhecida como CD105, é uma glicoproteína transmembranar presente em diferentes tipos celulares (ex. células endoteliais vasculares, monócitos, macrófagos, células mesenquimais, fibroblastos e células musculares lisas vasculares), e ela é essencial tanto para vasculogênese quanto para angiogênese (GALLARDO-VARA et al., 2018; ZHU et al., 2017). A deficiência de ENG em células endoteliais humanas inibiu a indução de angiogênese por VEGF, enquanto em células-tronco embrionárias *in vitro* não foram capazes de se organizarem em estruturas tubulares (ZHU et al., 2017). Embriões mutantes (*ENG* silenciado) morreram entre 10 e 12 dias durante a gestação (ZHU et al., 2017). O gene *ENG*

contém um elemento de resposta a hipóxia na região do *enhancer*, o que explica a indução da expressão de *ENG* durante hipóxia, porém, a princípio sua expressão não é induzida por *HIF-1*, mas pelas vias de sinalização *ERK-p38 MAPK* e *JNK* (ZHU et al., 2017). *ENG* está associado a diversas doenças relacionadas com endotélio (GALLARDO-VARA et al., 2018), como doenças cerebrovasculares (ZHU et al., 2017), cardiovasculares (SIER et al., 2021) e câncer (BERNABEU; LOPEZ-NOVOA; QUINTANILLA, 2009; LI et al., 2000; SIER et al., 2021). Também foi observado que super expressando a variante *S-Eng* em um camundongo modelo houve um retardamento do crescimento do tumor e da neovascularização em células transplantadas de carcinoma pulmonar de Lewis, e também houve uma redução de formação de papiloma em um modelo de carcinogênese quimicamente induzido em células de pele, o que aponta que a variante *S-Eng* desempenha uma função anti-angiogênica (MEURER et al., 2014; PÉREZ-GÓMEZ et al., 2005), o que explicaria o porquê essa variante foi inibida durante o pico de hipóxia no nosso modelo.

*Nestina (NES)* é uma proteína de filamento intermediário do tipo IV do citoesqueleto que foi primeiramente encontrado em células-tronco neuroepiteliais (BERNAL; ARRANZ, 2018; DUSART et al., 2018). Alguns estudos propuseram que *NES* é um biomarcador de angiogênese tumoral em diversos tipos de câncer (DUSART et al., 2018; ISHIWATA; MATSUDA; NAITO, 2011; MATSUDA; HAGIO; ISHIWATA, 2013; NOWAK et al., 2018). Porém, DUSART et al. (2018) apresentou evidências que *NES* é independente de angiogênese e não é biomarcador específico para células endoteliais associadas com tumores, uma vez que foi observado que a expressão do gene *NES* tanto em células endoteliais de tecidos normais quanto tumorais, e sua expressão não foi associada a um prognóstico ruim em câncer urotelial, câncer pulmonar, câncer do estômago e glioma. No nosso modelo, a variante *NES-203* estava regulada positivamente no dia 12 durante o pico de hipóxia, o que vai ao encontro dos estudos que apontam para uma função pró-angiogênica para esse gene.

*PIWI (P-element-induced wimpy testis)* pertence à família de proteínas argonautas, e tem como principal função se ligar a RNAs que interagem com *PIWI* (piRNAs), formando o complexo *PIWI-piRNA* que regula a expressão gênica a nível epigenético e pós-transcricional (HENAOUÏ et al., 2017; ILIEV et al., 2016; WANG et al., 2016). Essas proteínas são mais expressas em células germinativas, mas estudos recentes têm observado sua presença em células somáticas (HENAOUÏ et al., 2017; WANG et al., 2016). A expressão de *PIWIL4* foi associada com a formação e progressão com alguns tipos de câncer, como câncer de mama (WANG et al., 2016), carcinoma renal (ILIEV et al., 2016), câncer de ovário (CHEN; LIU; XU, 2013), câncer

gástrico (WANG et al., 2012) e câncer na cervical (SU et al., 2012). Em câncer de mama, por exemplo, a inibição de PIWI4L reduziu a migração e proliferação celular, e aumentou a apoptose (WANG et al., 2016).

Colágeno é uma família de proteínas de grande diversidade, com um total de 28 tipos diferentes, tendo em comum pelo menos um domínio tripla-hélice (BIELAJEW; HU; ATHANASIOU, 2020; RICARD-BLUM, 2011; XU et al., 2019). Essas proteínas são as mais abundantes na ECM e elas possuem diversas funções relacionadas com câncer (DISCHER et al., 2017; XU et al., 2019), como promover metástase e regular a imunidade tumoral (DISCHER et al., 2017; XU et al., 2019; YAMAUCHI et al., 2018). Colágeno tipo 2 alfa 1 (*COL2A1*) foi regulado positivamente em células de repovoamento de tumor (TRC) de melanoma linha B16F1, e assim foi proposto como um novo biomarcador, e um potencial alvo terapêutico, para melanoma TRC (TALLURI et al., 2020). As vias de sinalização JAK/STAT, hipóxia e *AKT* também estavam reguladas positivamente em TRC (TALLURI et al., 2020), o que pode indicar que a expressão de *COL2A1* pode ser regulada por vias similares às que promovem angiogênese. O alto nível de expressão de *COL2A1*, juntamente com inibição do pseudogene *SLC6A10P*, foram associados com a recorrência tardia de câncer de ovário (GANAPATHI et al., 2016).

*TNFAIP2* (*Tumor necrosis factor alpha induced proteins*) é um gene envolvido principalmente com processos de inflamação, migração celular, formação de nanotubo e angiogênese, e é expresso em abundância em células tumorais (CHEN et al., 2010; XIE; WANG, 2017). Em carcinomas de nasofaringe foi verificado que a expressão de *TNFAIP2* é correlacionada com a densidade de microvasos na célula tumoral (CHEN et al., 2010), o que antes já se havia observado *in vitro* para formação de tubos capilares endoteliais (CHEN et al., 2010; SARMA et al., 1992). A expressão elevada de *TNFAIP2* foi associada com o aumento da proliferação de células tumorais e metástase em câncer de nasofaringe e de células escamosas do esôfago (CHEN et al., 2010; XIE; WANG, 2017).

Como apresentado na seção 3.2.4 , as variantes *ENG-202* (Figura 15d) e *ADGRG6-201* (Figura 15f) foram as que mais divergiram dos resultados obtidos com RT-PCR. Como se trata de uma quantificação a nível de variante, em alguns casos uma variante difere da outra por alguns nucleotídeos apenas (uma das variantes pode apresentar um éxon menor), sendo todo restante da sequência idêntica a outra variante. Para esses casos, o alinhador pode ter dificuldade em definir de qual variante o *read* foi gerado. Mesmo usando quantificadores como RSEM, que conseguem lidar com *reads* ambíguos, para esses casos podem ocorrer algumas imprecisões na estimativa de abundância da variante. Realizar sequenciamento do tipo *paired-end*, ou

utilizar tecnologias que permitem gerar *reads* longos, pode auxiliar na precisão dos alinhadores e na estimativa dos quantificadores. Apesar desses casos, de maneira geral, as variantes escolhidas foram validadas com sucesso, pois existe uma correlação significativa entre os LFCs obtidos com RNA-Seq e RT-PCR (Figura 14).

## 4 Análise dos ncRNAs diferencialmente expressos durante angiogênese patológica

Por um período, regiões não codificadoras de proteína eram chamadas de "DNA lixo", e em humanos essas regiões correspondem a cerca de 98% do genoma (DHANOA et al., 2018). Com o passar dos anos, notou-se que essas regiões não só eram expressas, como também essas moléculas auxiliavam na regulação de processos biológicos e no ajuste fino da expressão gênica (VIEIRA et al., 2018), e então essas moléculas passaram a ser conhecidas como ncRNA.

Um ncRNA é definido como uma molécula de RNA sem potencial para codificar proteína (DHANOA et al., 2018). Além dos RNAs ribossomais (rRNA) e os RNAs transportadores (tRNA), que são os principais ncRNAs envolvidos na síntese proteica, existem também os lncRNAs, RNAs com mais de 200 nucleotídeos, e ncRNAs pequenos (tamanho menor que 200 nt), formado pelos miRNAs, pequenos RNAs nucleares (snoRNAs) e os *piwi interacting RNA* (piRNA) (VIEIRA et al., 2018).

No presente capítulo nós apresentamos as análises de identificação, quantificação e análise de expressão diferencial de 3 tipos de ncRNAs, os miRNAs (seção 4.1 ), os circRNAs (seção 4.2 ) e os lncRNAs (seção 4.3 ). No capítulo 5 nós realizamos a integração desses resultados com a rede de co-expressão de RNAs, a fim de se obter a rede de regulação gênica para angiogênese patológica. A Figura 16 apresenta um resumo do *pipeline* executado, o qual será mais bem detalhado no atual capítulo e no próximo.

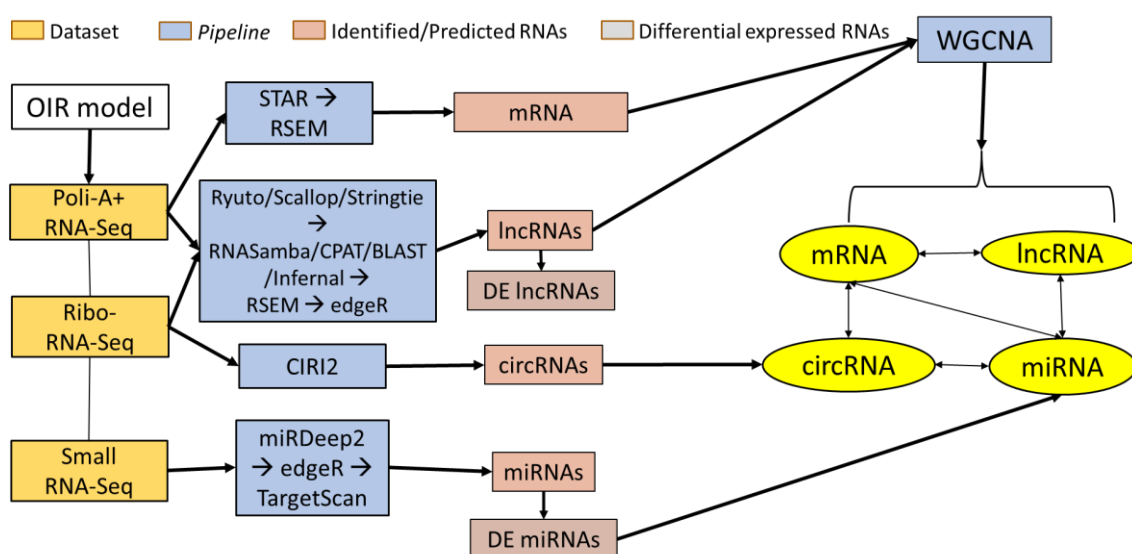


Figura 16: Resumo do pipeline executado para construção da rede de regulação gênica da angiogênese patológica.

## 4.1 miRNAs

Os miRNAs são ncRNAs com tamanho aproximado de 22 pares de base. "Cerca de metade dos miRNAs conhecidos são originados de íntrons e de éxons (mas são poucos) do gene hospedeiro. Os restantes são originados de regiões intergênicas, independente de gene hospedeiro, com seus próprios promotores" (O'BRIEN et al., 2018). A biogênese de miRNAs é classificada em canônica e não-canônica (Figura 17). Quando o miRNA é transcrito, parte da sequência se dobra (*hairpin*) formando o pri-miRNA. O *hairpin* então é reconhecido pelo complexo formado pela endonuclease Drosha e DGCR8, e é clivado, formando o pré-miRNA. Então, o pré-miRNA é enviado para o citoplasma, onde ele será processado pela endonuclease Dicer, que irá clivar a região do laço (loop), resultando em um RNA fita-dupla chamado de miRNA-miRNA duplex. Essa estrutura contém o miRNA maduro, que será alocado junto a família de proteínas argonauta (AGO2), formando o complexo de silenciamento induzido por miRNA (miRISC) (MCCREIGHT et al., 2017; O'BRIEN et al., 2018). Essa via caracteriza a via canônica, sendo essa a mais dominante (O'BRIEN et al., 2018). A via não-canônica pode ser dividida em 2: a via independente de Drosha/DGCR8, onde o pri-miRNA é diretamente exportado para o citoplasma e processado pela Dicer; e a via independente de Dicer, onde o pré-miRNA é alocado diretamente ao AGO2, sem ser processado pela Dicer (O'BRIEN et al., 2018).

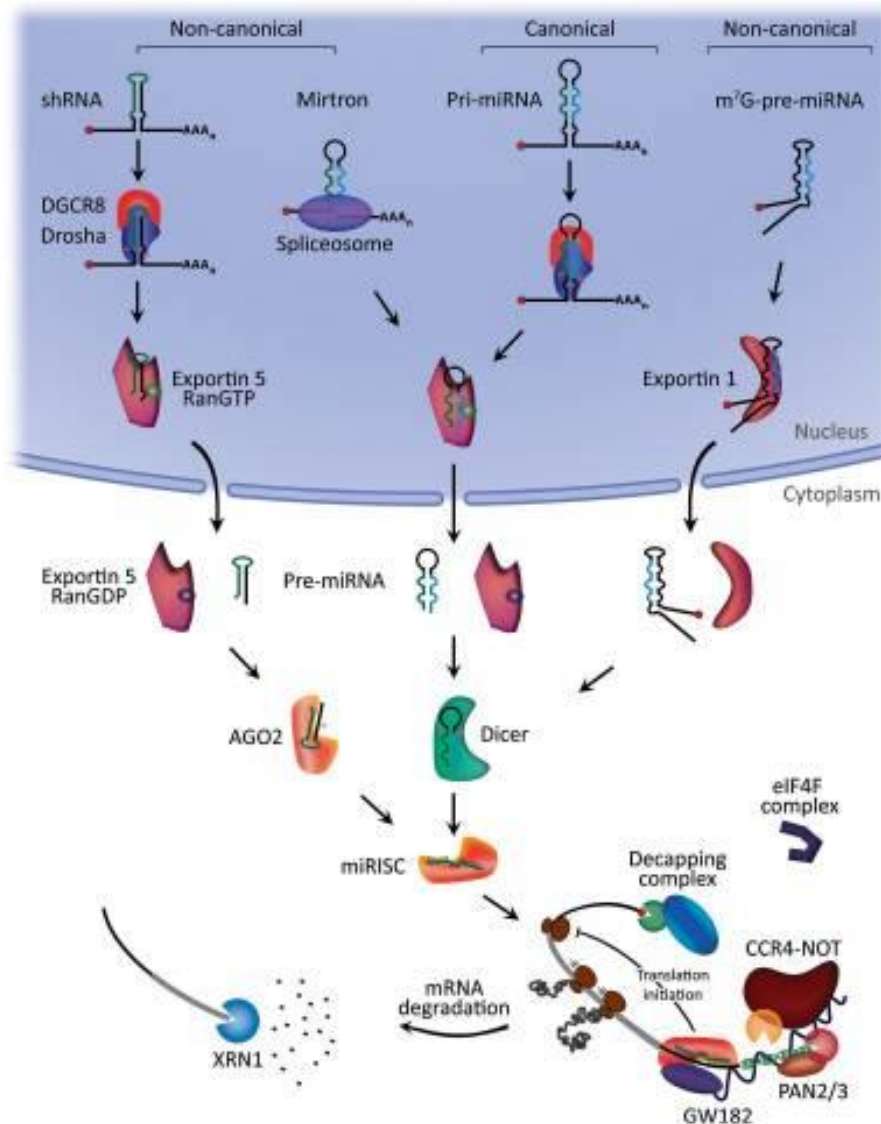


Figura 17: As Vias canônica e não canônica da biogênese de miRNAs. A via canônica envolve o processamento do pri-miRNA pela endonuclease Drosha/DGCR8, e o produto dessa reação (Pré-miRNA) é posteriormente processado no citoplasma pela endonuclease Dicer, gerando assim o miRNA maduro. Nas vias não-canônicas o miRNA será processado apenas por uma das endonucleases citadas anteriormente. Ambas as vias resultam na formação do complexo miRISC que irá atuar tanto na inibição da tradução ou na degradação do mRNA alvo. Figura retirada de O'BRIEN et al. (2018) – Figura 1.

Os miRNAs irão atuar principalmente no controle da regulação gênica, se ligando a região 3' UTR dos mRNAs alvo, reprimindo sua tradução, e removendo a cauda poli-A e o 5' CAP, mediado pelo miRISC (HA; KIM, 2014; O'BRIEN et al., 2018). Também existem evidências de que miRNAs também podem se ligar a outras regiões do mRNA (5' UTR, região codificante e região promotora), podendo também promover a indução da expressão do mRNA alvo, porém são casos mais raros de acontecer (O'BRIEN et al., 2018). Os miRNAs também podem ser secretados para fluidos extracelulares (como plasma e soro, saliva, urina, entre outros), associados a vesículas ou proteínas, e atuar como sinalizador celular (O'BRIEN et al., 2018). Os miRNAs extracelulares são



potenciais biomarcadores para doenças, como câncer e diabetes tipo 2 (CHEN et al., 2008).

Sabe-se que miRNAs também atuam na regulação da angiogênese, sendo esses chamados de angiomiR (TIWARI; MUKHERJEE; DIXIT, 2018). Eles podem atuar, até onde se conhece, de 2 maneiras: se ligando a reguladores negativos da angiogênese, promovendo a vascularização, ou se ligando a reguladores positivos, inibindo a angiogênese (TIWARI; MUKHERJEE; DIXIT, 2018). O miR-126 tem propriedades pro-angiogênicas, reprimindo reguladores negativos da via de sinalização do *VEGF*, como o *SPRED1* (*sprouty-related protein*) e *PIK3R2* (*phosphoinositol-3 kinase regulatory subunit 2*) (FISH et al., 2008). O miR-210 também possui propriedades pro-angiogênicas, estimulando a formação de capilares e a quimiotaxia de células endoteliais em resposta ao *VEGF*. O mecanismo utilizado pelo miR-210 é a inibição do gene Ephrin-A3 (*ENFA-3*) (FASANARO et al., 2008; TIWARI; MUKHERJEE; DIXIT, 2018). Em carcinoma hepatocelular, a inibição do miR-195 foi associada com o aumento da angiogênese, metástase, e uma pior sobrevida livre de recorrência, tendo, portanto, o miR-195 propriedades anti-angiogênicas. Foi verificado em seguida que o fator pró-angiogênico *VEGF*, e os fatores pro-metastáticos *VAV2* e *CDC42* são alvos diretos do miR-195 (TIWARI; MUKHERJEE; DIXIT, 2018; WANG et al., 2013b).

Algumas revisões (LANDSKRONER-EIGER; MONEKE; SESSA, 2013; SUÁREZ; SESSA, 2009; TIWARI; MUKHERJEE; DIXIT, 2018) apresentam estudos que exploraram as funções de miRNAs como agentes pró- e anti-angiogênicos, mostrando ser um tópico bem estudado e que ainda continua sendo explorado, e a principal justificativa é que os padrões de expressão dos miRNAs são específicos para cada célula/tecido e que mais estudos com diferentes fontes de células endoteliais com uma ampla cobertura de miRNAs irá fornecer uma compreensão mais completa na regulação de diferentes aspectos da angiogênese (LANDSKRONER-EIGER; MONEKE; SESSA, 2013). Por esse motivo, nós implementamos um *pipeline* para predição de miRNAs no modelo OIR, a fim de identificar quais miRNAs podem estar atuando na regulação da angiogênese patológica.

#### 4.1.1 Métodos

Para a análise dos miRNAs, nós utilizamos os dados de sequenciamento de miRNAs, que são compostos por 7 amostras, 1 para cada tempo (12, 12.5 [exclusivo do grupo retinopático], 15 e 17) e condição (controle e retinopatia). Para predição, anotação e quantificação de miRNAs utilizamos o software miRDeep2 (FRIEDLÄNDER et al., 2012). Uma vez identificados e quantificados, verificamos quais

miRNAs estavam DE com o edgeR (ROBINSON; MCCARTHY; SMYTH, 2010). No entanto, a ausência de réplicas biológicas impossibilita realizar esta análise, pois é a partir das réplicas que o software consegue estimar a dispersão dos dados e consequentemente calcular o coeficiente de variação biológica (BCV), que captura todas as fontes de variação técnica e biológica entre réplicas biológicas (MCCARTHY; CHEN; SMYTH, 2012). Quanto maior o BCV, maior será a variação da expressão dos mesmo genes entre as réplicas biológicas, e portanto maior deverá ser a diferença de expressão de um gene entre os grupos/tratamentos para que ela seja significativa e então o gene ser considerado como DE. Para estas situações, o pacote edgeR oferece a possibilidade de apontar manualmente o BCV, inclusive até fornece algumas sugestões de valores (0,4 para amostras humanas; 0,1 para animais geneticamente idênticos; 0,01 para réplicas técnicas) (CHEN et al., 2020). Para determinar o valor de BCV para a nossa análise, decidimos utilizar os dados de sequenciamento das bibliotecas poli A+ (bibliotecas enriquecidas com mRNA), as quais possuem 2 réplicas biológicas e no mínimo 6 réplicas técnicas por condição (com exceção do dia 15, onde as amostras do grupo controle e retinopático não possuem réplicas técnicas), e então calcular o BCV dessas amostras com o edgeR. O valor de BCV obtido foi de 0,17 aproximadamente, o que é bem próximo do valor esperado para amostras de animais modelos geneticamente idênticos. O passo seguinte foi rodar o software edgeR com os dados de sequenciamento de miRNA, definindo o BCV com o valor de 0,17. Os miRNAs com  $FDR \leq 0,05$  foram considerados como DE.

Para predição dos alvos dos miRNAs DE, nós utilizamos o algoritmo do TargetScan (AGARWAL et al., 2015), utilizando como referência as sequências 3' UTR dos transcritos de camundongo, obtidos com a ferramenta BioMart do ENSEMBL (7 de março 2021). Como a inibição da tradução, ou degradação, de um mRNA é principal mecanismo de regulação do miRNA, para reduzir o número de falsos positivos, consideramos que um transcrito só é um alvo válido se: 1) a região 3' UTR possuir 2 ou mais regiões de ligações de miRNA; e 2) O miRNA precisa ter um  $LFC \geq 1$  e o transcrito alvo precisa ter o  $LFC \leq -1$ , ou vice-versa.

#### 4.1.2 Expressão diferencial dos miRNAs

Ao todo foram identificados 390 miRNAs, dos quais 313 já são conhecidos e 77 são novos. Com a análise de expressão diferencial, identificamos 99 miRNAs DE, dos quais 91 já são conhecidos e 8 são novos (Tabela 8 e Apêndice IV: Lista de miRNAs diferencialmente expressos (DE)). O perfil de expressão desses miRNAs se encontra na Figura 18.

Tabela 8: Total de miRNAs DE em cada um dos dias em que o experimento foi conduzido.

	R12 x P12	R12.5 x P12	R15 x P15	R17 x P17
Regulados positivamente	42	26	0	29
Regulados negativamente	24	9	1	29

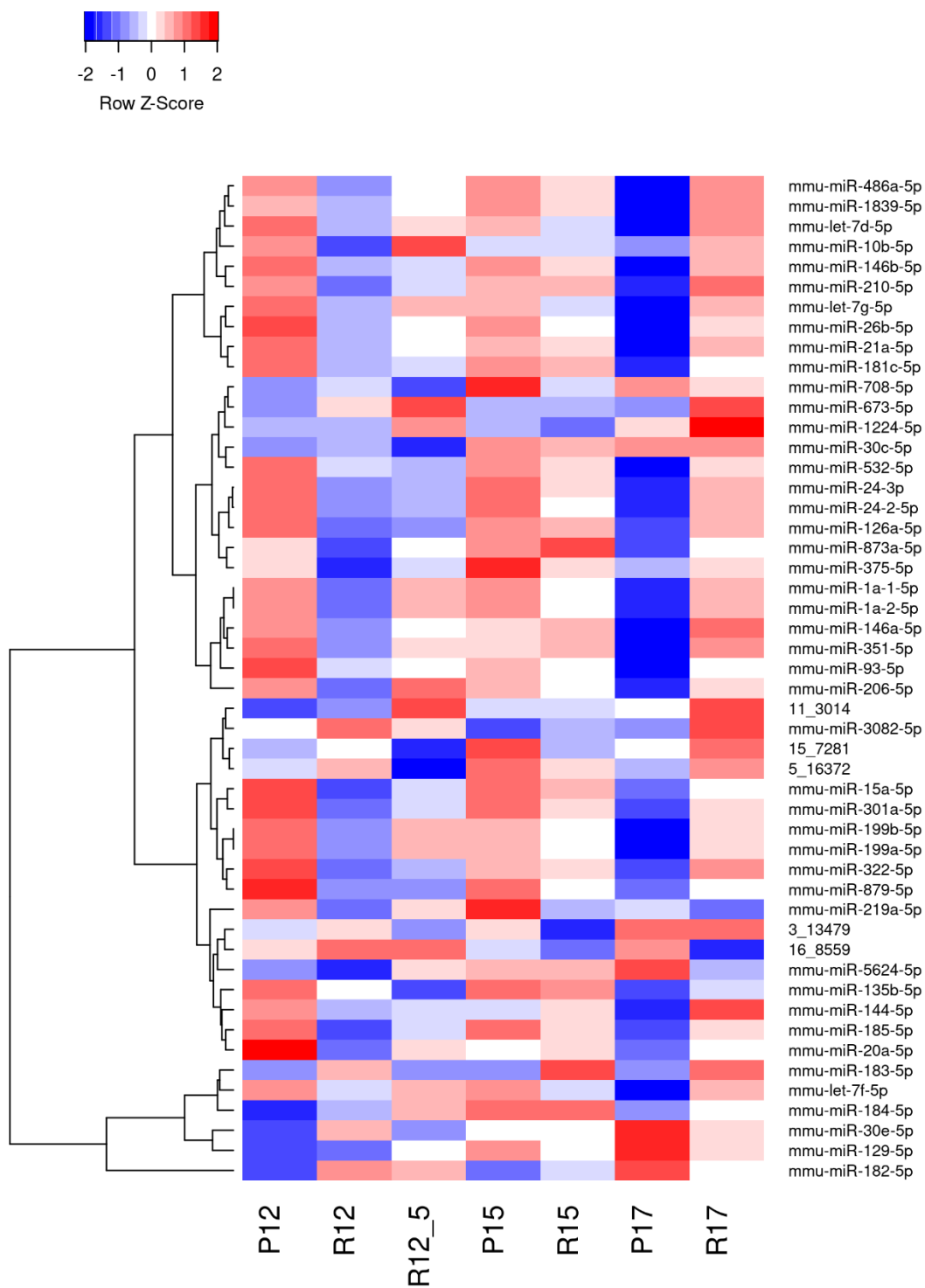


Figura 18: Heatmap do perfil de expressão dos 25 miRNAs DE com maior LFC e dos 25 miRNAs DE com menor LFC, considerando todas as comparações realizadas (R12 vs P12, R12.5 vs P12, R15 vs P15 e R17 vs P17). A letra P representa as amostras do grupo controle, e a letra R representa as amostras do grupo retinopático. É possível observar que a maioria dos miRNAs apresentados no heatmap tem sua expressão reduzida no dia 12, quando comparamos R12 vs P12, e que depois no dia 17 esse padrão é invertido.

### 4.1.3 Predição dos alvos

Dos 99 miRNAs DE, 97 possivelmente regulam a expressão de 2.358 transcritos. Para verificar o provável papel desses miRNAs na regulação da angiogênese, nós realizamos uma análise de enriquecimento funcional com Gene Ontology (ASHBURNER et al., 2000) (Tabela 9) e vias do KEGG (KANEHISA; GOTO, 2000) (Tabela 10) para os genes alvos desses miRNAs utilizando a ferramenta DAVID (SHERMAN et al., 2022).

Tabela 9: Os 10 termos GO (ontologia “*Biological process*”) mais enriquecidos para os genes alvos dos miRNAs DE. A análise de enriquecimento foi realizada com a ferramenta web DAVID, e o método de correção do p-valor utilizado foi a de Benjamini-Hochberg.

<b>Termo GO</b>	<b>Contagem de genes</b>	<b>p-valor</b>	<b>Correção do p-valor</b>
Angiogenesis	60	9,0E <sup>-13</sup>	4,6E <sup>-9</sup>
Positive regulation of cell migration	51	5,4E <sup>-11</sup>	1,4E <sup>-7</sup>
Positive regulation of cell proliferation	91	1,7E <sup>-8</sup>	2,9E <sup>-5</sup>
Cell adhesion	83	3,3E <sup>-8</sup>	4,2E <sup>-5</sup>
Positive regulation of gene expression.	71	7,6E <sup>-8</sup>	7,8E <sup>-5</sup>
Blood vessel development	23	1,9E <sup>-7</sup>	1,6E <sup>-4</sup>
Positive regulation of transcription from RNA polymerase II promoter	140	2,1E <sup>-7</sup>	1,6E <sup>-4</sup>
Positive regulation of smooth muscle cell proliferation	24	4,5E <sup>-7</sup>	2,5E <sup>-4</sup>
Negative regulation of transcription from RNA polymerase II promoter	108	5,5E <sup>-7</sup>	2,5E <sup>-4</sup>
Protein phosphorylation	90	5,6E <sup>-7</sup>	2,5E <sup>-4</sup>

Tabela 10: As 10 vias do KEGG mais enriquecidas para os genes alvos dos miRNAs DE. A análise de enriquecimento foi realizada com a ferramenta web DAVID, e o método de correção do p-valor utilizado foi a de Benjamini-Hochberg.

Via do KEGG	Contagem de genes	p-valor	Correção do p-valor
Focal adhesion	55	2,2E <sup>-13</sup>	6,2E <sup>-11</sup>
PI3K-Akt signaling pathway	75	1,1E <sup>-12</sup>	1,5E <sup>-10</sup>
ECM-receptor interaction	30	2,4E <sup>-10</sup>	2,2E <sup>-8</sup>
Pathways in cancer	75	4,1E <sup>-10</sup>	2,8E <sup>-8</sup>
Osteoclast differentiation	35	2,2E <sup>-9</sup>	1,5E <sup>-7</sup>
Proteoglycans in cancer	44	6,4E <sup>-8</sup>	3,0E <sup>-6</sup>
Rap1 signaling pathway	42	2,1E <sup>-6</sup>	8,4E <sup>-5</sup>
Dilated cardiomyopathy	23	2,5E <sup>-6</sup>	8,4E <sup>-5</sup>
Transcriptional misregulation in câncer	35	3,0E <sup>-6</sup>	8,4E <sup>-5</sup>
Small cell lung cancer	23	3,1E <sup>-6</sup>	8,4E <sup>-5</sup>

#### 4.1.4 Discussão

O estudo de miRNA do modelo OIR permitiu identificar 99 miRNAs DE na angiogênese patológica, além, de 77 novos miRNAs. Um dos limitantes do nosso estudo foi a falta de réplicas biológicas. Mesmo assim, com a estratégia adotada na seção 4.1.1, observamos que os miRNA identificados têm como alvos moléculas sabidamente envolvidas em angiogênese, o que de certa forma valida nossa estratégia. Porém, estudos de validação experimental serão essenciais para confirmar nossos resultados.

Nós procuramos na literatura o que já se conhecia para o miRNA com maior LFC e para o miRNA com menor LFC entre todos os dias do experimento. O miR-144-5p foi o miRNA com maior LFC. Esse miRNA foi identificado como DE no dia 17, com expressão aproximadamente 7 vezes maior no grupo retinopático. O miR-144-5p atua como um supressor tumoral, inibindo proliferação, migração e invasão celular em câncer renal, câncer de bexiga e em células endoteliais de veia umbilicais humanas (HUVEC) (FU et al., 2020; MATSUSHITA et al., 2015; YAMADA et al., 2018). A baixa expressão desse miRNA foi associado significativamente com um prognóstico ruim para pacientes com câncer renal (YAMADA et al., 2018). O miRNA com o menor LFC foi o miRNA miR-20a-5p, com expressão quase 8 vezes menor no grupo retinopático comparado ao controle. Em neuroblastoma, o miR-20a-5p regula negativamente o

gene *ATG7* se ligando a região 3' UTR do gene, inibindo autofagia e proliferação celular, e promovendo apoptose (YU et al., 2018). Por outro lado, outros estudos atribuem a esse miRNA um papel pró-tumor por promover a proliferação, migração e invasão celular em câncer de bexiga e em carcinomas de células escamosas da cabeça e do pescoço (WU et al., 2018; YANG; CHEN; LIU, 2021). Ambos miRNAs são bons candidatos para validação em bancada para confirmar seus níveis de expressão durante angiogênese patológica e para realizar ensaios que confirmem seus possíveis alvos nesse processo.

## 4.2 circRNAs

Os circRNAs já foram observados a algumas décadas atrás, mas se acreditava que se tratava de erros de splicing com baixo potencial funcional (CHEN, 2016; KRISTENSEN et al., 2019). Só após o desenvolvimento das novas tecnologias de sequenciamento de RNAs não poliadenilados é que se observou que circRNAs possuem um padrão de expressão tecido-específico, apesar de ainda ser em menor nível que os mRNAs lineares (CHEN, 2016; KRISTENSEN et al., 2019).

Os circRNAs são produzidos a partir do *backsplicing* de pré-mRNAs, e esse processo é dependente da maquinaria de *splicing* canônica. A hipótese é que a complementaridade de elementos repetidos invertidos nas regiões dos íntrons flanqueadores, ou a dimerização por proteínas ligantes de RNA (RBPs) que se ligam a motivos nas regiões dos íntrons flanqueadores, auxiliam na circularização, aproximando as regiões doadoras das regiões receptoras (Figura 19a) (CHEN, 2016; KRISTENSEN et al., 2019). Outra forma que dá origem aos circRNAs é durante a formação do laço (*lariat*) durante *éxon skipping* no AS, onde os éxons que ficam no interior do laço sofrem *backsplicing*, ou quando laços intrônicos escapam da desramificação e acabam originando circRNAs intrônicos (ciRNAs) (Figura 19b) (KRISTENSEN et al., 2019).

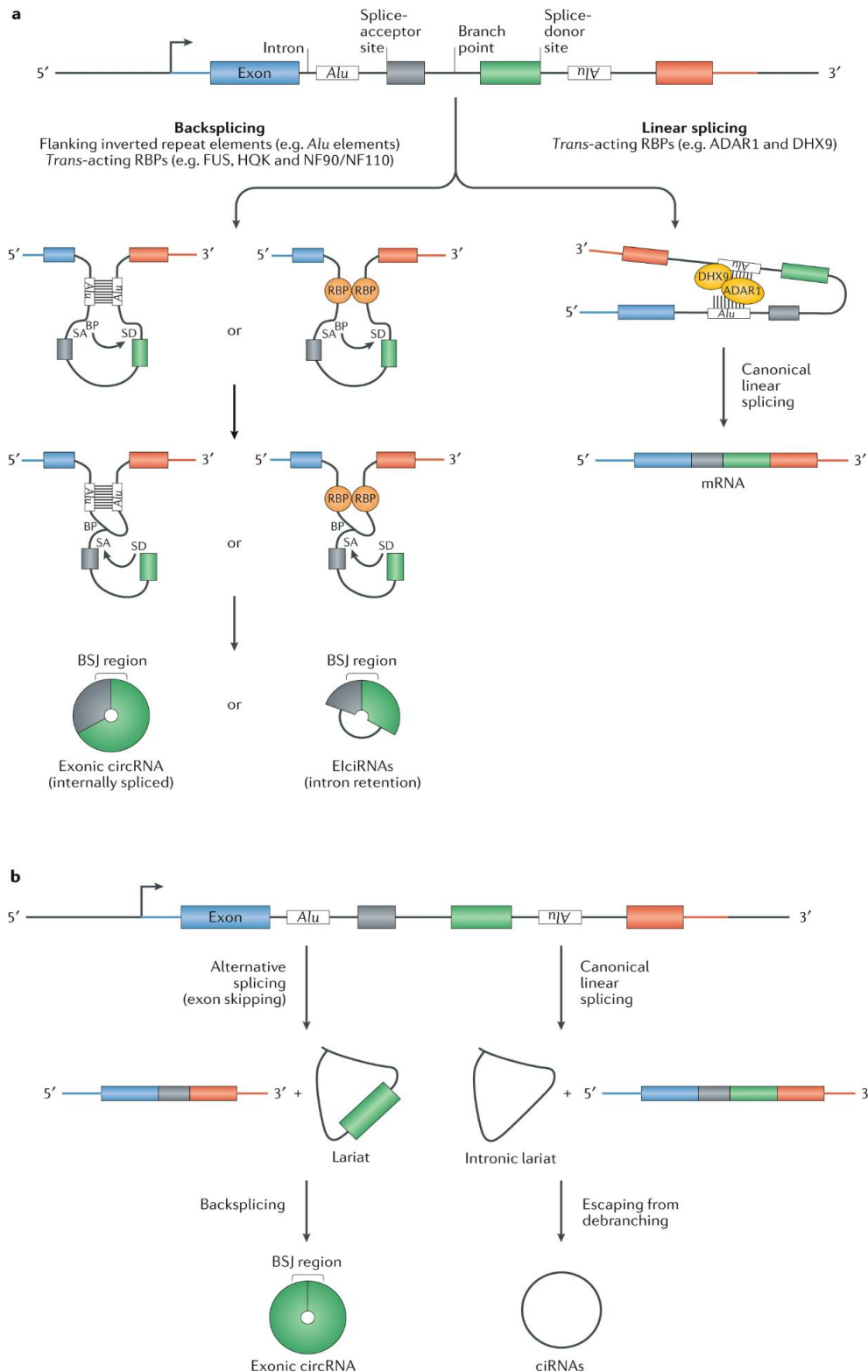


Figura 19: Vias de biogênese de circRNAs: a) regiões de repetição nos íntrons flanqueadores ou os RBPs auxiliam na circularização do RNA aproximando as porções receptoras e doadoras. b) Formação do laço (*lariat*) durante AS, podem ocorrer o *backsplicing* dos éxons internos ou a circularização do *lariat*, formando um circRNA intrônico. Figura retirada de KRISTENSEN et al. (2019) – Figura 1.

Foi observado que ao desacelerar o processamento do pré-mRNA, o processo de *backsplicing* é favorecido, aumentando a concentração de circRNAs em relação aos



RNAs lineares (KRISTENSEN et al., 2019). A maior parte dos circRNAs são formados por éxons (um ou múltiplos), originados de genes codificantes de proteína, sem CAP 5' e a cauda poli-A. Existem circRNAs que retêm os íntrons, sendo esses conhecidos como éxon-íntron circRNAs. Éxon-íntron circRNAs e ciRNAs permanecem no núcleo, enquanto os circRNAs exônicos são exportados para o citoplasma. A estrutura circular garante maior estabilidade à molécula, pois a protege da degradação por exonucleases.

A função mais conhecida dos circRNAs é a inibição de miRNAs, em que os miRNAs se ligam ao circRNA ao invés do alvo RNA linear, atuando como uma "esponja". Mas a interação de circRNAs com o complexo RBPs, sugere que eles também possam atuar como "esponja" de proteínas, inibindo suas funções, ou atuando como um reservatório, permitindo uma rápida resposta a um estímulo extracelular. O próprio processo de formação de circRNA pode ser um mecanismo de controle de expressão de mRNAs lineares, uma vez que o processo de *backsplicing* dos circRNAs compete com o *splicing* do pré-mRNA, diminuindo os níveis de mRNA linear (CHEN, 2016). Apesar da grande maioria dos circRNAs atuarem como RNAs não codificantes, alguns são capazes de serem traduzidos mesmo sem o CAP 5', apesar de os peptídeos oriundos de circRNAs ainda não terem sua função definida.

Alguns estudos apontam que alguns circRNAs podem atuar na regulação da angiogênese em alguns tipos celulares humanos (LI; MA; YU, 2017; LIU et al., 2017b). Um deles é o circHIPK3, cuja expressão foi induzida em retinas diabéticas e células endoteliais de retina seguida por estressores relacionados com diabetes mellitus. Descobriu-se que o circHIPK3 atua como uma "esponja" para o miR-30a-3p, sequestrando-o e inibindo sua atividade, o que leva ao aumento da expressão de *VEGF-C*, *FZD4* e *WNT2*. Em um ensaio *in vivo*, o circHIPK3 foi silenciado e se observou uma diminuição da disfunção vascular na retina, devido a inibição da formação de capilares acelulares na retina e inflamação (SHAN et al., 2017). Para o modelo murino OIR foi realizado o silenciamento do circRNA *cZNF609*, e se observou que o processo de angiogênese foi inibido *in vivo*. O provável mecanismo é o sequestro do miR-615-5p, inibindo sua atividade, o que leva ao aumento da expressão do gene *MEF2A* (LIU et al., 2017a). O papel dos circRNAs na regulação da angiogênese ainda precisam ser explorados, uma vez que os trabalhos que se propõem a investigá-los são relativamente recentes, mas apresentam resultados promissores.

#### 4.2.1 Métodos

Para essa análise utilizamos os dados de sequenciamento das bibliotecas Ribo-. Os *reads* de todas as bibliotecas foram agrupados em um único arquivo e processadas simultaneamente, garantindo uma maior cobertura, auxiliando a ferramenta na identificação de novos circRNAs. Os *reads* então foram alinhados ao genoma utilizando o BWA (LI; DURBIN, 2009), e para a predição dos circRNAs foi utilizado o *software* CIRI2 (GAO; ZHANG; ZHAO, 2018). Os íntrons dos circRNAs exônicos foram removidos da sequência pseudolinar, conforme descrito em LIU et al. (2019). Utilizando o banco de dados circBase (GLAŽAR; PAPAVALASILEIOU; RAJEWSKY, 2014) e o BLAST (ALTSCHUL et al., 1990) para fazer os alinhamentos, nós verificamos quais circRNAs já haviam sido descritos anteriormente por outros estudos. Para verificar quais circRNAs podiam estar atuando como “esponjas” para miRNAs, utilizamos o TargetScan para predizer sítios de ligação de miRNAs nas sequências pseudolineares.

Para análise de expressão diferencial dos circRNAs consideramos apenas os *reads* que mapeiam na região chamada de “*back-spliced junction*” (BSJ) (Figura 20), que são identificadas pelos *reads* em que a porção próxima a ponta 5' se alinha a um éxon/íntron de transcrito mais próximo da ponta 3', e a outra porção do *read* (mais próxima da ponta 3') se alinha a um éxon/íntron de um transcrito mais próximo da ponta 5' (KRISTENSEN et al., 2019). Após a quantificação, a análise de expressão diferencial foi realizada com edgeR.

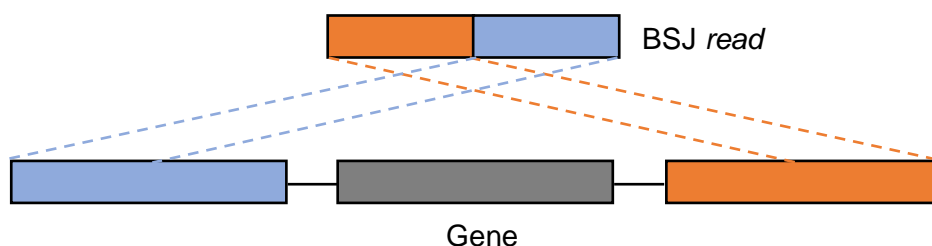


Figura 20: Representação de um BSJ *read* de um circRNA exônico. A porção 3' do *junction read* alinha com o éxon da ponta 5' do gene, enquanto a porção 5' do *junction read* alinha com o éxon da ponta 3' do gene. Figura baseada na figura 1A publicada em GAO et al. (2015).

#### 4.2.2 circRNAs preditos

Ao todo foram preditos 5691 circRNAs oriundos de 2484 genes parentais. A Figura 21 sumariza os tipos de circRNAs preditos nas retinas dos camundongos. Quase metade (~43%) dos circRNAs preditos estão presentes no circBase quando

consideramos apenas os alinhamentos que resultaram em 100% de identidade e cobertura (Figura 22).

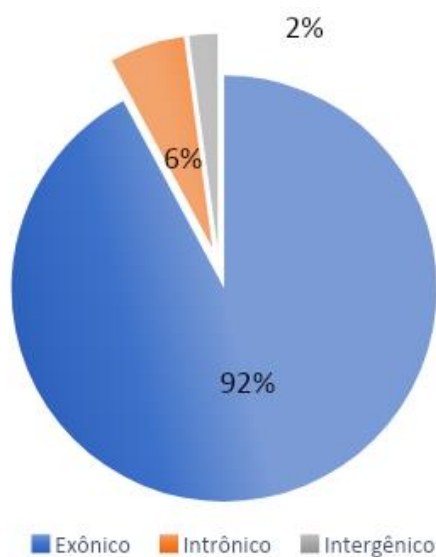


Figura 21: Classificação dos circRNAs dada pelo software CIRI2. Os circRNAs exônicos são aqueles em que seus limites ocorrem dentro da região de um éxon. Os circRNAs intrônicos são aqueles em que seu início ou fim se encontra na região de um íntron. Os circRNAs intergênicos são aqueles em que seus limites se encontram fora dos limites de um gene.

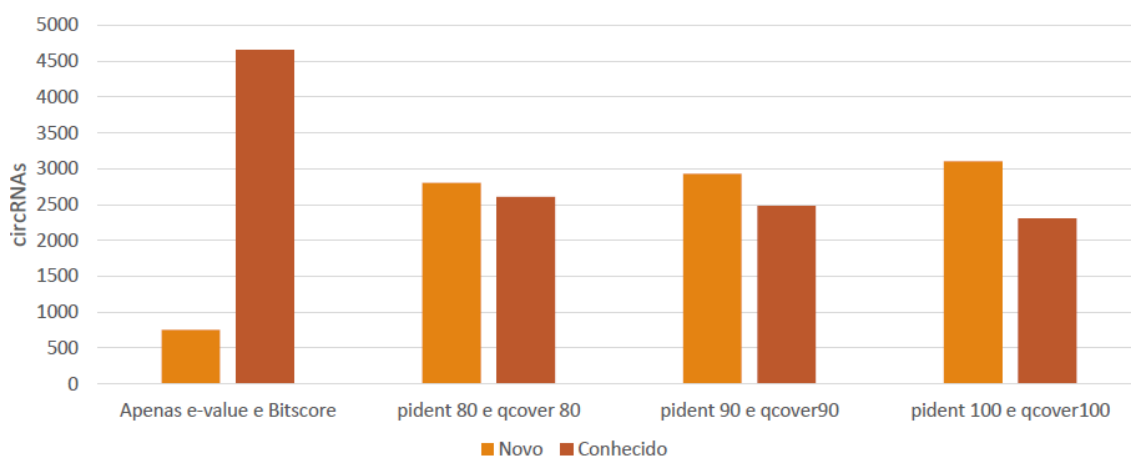


Figura 22: Resultados do BLAST entre os circRNAs preditos e o banco de dados do circBase. Os nomes *pident* e *qcover* se referem ao percentual de identidade e de cobertura do alinhamento, respectivamente.

#### 4.2.3 Expressão diferencial dos circRNAs

Devido à baixa cobertura de *reads* na região BSJ (Figura 23), a análise de expressão diferencial não apontou nenhum circRNA como DE.

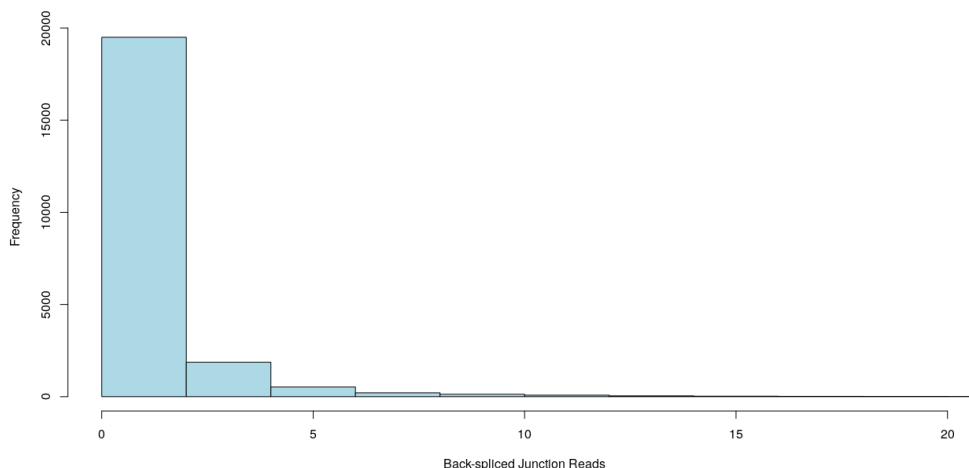


Figura 23: Histograma do número de *reads* BSJ alinhados nos circRNAs preditos. A quantificação foi realizada pelo próprio software CIRI2. A grande maioria dos circRNAs tem no máximo 2 BSJ *reads*, impossibilitando a análise de expressão diferencial.

#### 4.2.4 Predição de sítios de ligação de miRNAs

Encontramos sítios de ligação para todos os 99 miRNAs DE em 4646 circRNAs utilizando o TargetScan para fazer as predições. A Tabela 11 apresenta alguns exemplos de miRNAs que podem ser alvos de circRNAs preditos em nossos dados.

Tabela 11: Alguns dos circRNAs preditos e seus respectivos miRNAs alvos.

circRNA ID	Transcript ID	Gene	miRNA ID
1:52479115 52490632	ENSMUST00000069792	NAB1	mmu-miR-1968-5p
15:25800032 25801271	ENSMUST00000110457	MYO10	mmu-miR-181b-5p
18:35221976 35223486	ENSMUST00000042345	CTNNA1	mmu-miR-133a-5p
3:122747816 122748404	ENSMUST00000200389	PDE5A	mmu-miR-532-5p
3:122781494 122817999	ENSMUST00000200389	PDE5A	mmu-miR-383-5p
7:134724030 134876012	ENSMUST00000084488	DOCK1	mmu-miR-135b-5p
7:134782048 134788984	ENSMUST00000084488	DOCK1	mmu-miR-3099-5p
7:134847839 134876012	ENSMUST00000084488	DOCK1	mmu-miR-378a-5p

#### 4.2.5 Discussão

A partir dos dados de sequenciamento de RNA total com depleção de rRNA, nós conseguimos predizer 5691 circRNAs, e identificamos quase metade deles no banco de dados do circBase. No entanto, as características dos nossos dados de sequenciamento não nos permitiram explorar o potencial máximo do *pipeline* de predição dessas sequências por se tratar de um sequenciamento *single-end*. Os algoritmos de *predição* de circRNAs fazem melhor uso de sequenciamentos do tipo *paired-end* podendo aplicar filtros mais restritivos na hora de determinar se o circRNA predito é válido ou não. A informação do *read* par também auxilia na estimativa de abundância desses circRNAs, o que não foi bem-sucedido em nossas análises (Figura 23). Outro fator que contribuiu para esse cenário foi o fato de que a maioria dos circRNAs são significativamente menos abundantes que a forma linear canônica do gene parental (ENUKA et al., 2016). Hoje existem métodos de enriquecimento de circRNAs nas amostras para um melhor aproveitamento do sequenciamento para esses tipos de RNAs.

Nossos resultados mostraram que existe um potencial para exploração de circRNAs no modelo OIR, mas para se ter melhores resultados nessas análises, faz-se necessário utilizar métodos para enriquecer as amostras com circRNAs e realizar um novo sequenciamento do tipo *paired-end*.

### 4.3 lncRNAs

Os lncRNAs podem ser transcritos a partir de regiões intrônicas, intergênicas, exônicas e de regiões codificantes de proteínas distantes, pela RNA polimerase II (DHANOA et al., 2018), e são classificados em 5 tipos (Figura 24), de acordo com sua localização no genoma: senso, quando o lncRNA se encontra na mesma fita do gene codificante de proteína mais próximo; antisense, quando se encontra na fita oposta ao gene codificante de proteína mais próximo; bidirecional, quando se iniciam na região promotora do gene codificante de proteína; intrônico, quando são encontrados nos íntrons de genes codificantes de proteína; e intergênico, quando se encontram na região intergênica de dois genes codificantes de proteína (KUMAR; GOYAL, 2017). Eles ainda podem ser processados de forma semelhante aos mRNAs (recebem o CAP 5', cauda poli-A e sofrem AS) (DHANOA et al., 2018). Os lncRNAs possuem mecanismos de ação cis, quando seu alvo se encontra no mesmo *locus* gênico, e

trans, quando seu alvo se encontra em um *locus* gênico diferente (DHANOA et al., 2018).

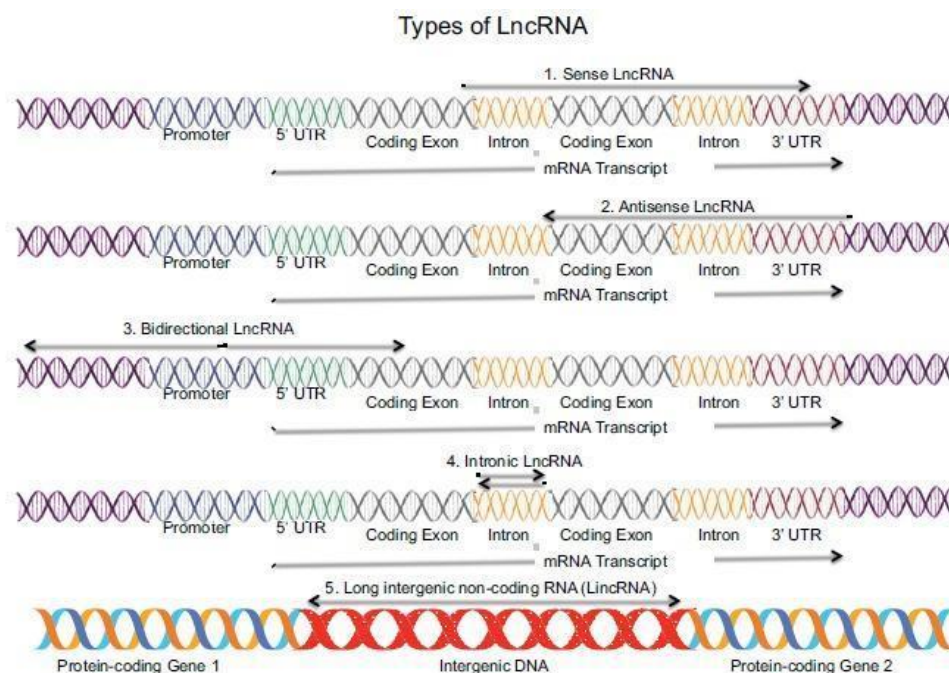


Figura 24: As 5 classes de lncRNAs: Senso, quando o lncRNA se encontra na mesma fita do gene codificante de proteína mais próximo; antisense, quando se encontra na fita oposta ao gene codificante de proteína mais próximo; bidirecional, quando se iniciam na região promotora do gene codificante de proteína; intrônico, quando são encontrados nos íntrons de genes codificantes de proteína; e intergênico, quando se encontram na região intergênica de dois genes codificantes de proteína. Figura retirada de KUMAR and GOYAL (2017) – Figura 1.

Algumas das funções desempenhadas pelos lncRNAs são: auxiliar na formação de estruturas quaternárias de algumas proteínas ou RNA regulatórios; precursores de miRNAs; regulação pós-transcricional de RNAs; e regulação epigenética, onde o estado de condensação da cromatina é alterado através da modificação de histonas, a fim de se silenciar ou ativar o gene alvo (KUMAR; GOYAL, 2017).

A revisão de Kumar and Goyal (2017) lista diversos estudos reportando evidências experimentais do papel dos lncRNAs na regulação de genes anti- e pró-angiogênicos. Um deles é o *LUNAR1*, um lncRNA regulado pelo oncogene *NOTCH1* em células-T humanas de leucemia linfoblástica aguda (T-ALL), e tem como função aumentar a expressão do receptor do fator de crescimento similar a insulina tipo 1 (*IGFR-1*) por meio do mecanismo de ativação cis, que por sua vez promove a angiogênese e sobrevivência da célula (KUMAR; GOYAL, 2017; TRIMARCHI et al., 2014). Outro lncRNA conhecido por promover angiogênese é o *HOTAIR*. Em células de câncer de nasofaringe, o silenciamento do *HOTAIR* suprimiu a proliferação celular e a angiogênese tanto *in vitro* quanto *in vivo*. O *HOTAIR* é capaz de inativar

diretamente a transcrição do *VEGFA*, assim como suprimir a expressão da proteína-78 regulada pela glicose (*GRP78*), regulando indiretamente a expressão de *VEGFA* e angiopoietina 2 (*ANG2*) (FU et al., 2016; KUMAR; GOYAL, 2017). Outros lncRNAs que apresentam funções importantes na regulação da angiogênese são *MEG3* (*maternally expressed gene 3*) e *MALAT1* (*metastasis associated lung adenocarcinoma transcript 1*). A inibição da expressão de *MEG3* foi associada com a supressão da migração de células endoteliais e angiogênese induzida por *VEGF* em HUVEC e em células de adenocarcinoma pulmonar (LI et al., 2019; RUAN et al., 2018; SHEN et al., 2022). Já *MALAT1* possui propriedades pró-angiogênicas, uma vez que foi observado que a supressão da expressão de *MALAT1* reduziu a angiogênese em células endoteliais na presença de altas concentrações de glicose (KUMAR; GOYAL, 2017; MICHALIK et al., 2014).

Em resumo, esses trabalhos demonstram que os lncRNAs têm um papel central na neovascularização e são, portanto, potenciais alvos terapêuticos no tratamento de doenças promovidas por angiogênese patológica. Claro, mais estudos nesse campo são necessários para elucidar os papéis de muitos desses lncRNAs, moléculas cujas funções moleculares na célula ainda são pouco conhecidas.

#### 4.3.1 Métodos

Nós realizamos a predição de novos transcritos utilizando dois conjuntos de dados distintos, os dados do sequenciamento da biblioteca poli-A+ e da biblioteca Ribo-, pois lncRNAs podem ou não possuir caudas poli-A (WILUSZ, 2016; ZHANG; YANG; CHEN, 2014). Para realizar a montagem dos transcritos utilizamos os softwares StringTie (PERTEA et al., 2015), Ryüto (GATTER; STADLER, 2019) e Scallop (SHAO; KINGSFORD, 2017). Com a função “merge” do StringTie, nós montamos um conjunto de transcritos não redundantes utilizando os resultados dos 3 softwares, tendo como referência a anotação do genoma do camundongo. Realizamos um BLAST (identidade e cobertura acima de 90%) contra as sequências de mRNAs e ncRNAs do GENCODE (GRCm38.p6) (FRANKISH et al., 2019; HARROW et al., 2012) de humano, camundongo e rato, e utilizamos o INFERNAL (NAWROCKI; EDDY, 2013) utilizando como referência os modelos de covariância depositados no RFAM (GRIFFITHS-JONES et al., 2003), para identificar se algum dos novos transcritos é semelhante a outro lncRNA já descrito. Também medimos o potencial codificador dos novos transcritos com as ferramentas CPAT (WANG et al., 2013a) e RNASamba (CAMARGO et al., 2020).

Uma vez identificados os potenciais novos lncRNAs, nós estimamos a abundância dos novos lncRNAs e dos já descritos em camundongo com o *software* RSEM. Para que não houvesse uma superestimativa da expressão dos lncRNAs, a quantificação foi feita simultaneamente com RNAs codificadores de proteína. Ao todo foram analisados 19.488 lncRNAs (8.996 já descritos e 10.492 novos) com o pacote edgeR para identificar quais lncRNAs estavam DE. Ao realizar a análise de componentes principais (PCA) para observar como as amostras se agrupavam, verificamos que elas se separavam de acordo com a técnica de sequenciamento (Figura 25). Entre as amostras da biblioteca Ribo- é ainda pior, pois como as réplicas foram sequenciadas em *lanes* diferentes na plataforma Illumina, isso induziu uma diferença entre as réplicas biológicas. Por isso decidimos prosseguir as análises apenas com as amostras da biblioteca poli-A+ que possuem uma profundidade de sequenciamento muito maior que a biblioteca Ribo-, e o sequenciamento de réplicas técnicas em diferentes *lanes* garantiu o não enviesamento dos dados observado na biblioteca Ribo-

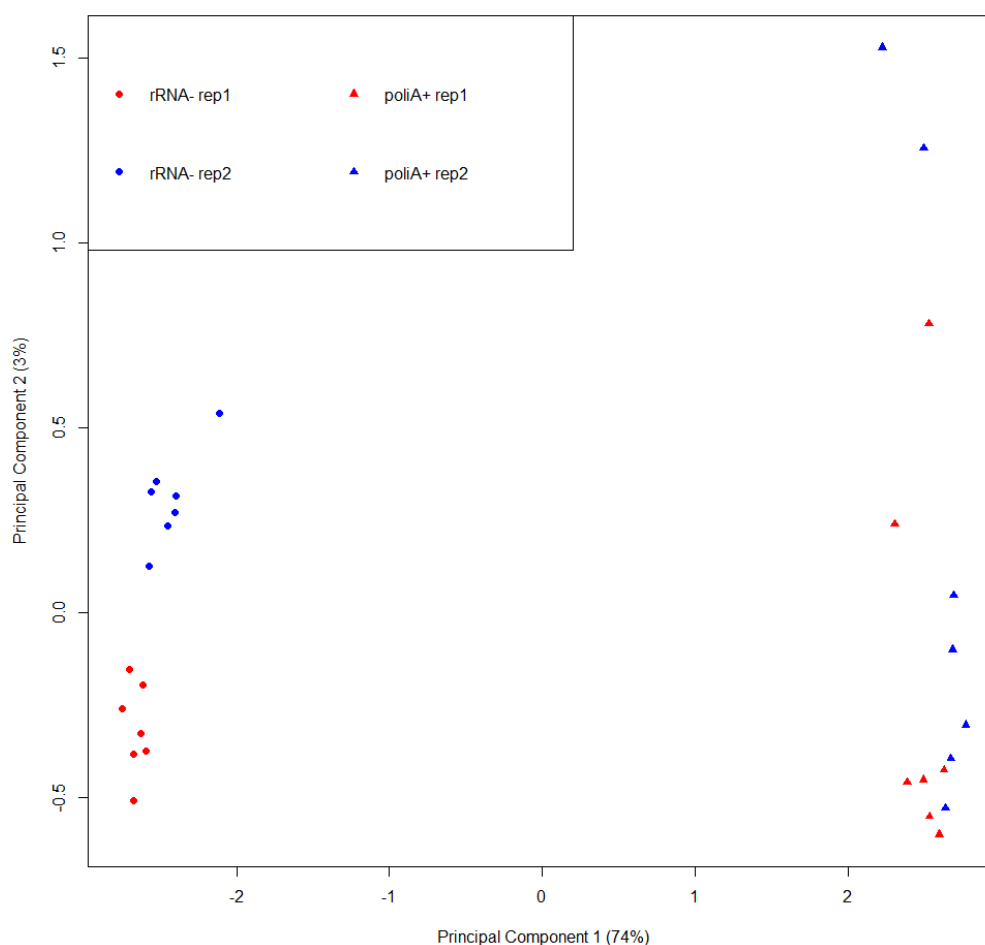


Figura 25: Análise de componentes principais (PCA) com as amostras das bibliotecas Ribo- e poliA+. É possível ver uma separação clara das amostras Ribo- e poliA+ com relação ao componente principal 1. É possível observar também que existe uma segregação entre as réplicas biológicas 1 e as réplicas biológicas 2 da biblioteca Ribo-, o que aponta para um efeito de *batch* introduzido pelo sequenciamento dessas amostras em *lanes* diferentes do Illumina.



### 4.3.2 Detecção dos lncRNAs

Utilizando os 3 montadores de transcritos, e levando em conta o genoma do camundongo como referência, nós identificamos 52.926 novos transcritos, mas apenas 10.492 são potencialmente lncRNAs (Tabela 12).

Tabela 12: Resumo dos resultados obtidos com o *pipeline* para detecção de potenciais novos lncRNAs.

	<b>Total</b>
Novos transcritos	52.926
Tamanho médio	5.696 pb (min: 147 pb; máx: 111.912 pb)
Potenciais lncRNAs	
CPAT + RNAsamba	10.394
BLAST	181
Infernal	89
Total de potenciais lncRNAs (não redundante)	10.492

### 4.3.3 Expressão diferencial dos lncRNAs

Ao todo foram 218 lncRNAs DE, os quais 57 já são conhecidos e 161 são potenciais novos lncRNAs. A Tabela 13 apresenta o número de lncRNAs DE em cada um dos dias (a lista completa se encontra no Apêndice V: Lista de lncRNAs diferencialmente expressos (DE)). A Figura 26 apresenta o perfil de expressão gênica dos 50 lncRNAs de maior LFC absoluto nas diferentes amostras.

Tabela 13: Total de lncRNAs DE em cada um dos dias em que o experimento foi conduzido.

	R12 x P12	R12.5 vs P12	R15 vs P15	R17 vs P17
Regulado Positivamente	2	26	48	110
Regulado Negativamente	10	3	28	34

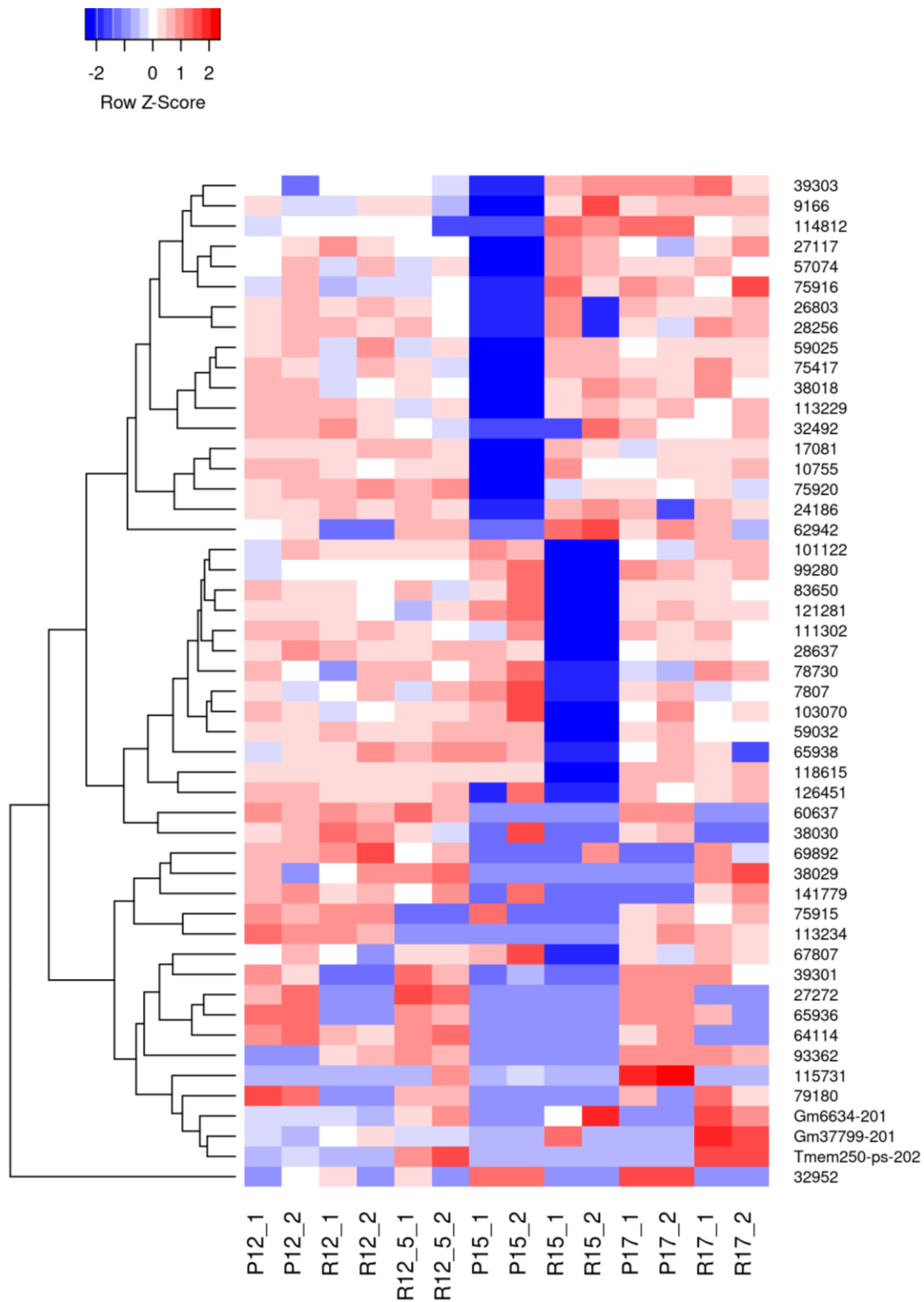


Figura 26: Heatmap do perfil de expressão de 25 lncRNAs DE com maior LFC e 25 lncRNAs DE com menor LFC, considerando todas as comparações realizadas (R12 vs P12, R12.5 vs P12, R15 vs P15 e R17 vs P17). A letra P representa as amostras do grupo controle, e a letra R representa as amostras do grupo retinopático. A grande maioria dos lncRNAs com maior LFC absoluto são novos candidatos preditos em nosso *pipeline*. Apenas 3 lncRNAs previamente descritos se encontram entre eles (Gm6634-201, Gm37799-201 e Tmem250-os-202)

#### 4.3.4 Discussão

A grande maioria dos lncRNAs com maior LFC são novos candidatos que foram preditos *in silico* com os nossos dados de sequenciamento (Figura 26). O lncRNA “62942” apresentou o maior LFC entre os lncRNAs DE. Ele foi identificado como DE no dia 12, onde sua expressão era aproximadamente 10 vezes menor no grupo retinopático, e no dia 15, sendo encontrado 10 vezes mais expresso no grupo retinopático (Figura 26). Sua posição cromossômica coincide com a posição do gene *MTPAP*, um dos fatores responsáveis pela poliadenilação da ponta 3' dos mRNAs e também é envolvido na degradação de histonas (“NCBI”, 2022), podendo ser uma variante não-codificante do gene. Já o lncRNA “32952” foi o que apresentou o menor LFC entre os lncRNAs DE, tendo sua expressão reduzida em até 17 vezes no grupo retinopático no dia 17 (Figura 26). A sua posição cromossômica coincide com lincRNA *MEG3*, podendo ser uma nova variante do gene. Foi demonstrado em células HUVEC sob hipóxia, que a expressão do *MEG3* é regulada por *HIF-1 $\alpha$* , e ele é responsável por manter a expressão do *VEGFR2* (RUAN et al., 2018). A inibição da expressão desse *lincRNA* foi associada com a supressão da migração de células endoteliais e angiogênese induzida por *VEGF* em HUVEC e em células de adenocarcinoma pulmonar (LI et al., 2019; RUAN et al., 2018; SHEN et al., 2022). Existem 3 lncRNAs que apresentaram um alto LFC e que já haviam sido descritos anteriormente, *Gm6634*, *Gm37799* e o *Tmem250-ps* (Figura 26). Todos os 3 são lincRNAs (lncRNAs intergênicos), e estão localizados nos cromossomos 3, 1 e 2, respectivamente. Até o presente momento, não identificamos estudos mais detalhados com relação à função desses lincRNAs na angiogênese ou no câncer.

Os lncRNAs discutidos nessa seção são bons candidatos para validações por RT-PCR. Uma vez validados, eles são alvos interessantes para estudos mais específicos a fim de se entender melhor sua função em angiogênese, e podem acabar se tornando alvos para desenvolvimento de novos fármacos para tratamento de doenças dependentes de angiogênese.

## 5 Rede de regulação gênica da angiogênese patológica

Uma compreensão mais detalhada dos processos moleculares da angiogênese é um passo importante para o desenvolvimento de novas terapias e métodos diagnósticos para as doenças dependentes da angiogênese. Nesse sentido, as redes de regulação gênica (GRNs) são modelos úteis e importantes, utilizados para descrever e prever dependências entre entidades moleculares complexas (genes, RNA, proteínas, metabólitos...). Geralmente são representados por grafos, onde os nós representam moléculas, e as arestas que conectam os nós representam as interações entre elas (DELGADO; GÓMEZ-VELA, 2019). As GRNs possibilitam visualizar processos biológicos complexos, auxiliando na identificação de relações causais entre seus componentes, algo que não é possível de ser feito analisando, por exemplo, individualmente os genes, como numa análise de expressão diferencial. Existem diferentes métodos de inferência de GRNs, sendo uma delas as redes de coexpressão (DELGADO; GÓMEZ-VELA, 2019). As interações nessas redes são determinadas por medidas de dependências entre os nós da rede, que podem ser coeficientes de correlação (Pearson, Spearman ou Kendall), medidas de distância (euclidiana por exemplo) e informação mútua. Dentre as vantagens desse modelo se encontram a sua simplicidade, baixo custo computacional e um baixo número de amostras requisitadas.

Um estudo recente, realizado no Hospital infantil da Universidade de Soochow na China (ZHOU et al., 2019), propôs redes de correlação circRNA-miRNA-mRNA para o modelo OIR. Foram feitos sequenciamentos de alto rendimento (RNA-Seq) com bibliotecas de RNA total e miRNAs. O estudo trouxe como contribuição a identificação de 9 circRNAs candidatos que podem atuar na regulação de 3 miRNAs, que por sua vez podem estar relacionados com o controle da expressão de genes relacionados com angiogênese. No entanto, o trabalho se restringiu a analisar o perfil de expressão gênica dos circRNAs, miRNAs e mRNAs apenas no dia 17, e os autores utilizaram como *threshold* para classificar como DE o *p-value* menor que 0.05, mas não está explícito no texto se foi utilizado algum método de correção para múltiplos testes, o que pode elevar o número de falsos-positivos nos resultados. Os dados também não se encontram disponíveis publicamente, e não foi divulgado o número de *reads* sequenciados, não sendo possível saber a profundidade do sequenciamento.

Nossos dados de sequenciamento são de alta cobertura (>100 milhões *reads*/amostra) e nossas duplicatas biológicas apresentam alta correlação entre si. Num estudo de validação com mais de 50 genes individuais, os resultados de

expressão obtidos por qRT-PCR mostram altíssima correlação com os dados de RNA-seq ( $R^2= 0.93$ ;  $p_R= 2.31 \times 10^{-104}$ ) (GUARISCHI-SOUSA et al., 2019). Esta alta correlação é esperada, pois nosso modelo se baseia em animais isogênicos (com mesmo *background* genético), bastante diferentes de amostras humanas ou mesmo de linhagens celulares, que apresentam *backgrounds* genéticos bastante diversos. O modelo do OIR é um modelo sincronizado, onde a angiogênese é induzida ao mesmo tempo em todos os animais, outro aspecto que diferencia este modelo de amostras humanas, por exemplo, de tumores. Por todas estas razões, tudo indica que as redes de coexpressão gênica geradas no presente trabalho são robustas e oferecem uma boa representação do processo de regulação gênica da angiogênese patológica na retina do camundongo.

## 5.1 Métodos

### 5.1.1 Identificação dos módulos de coexpressão de genes

Para a construção da rede de coexpressão de genes, utilizamos o pacote WGCNA (LANGFELDER; HORVATH, 2008) para o R. O WGCNA começa primeiro calculando a correlação entre todos os pares de genes, montando a matriz de similaridade. Em seguida, os valores de correlação são elevados a uma potência ( $\beta$ ), chamada de *soft threshold*, transformando a matriz de similaridade em uma matriz de adjacência. Uma vez calculado a matriz de adjacência, então é calculada a matriz de sobreposição topológica (TOM), que é determinada pelo número de nós vizinhos que são compartilhados por 2 nós conectados, de modo que, quanto maior esse número, maior será o valor de TOM (YIP; HORVATH, 2007). E por fim é calculada a matriz de dissimilaridade ( $1 - TOM$ ). Com a matriz de dissimilaridade calculada, um método de clusterização hierárquica é aplicado, e os ramos do dendrograma resultante são “cortados” pelo algoritmo implementado no pacote *dynamicTreeCut* (LANGFELDER; ZHANG; HORVATH, 2008), definindo-se assim os clusters (módulos) de genes co-expressos (FARHADIAN et al., 2021; LANGFELDER; HORVATH, 2008; NIU et al., 2019).

Para esta análise consideramos apenas a biblioteca poli-A+, pois como observado na seção 4.1.1, não é possível utilizar os dados de sequenciamento da biblioteca Ribo-, pois o sequenciamento introduziu uma diferença artificial entre as réplicas. A quantificação foi realizada com o software RSEM, considerando apenas como referência as variantes principais de cada gene, classificada segundo o banco de

dados APPRIS. Na ausência de uma classificação, foi considerada a variante de maior tamanho como a principal. Além das variantes principais de cada gene, também foram adicionados a referência os 10.492 novos lncRNAs apresentados na seção 4.3.2. Os dados de expressão foram normalizados e transformados para estabilização da variância com o método aplicado pelo pacote DESeq2 (LOVE; HUBER; ANDERS, 2014). Descartamos os transcritos cujas variâncias estimadas fossem menores que a mediana das variâncias estimadas de todos os transcritos, resultando em um total de 35.331 transcritos.

Os módulos de co-expressão foram determinados utilizando a função *blockwiseModules* (*networkType* = "signed", *corType* = "bicor", *corOptions* = *list* (*maxPOutliers* = 0.05), *soft threshold power* = 20) do WGCNA. O *soft threshold power* ( $\beta$ ) foi determinado usando a função *pickSoftThreshold* do WGCNA, onde foi escolhido o menor valor capaz de gerar uma rede com uma topologia livre de escala (BARABÁSI; OLTVAI, 2004) (Figura 27).

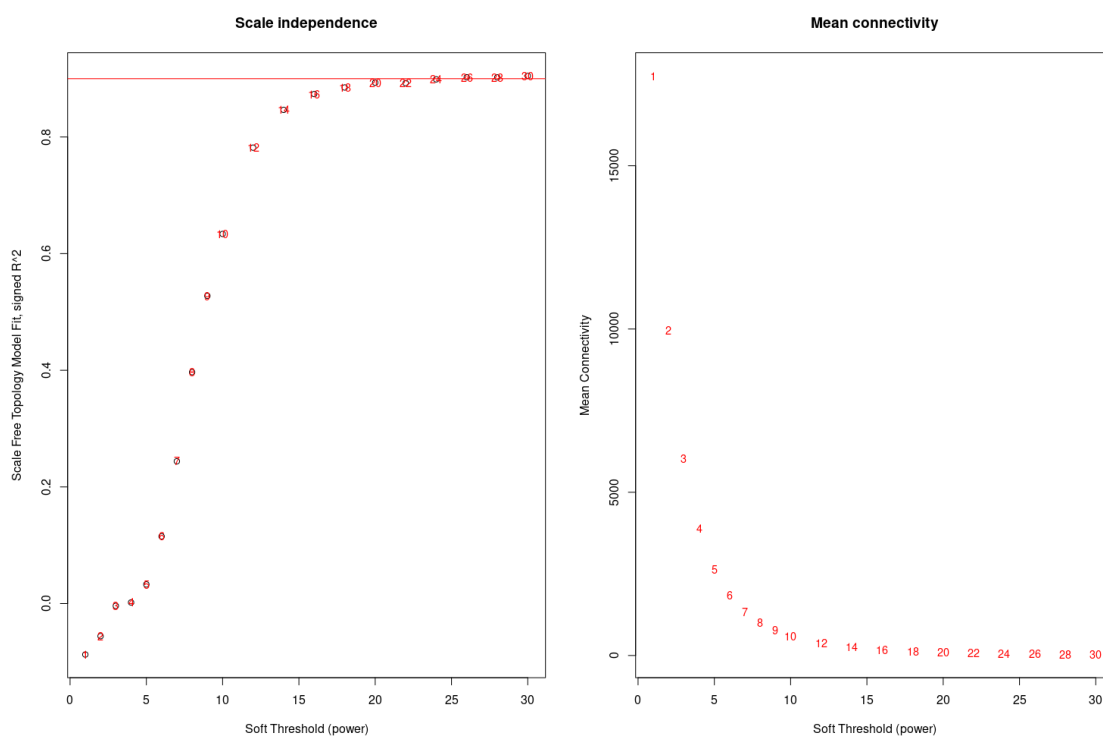


Figura 27: Seleção do valor do *soft threshold* ( $\beta$ ), que foi utilizado para gerar a matriz de adjacência. A linha horizontal vermelha marca o coeficiente de determinação ( $R^2$ ) = 0.9 de um modelo de topologia livre de escala. O melhor valor de  $\beta$  é o menor valor mais próximo da linha horizontal vermelha.

Uma vez identificados os módulos, calculamos a correlação da expressão do *eigengene*, que é o primeiro componente principal de todos os genes que compõem o módulo, com cada grupo e tempo, para verificar quais módulos e seus respectivos transcritos estão mais associados com o grupo retinopático. Os módulos com

correlação positiva ( $\rho > 0$ ) e significativa ( $p\text{-valor} \leq 0.05$ ) com amostras do grupo retinopático (R12, R12.5, R15 e R17) foram selecionados para análise de enriquecimento funcional com Gene Ontology (GO) e KEGG com a ferramenta online DAVID.

### 5.1.2 Análise de preservação dos módulos de interesse

Para avaliar a robustez dos módulos selecionados, calculamos o índice de preservação *Zsummary* com teste de permutação ( $n = 100$ ). Utilizamos como referência para a análise o conjunto de dados utilizados para construção da rede (transcritos com variância maior que a mediana). Para o conjunto teste, construímos dois conjuntos de dados: 1) transcritos com variância maior que o primeiro quartil (49.618 transcritos, ou seja, um conjunto maior que a referência); 2) transcritos com variância maior que o terceiro quartil (17.666 transcritos, ou seja, um conjunto menor que a referência).

### 5.1.3 Seleção dos genes mais relevantes para angiogênese

O próximo passo então é selecionar os transcritos de maior relevância para o módulo, e assim poupar tempo de processamento nos passos seguintes. Para isso calculamos 2 métricas dentro do WGCNA que permitem fazer essa seleção: 1) *gene significance* (GS), que é a correlação entre a expressão do transcrito e o grupo de tratamento, 2) *module membership* (MM), que é a correlação do transcrito com o *eigengene* do módulo. Apenas os transcritos com GS e MM significativos ( $p\text{-valor} \leq 0,05$ ) foram selecionados.

Para estabelecer as interações entre os transcritos selecionados de cada módulo, submetemos a lista de genes para o banco STRING para gerar uma rede de interações proteína-proteína (PPI) para cada módulo. Os parâmetros utilizados foram: '*Network Type*' = *physical subnetwork, medium confidence (0.400) score* e *FDR stringency = medium (5%)*.

Uma vez obtida a rede de cada módulo, precisamos verificar se os genes que participam dessa rede possuem alguma relevância biológica. Para esta tarefa utilizamos os dados do METABRIC (*Molecular Taxonomy of Breast Cancer International Consortium*) (CURTIS et al., 2012; PEREIRA et al., 2016) baixados do banco de dados do cBioPortal (CERAMI et al., 2012; GAO et al., 2013), que contém dados clínicos e de expressão gênica de 1.904 pacientes diagnosticadas com câncer



de mama e que não receberam tratamento com drogas anti-HER2. O recente estudo publicado pelo nosso grupo (GUARISCHI-SOUSA et al., 2019) mostrou que a assinatura gênica para angiogênese obteve uma boa resposta preditiva para as pacientes com câncer de mama, sendo portanto uma ótima doença modelo para se investigar a angiogênese patológica. Os pacientes foram divididos em 2 conjuntos de dados (conjunto de treinamento com 961 pacientes, e conjunto de validação com 943 pacientes), e o tempo de sobrevivência dos pacientes foi truncado para 15 anos, os mesmos passos seguidos por (GUARISCHI-SOUSA et al., 2019). A validação foi realizada por meio da análise de regressão de Cox univariada, onde se verificou a associação entre o perfil de expressão de cada gene com o tempo de sobrevida da paciente. Para realizar o teste, nós identificamos os homólogos humanos para os genes de camundongo selecionados e verificamos quais desses genes estavam presentes nos dados do METABRIC. Além dos valores contínuos de expressão dos genes, adicionamos também valores discretos para cada gene, sendo 1 para expresso e 0 para não-expresso. Um gene era considerado expresso se o valor contínuo de expressão estivesse acima da mediana da expressão do mesmo gene em todas as amostras, e não-expresso se esse valor estivesse abaixo da mediana (GUARISCHI-SOUSA et al., 2019). Os genes que tiveram uma associação significativa ( $p$ -valor  $\leq 0,05$ ) foram considerados validados e por isso foram selecionados para as demais análises.

#### 5.1.4 Construção da rede de regulação gênica para angiogênese

Para os genes que foram validados na etapa anterior, nós recuperamos as redes PPI geradas pelo STRING (SZKLARCZYK et al., 2021), e verificamos quais miRNAs regulam esses genes, quais lncRNAs estão co-expressos com esses genes, e se existem circRNAs preditos também para esses genes, e quais miRNAs são seus possíveis alvos. As redes obtidas foram visualizadas e analisadas pelo software Cytoscape (SHANNON et al., 2003).

## 5.2 Resultados

### 5.2.1 Identificação dos módulos de coexpressão de genes

Para construir a rede foram utilizados os dados de expressão de 25.218 mRNAs e 10.133 lncRNA (3436 conhecidos e 6.667 novos) como input para o programa

WGCNA. Ao todo foram identificados 46 módulos, onde cada módulo é identificado por uma cor (Figura 28). Desses 46, 12 módulos estão positivamente correlacionados com amostras do grupo retinopático, sendo eles: *blue*, *brown*, *darkgrey*, *darkred*, *floralwhite*, *grey60*, *lightsteelblue1*, *mediumpurple3*, *sienna3*, *violet* e *yellowgreen* (Figura 29). A análise de enriquecimento funcional com Gene Ontology (GO) mostrou que termos GO relacionados com angiogênese estão entre as funções mais enriquecidas nos módulos “darkred” e “brown” e por isso esses módulos foram selecionados para as próximas etapas. Os termos GO mais enriquecidos para cada um dos módulos, respectivamente, foram “*cellular response to hypoxia*” (Tabela 14) e “*angiogenesis*” (Tabela 15), e como é possível observar na Figura 29, o módulo “darkred” está principalmente associado com o grupo R12.5, quando ocorre o pico de hipóxia no tecido, e o módulo “brown” está principalmente associado com o grupo R17, que é quando ocorre o pico de neovascularização do tecido.

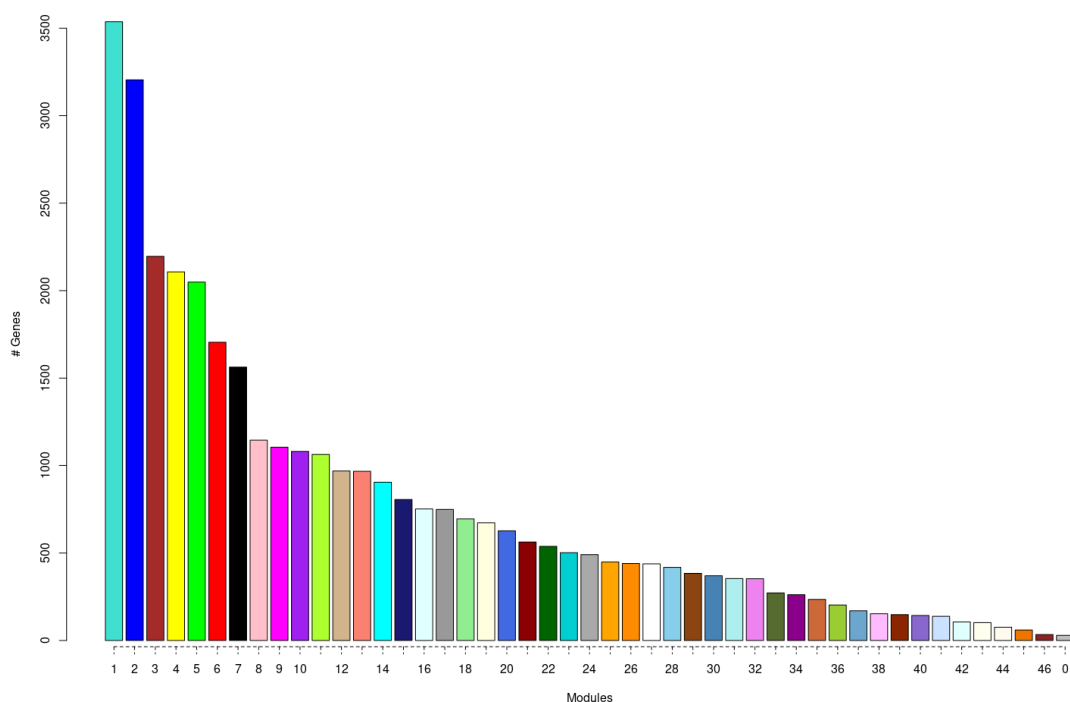


Figura 28: Tamanho, em número de transcritos, de cada módulo identificado. No módulo 0 (cor cinza) se encontram os transcritos que não se agruparam em nenhum dos módulos.

Tabela 14: Os 5 termos GO (ontologia "Biological process") mais enriquecidos entre os genes do módulo "Darkred". A análise de enriquecimento foi realizada com a ferramenta web DAVID, e o método de correção do p-valor utilizado foi a de Benjamini-Hochberg.

<b>Termo GO</b>	<b>Contagem de genes</b>	<b>P-valor</b>	<b>Correção do p-valor.</b>
Celullar response to hypoxia	19	5,7E <sup>-11</sup>	1,5E <sup>-7</sup>
Glycolytic process	12	1,2E <sup>-10</sup>	1,6E <sup>-7</sup>
Response to hypoxia	20	5,0E <sup>-8</sup>	4,5E <sup>-5</sup>
Antigen processing and presentation of endogenous peptide antigen via MHC class I via ER pathway, TAP-independent	9	2,5E <sup>-7</sup>	1,3E <sup>-4</sup>
Antigen processing and presentation of endogenous peptide antigen via MHC class Ib	9	2,5E <sup>-7</sup>	1,3E <sup>-4</sup>

Tabela 15: Os 5 termos GO (ontologia "Biological process") mais enriquecidos entre os genes do módulo "Brown". A análise de enriquecimento foi realizada com a ferramenta web DAVID, e o método de correção do p-valor utilizado foi a de Benjamini-Hochberg.

<b>Termo GO</b>	<b>Contagem de genes</b>	<b>P-valor</b>	<b>Correção do p-valor.</b>
Angiogenesis	19	5,7E <sup>-11</sup>	1,5E <sup>-7</sup>
Inflammatory response	12	1,2E <sup>-10</sup>	1,6E <sup>-7</sup>
Regulation of cell shape	20	5,0E <sup>-8</sup>	4,5E <sup>-5</sup>
Positive regulation of angiogenesis	9	2,5E <sup>-7</sup>	1,3E <sup>-4</sup>
Cell migration	9	2,5E <sup>-7</sup>	1,3E <sup>-4</sup>

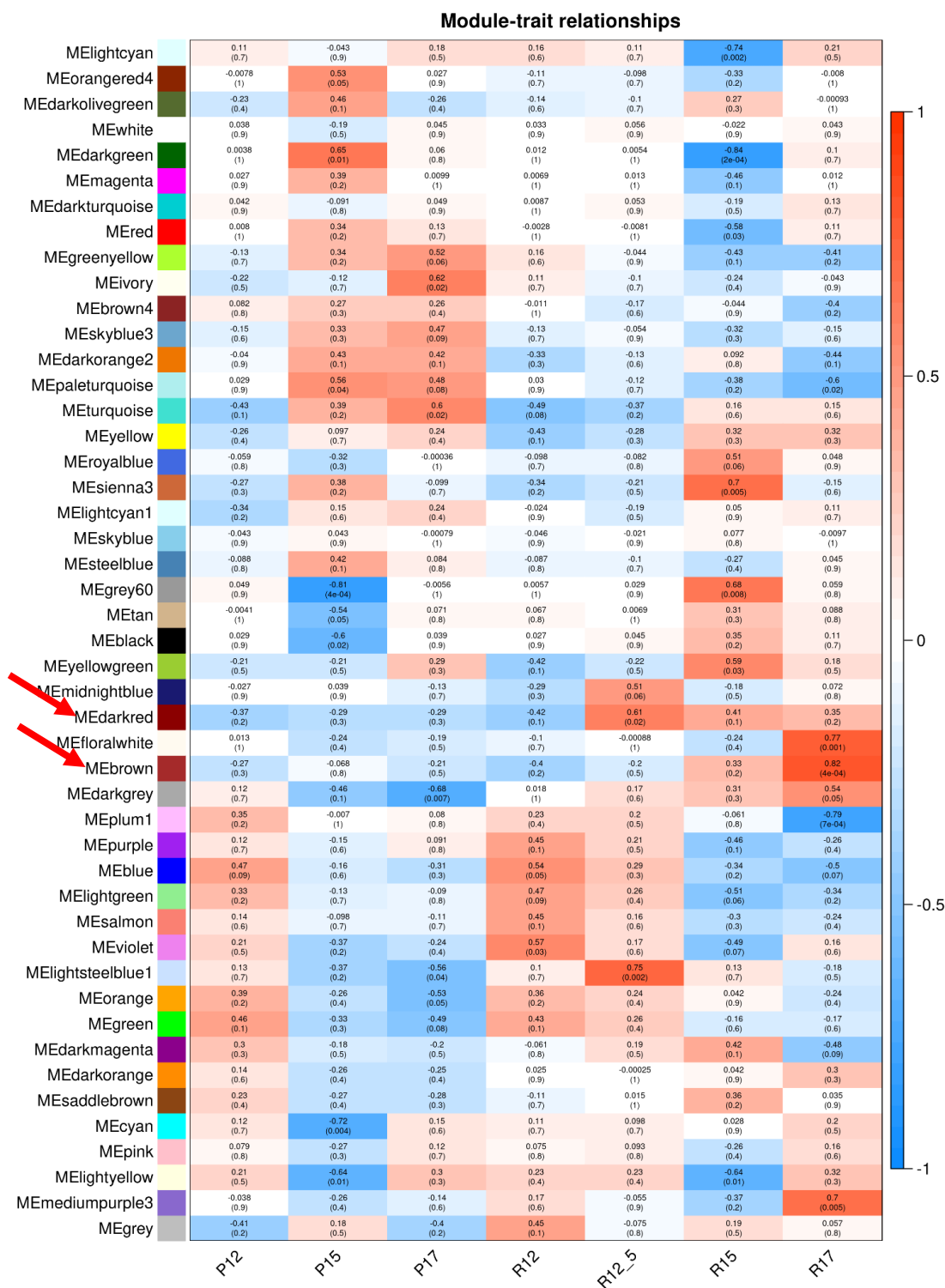


Figura 29: Heatmap dos valores de correlação (*Pearson*) dos *eigengene* de cada módulo com os grupos de tratamento. O *eigengene* corresponde a primeira componente principal do módulo e representa o perfil de expressão de todos os genes do módulo. As setas indicam os módulos “Darkred” e “Brown” que foram selecionados para as etapas seguintes da análise. É possível ver que o módulo “Darkred” tem uma correlação positiva e significativa ( $p$ -valor < 0,05) com as amostras do grupo R12\_5, enquanto o módulo “Brown” possui uma correlação positiva e significativa com as amostras do grupo R17.

## 5.2.2 Análise de preservação dos módulos de interesse

A análise de preservação dos módulos *darkred* e *brown* mostrou que em ambas as comparações realizadas os módulos se mantiveram entre os mais preservados, mesmo variando o tamanho do conjunto de dados, uma vez que ambos estão com *Zsummary* acima de 10 (Figura 30).

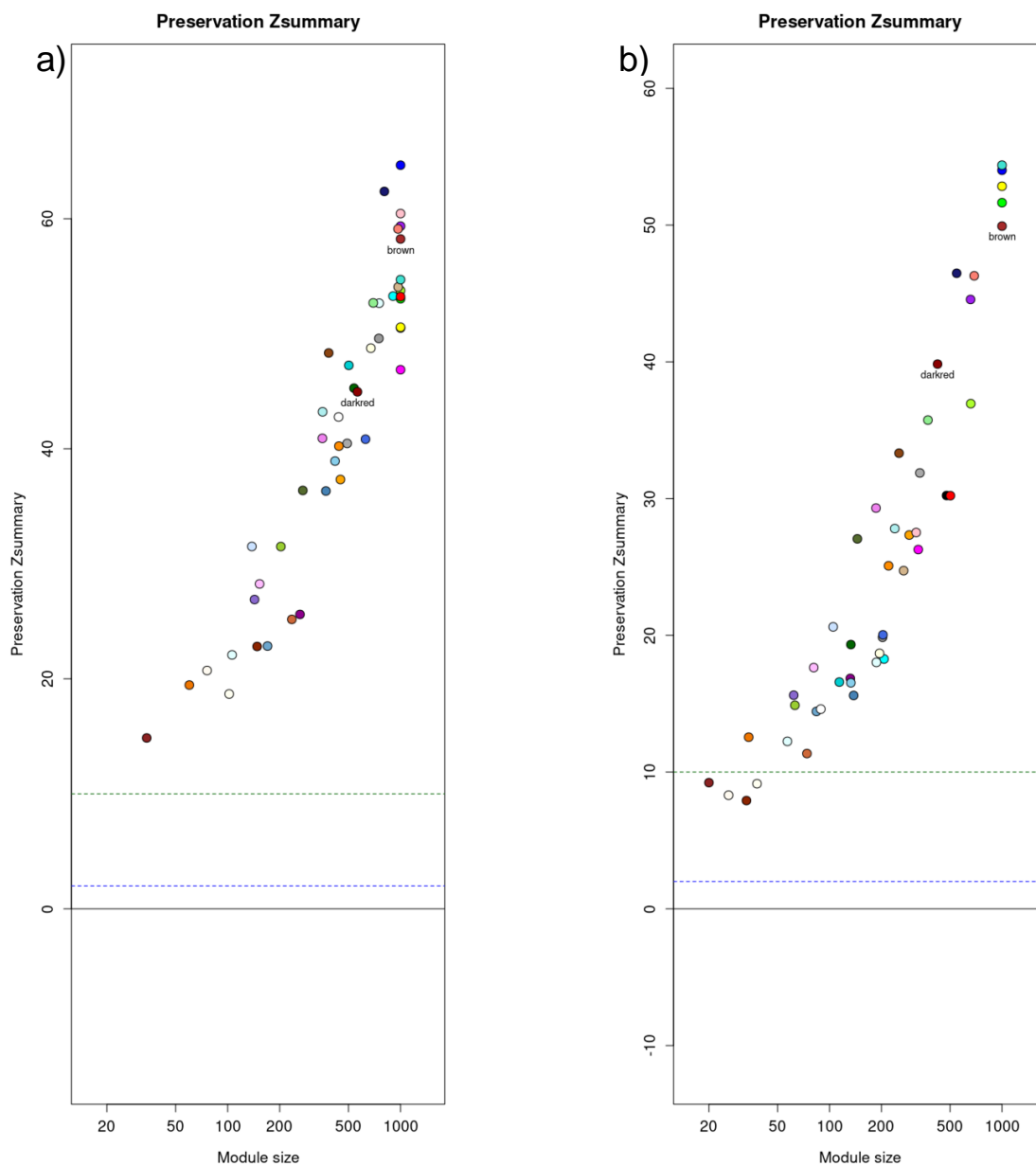


Figura 30: Índice de preservação Zsummary: a) entre o conjunto de referência (35.331 transcritos) e o conjunto 1º quartil (transcritos com variância maior que o primeiro quartil, um total de 49.618 transcritos) e b) entre o conjunto de referência e o conjunto 3º quartil (transcritos com variância maior que o terceiro quartil, um total de 17.666 transcritos).

### 5.2.3 Seleção dos genes mais relevantes para angiogênese

Após filtrar os transcritos com  $p$ -valor significativo ( $\leq 0,05$ ) em GS e MM, restaram 204 e 1407 transcritos nos módulos *darkred* e *brown*, respectivamente. Com a ferramenta STRING nós geramos duas redes PPI, uma para cada módulo. A rede PPI do módulo “darkred” possui 55 transcritos e 71 interações (Figura 31), enquanto o módulo “brown” possui 631 transcritos e 1700 interações (Figura 32).

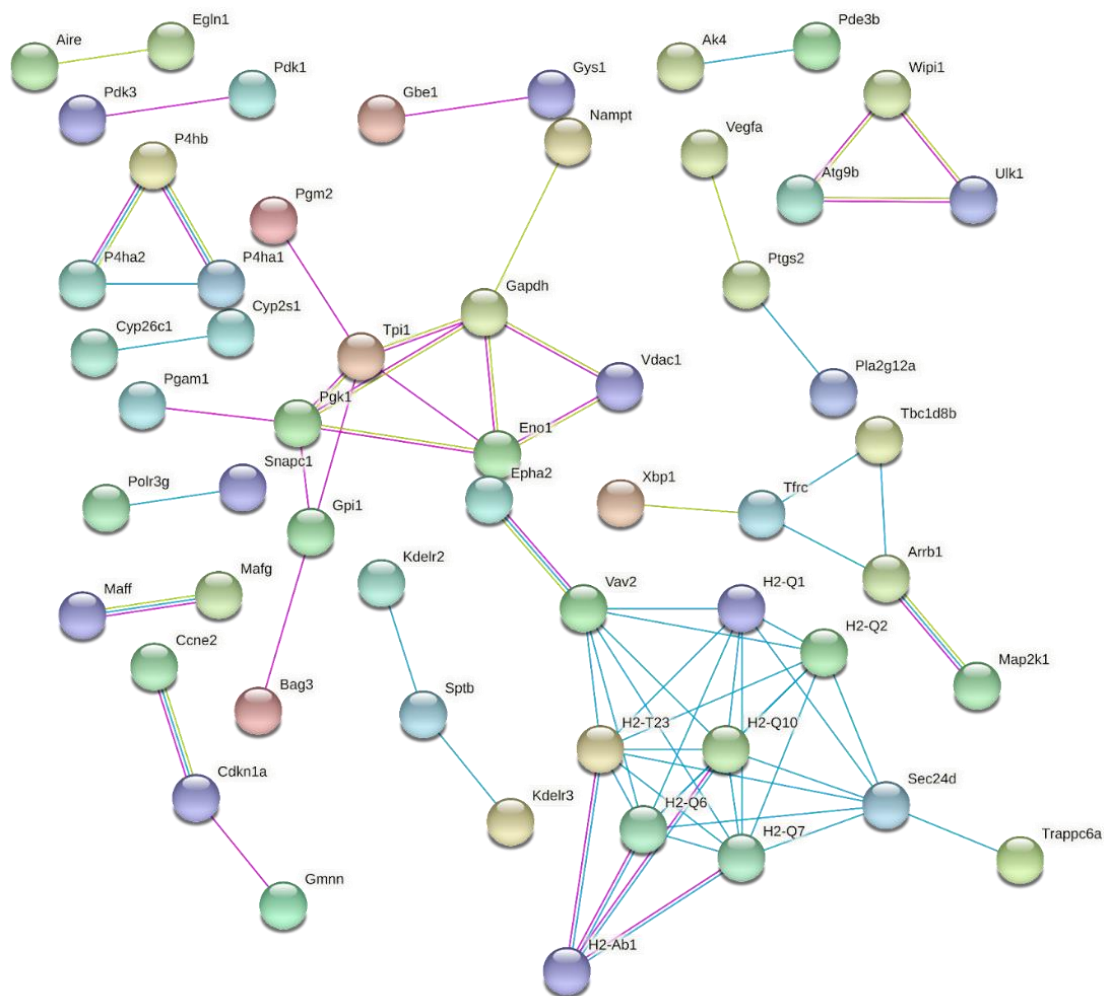


Figura 31: Rede PPI dos genes do módulo “darkred” gerada pelo software STRING.

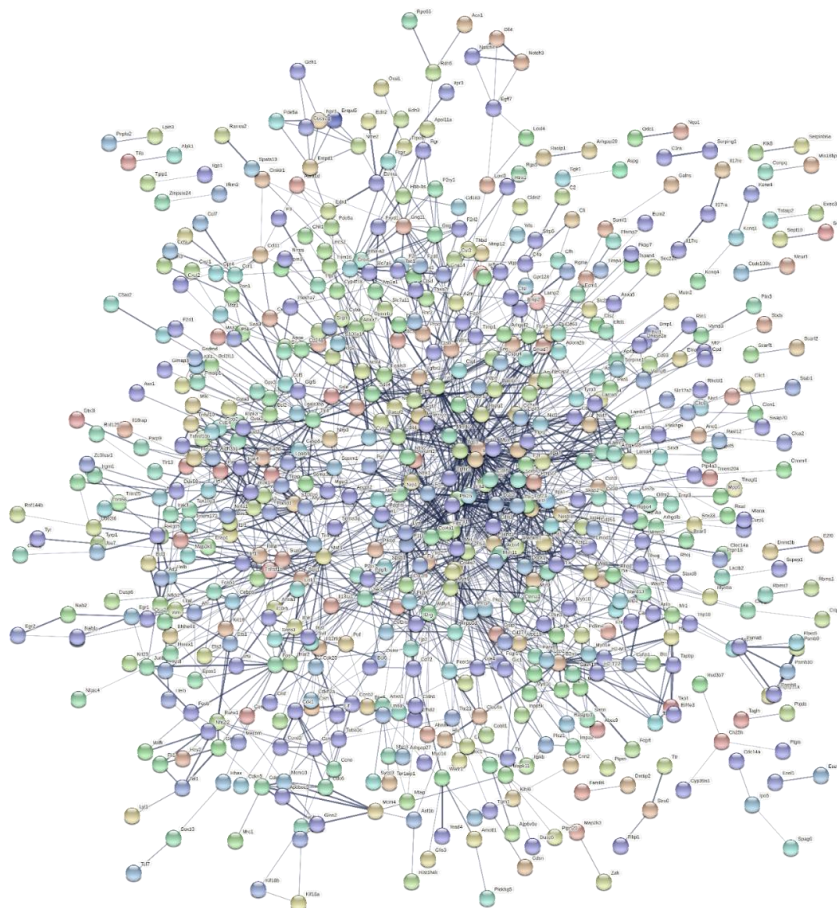


Figura 32: Rede PPI dos genes do módulo “brown” gerada pelo software STRING.

Encontramos 50 e 615 homólogos humanos para os genes das redes PPI dos módulos “darkred” e “brown”, respectivamente. Destes, 49 e 573 genes estão presentes nos dados do METABRIC. A partir da análise de regressão de Cox verificamos que 18 e 220 genes dos módulos “darkred” e “brown”, respectivamente, estão associados significativamente com o tempo de sobrevivência das pacientes com câncer de mama (Tabela 16 Tabela 17).

Tabela 16: Os 10 genes mais significativos (em ordem decrescente) do módulo “darkred”.

Posição	Gene	Posição	Gene
1º	<i>ENO1</i>	6º	<i>PDK1</i>
2º	<i>GPI</i>	7º	<i>TRFC</i>
3º	<i>GAPDH</i>	8º	<i>P4HA1</i>
4º	<i>VEGFA</i>	9º	<i>P4HB</i>
5º	<i>CCNE2</i>	10º	<i>PTGS</i>

Tabela 17: Os 10 genes mais significativos (em ordem decrescente) do módulo “brown”.

Posição	Gene	Posição	Gene
1º	CST3	6º	F2RL2
2º	CDKN2C	7º	NQO1
3º	BCL2	8º	EZR
4º	EPHX1	9º	EFEMP2
5º	KIF14	10º	DNMT3B

#### 5.2.4 Construção da rede de regulação gênica

Verificamos que dos 99 miRNAs DE na retina do nosso modelo animal, 30 são potencialmente reguladores de 7 genes do módulo “darkred”. Já no módulo “brown” identificamos 58 miRNAs potencialmente regulando 67 genes (Tabela 18).

Tabela 18: Resumo com o total de miRNAs DE (conhecidos e novos) interagindo com genes validados dos módulos “darkred” e “brown”.

Módulos	miRNAs conhecidos	novos miRNAs
Darkred	26	4
Brown	52	6

Verificamos também se para os genes mais significativos dos módulos foram preditos circRNAs e se eles possuem sítios de ligação para os miRNAs que regulam seus genes parentais. A predição dos sítios de ligação foi realizada com TargetScan (Tabela 19).

Tabela 19: Resumo com o total de circRNAs preditos para os genes validados nos módulos “darkred” e “brown”.

Módulos	circRNAs	miRNAs DE alvos
Darkred	3	0
Brown	17	0

Para determinar se um lncRNA está coexpresso com os genes mais significativos dos módulos, descartamos as interações fracas apontadas pelo WGCNA, para as quais a probabilidade de serem verdadeiras é menor. Consideramos como interação





e, portanto, mais caminhos (arestas) passam por ele. Esta imagem, assim como as Figura 34 e Figura 35 foram geradas pelo software Cytoscape.

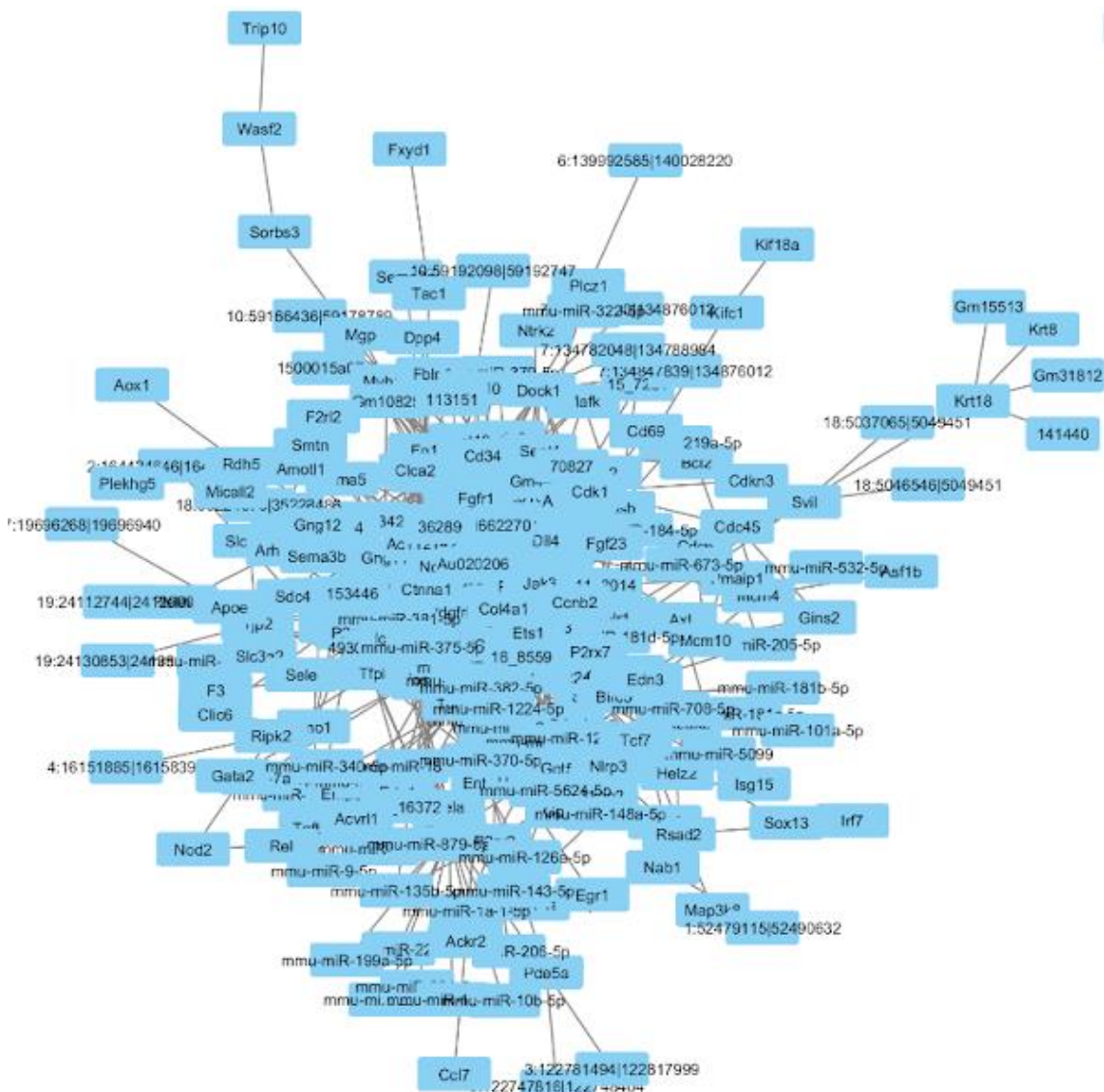


Figura 34: Rede de regulação gênica baseada no módulo “brown”.

O módulo “brown” ficou ilegível devido à sua grande quantidade de nós e arestas. Por isso, nós calculamos com o Cytoscape algumas estatísticas relacionadas com a rede. Uma delas é a *betweenness centrality*, que quantifica a centralidade de um nó na rede. Quanto maior a centralidade de um nó, maior a importância dele na rede, uma vez que mais caminhos na rede passam por esse nó. Para o módulo “brown” nós geramos uma sub-rede com os 20 nós (RNAs) mais centrais da rede (Figura 35).

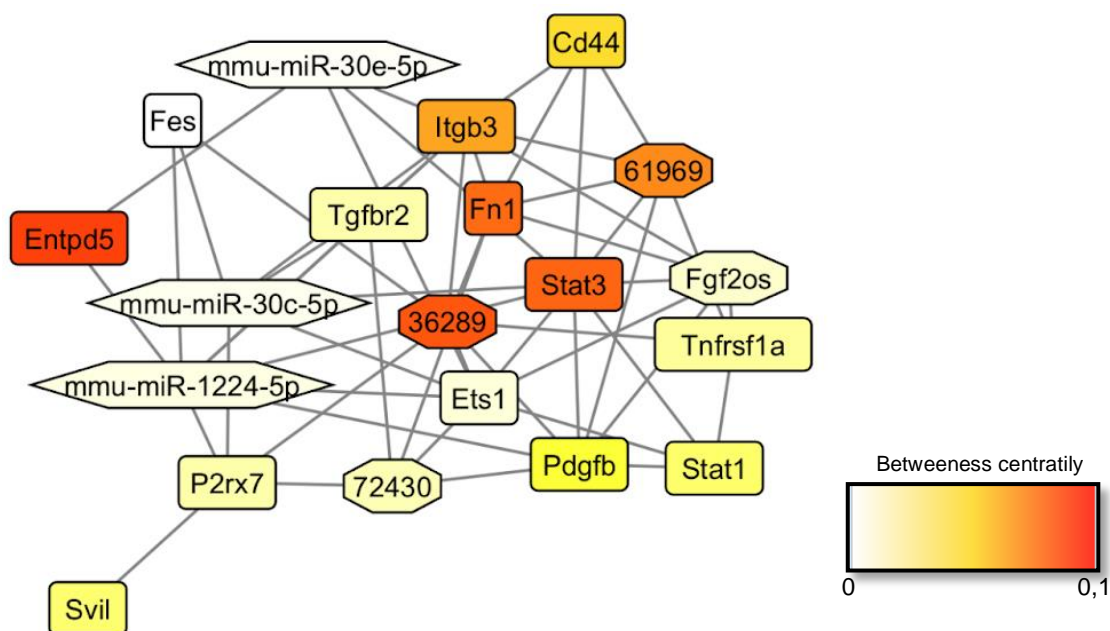


Figura 35: Os 20 RNAs de maior centralidade (*betweenness centrality*) na GRN do módulo “brown”. Os mRNAs são os retângulos, os miRNAs são os hexágonos, os lncRNAs são os octógonos e os circRNAs são as elipses. Os mRNAs estão representados pelos retângulos, os miRNAs pelos hexágonos, os lncRNAs pelos octógonos e os circRNAs pelas elipses. A cor do nó representa a centralidade do RNA na rede. Quanto mais quente, maior a centralidade do nó, e, portanto, mais caminhos (arestas) passam por ele.

### 5.3 Discussão

A partir dos resultados gerados com WGCNA, juntamente com as análises de enriquecimento funcional com o GO, nós identificamos 2 módulos de coexpressão de genes que recapitulam 2 distintas etapas da angiogênese em nosso modelo animal: o estágio inicial de resposta ao estado de hipóxia, onde fatores pró-angiogênicos são expressos para ativar a via de sinalização da angiogênese, sendo essa etapa representada pelo módulo “darkred”; e a fase da neovascularização em si, onde células endoteliais especializadas se proliferam e migram para formar novos vasos sanguíneos, fase essa representada pelo módulo “brown” (ADAIR; MONTANI, 2010). Ambos os módulos são robustos quanto a sua composição de genes, pois eles se mantiveram preservados mesmo alterando o tamanho do conjunto de dados de entrada para o WGCNA (Figura 30), o que aumenta a confiabilidade dos nossos resultados. Após selecionar os genes mais relevantes para ambos os modelos (utilizando as métricas GS e MM, recuperando as redes PPI com STRING e validando com os dados do METABRIC), e cruzar com os resultados obtidos com as análises de miRNAs, circRNAs e lncRNAs (Seção 5.2.4), nós obtivemos duas redes de regulação gênica, cada uma representando uma fase da angiogênese do nosso modelo animal.

Entre os RNAs mais centrais da rede do módulo “darkred” (Figura 33), nós temos os mRNAs dos genes *VEGFA*, *PGK1*, *P4HB* e *PTGS2*, sendo todos esses anotados com o termo GO “*cellular response to hypoxia*”, e os mRNAs dos genes *Gapdh*, *PDK1*, *ENO1*, *PGK1* e *GPI1*, que são genes essenciais da via glicolítica. Já é conhecida a relação da expressão de *HIF-1A* com a expressão de genes da via glicolítica (DU et al., 2021; KIERANS; TAYLOR, 2021; ROBEY et al., 2005), assim como a importância da via glicolítica para a progressão da angiogênese, sendo essa via a principal fonte de energia das células endoteliais em angiogênese (DRAOUI; DE ZEEUW; CARMELIET, 2017; DU et al., 2021; STAPOR et al., 2014; YETKIN-ARIK et al., 2019). É provável que o gene *PGK1* seja um dos responsáveis por fazer a ponte entre essas 2 vias, pois o gene *PGK1* não só é induzido por hipóxia (LI; KO; WHITLOCK, 1996; ZHANG et al., 2020c), mas como também contribui para o crescimento e progressão de tumores (HE et al., 2022; ZHANG et al., 2020c), sendo ele um potencial candidato para alvo de terapias anti-angiogênicas.

Na rede “darkred” nós observamos 5 lncRNAs, 4 preditos *in silico* em nossas análises e um previamente anotado, o Gm15283. Esse último é um lncRNA do tipo antisense, e se encontra na fita oposta ao gene *HIF-1A* (Figura 36).

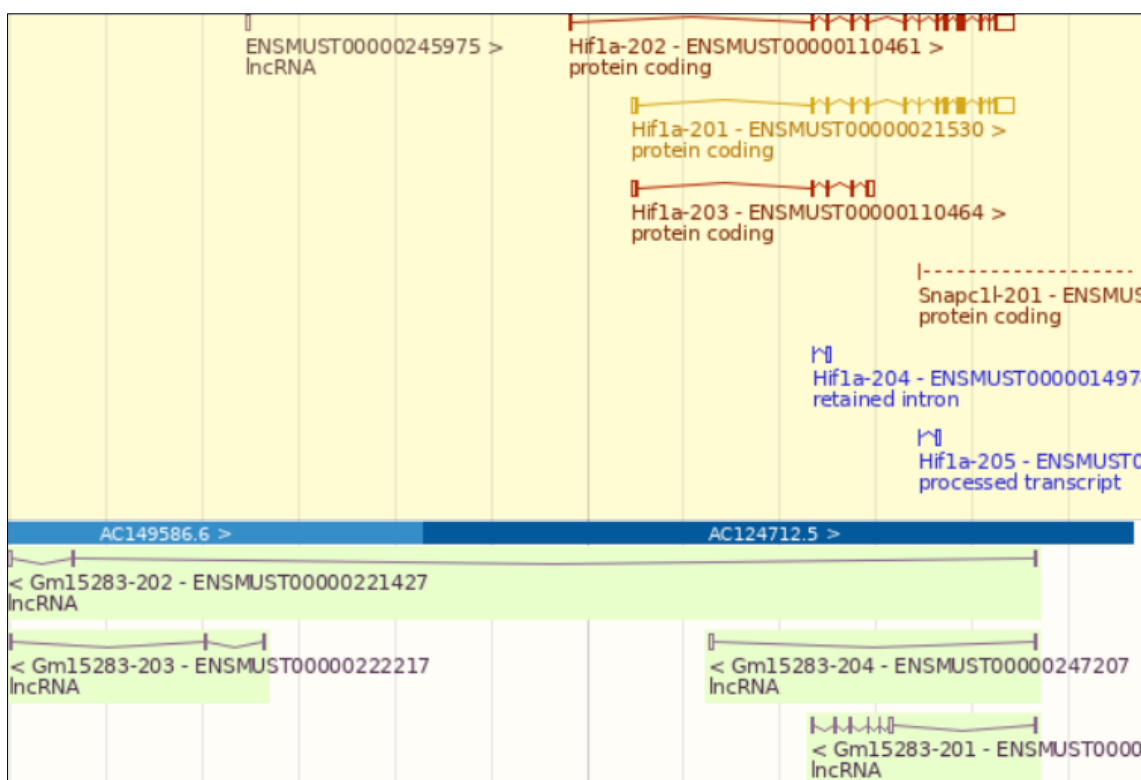


Figura 36: O lncRNA Gm15283 se encontra na fita oposta ao gene HIF-1 $\alpha$ . Imagem retirada do ENSEMBL (ENMUSG00000087700)

Em humano, existem 3 *HIF-1A* lncRNAs antisenso, *HIF1A-AS1*, *-AS2* e *-AS3*. *HIF1A-AS3*, chamado pelos autores como *HIFAL*, auxilia o *HIF-1A* a se ligar aos elementos de resposta a hipóxia (HRE) dos seus genes alvos, tornando o lncRNA essencial na manutenção da expressão desses genes, incluindo os que pertencem a via glicolítica (ZHENG et al., 2021). O mesmo estudo apontou que alto níveis de expressão desse lncRNA estão associados com um prognóstico ruim para pacientes com câncer de mama. *HIF1-AS1* apresenta funções similares ao *HIF1-AS3*. Um estudo relacionou a resistência a droga Gemcitabina (GEM) no tratamento de pacientes com câncer de pâncreas com a expressão do *HIF1-AS1*. Eles observaram que o lncRNA induz a expressão do *HIF-1A*, o que promove a glicólise, aumentando a resistência das células de câncer pancreático contra GEM (XU et al., 2021). *HIF1-AS2* já foi associado com aumento da progressão associada a hipóxia de diversos tumores e doenças (MINEO et al., 2016; WU et al., 2019a; ZHANG et al., 2022, 2020b; ZHU et al., 2020). *HIF1-AS2* foi capaz de induzir angiogênese em células endoteliais da veia umbilical humana (HUVEC) sob hipóxia. Esse lncRNA sequestra o miR-153-3p, impedindo o silenciamento do gene *HIF-1A* (LI et al., 2017). Nossa rede mostra a correlação entre a expressão do lncRNA Gm15283 com os genes de resposta a hipóxia e da via glicolítica, corroborando com o que foi observado pelos estudos apresentados acima. Isso torna esse lncRNA um bom candidato para futuros estudos de desenvolvimento de novas terapias anti-angiogênicas.

Quantos aos novos lncRNAs preditos em nossos dados, o lncRNA “19170” se encontra na mesma região do gene *P4HA2*, podendo ser uma variante não codificante do gene. O lncRNA “42839” se encontra na fita oposta ao gene *ZFH2* (*Zinc finger homeobox 2*), que é um fator de transcrição expresso principalmente no cérebro, e está associado a funções como sensibilidade a dor (AKSHAYAA et al., 2019; HABIB et al., 2018), diferenciação de neurônios e distúrbios comportamentais (KOMINE et al., 2012). O lncRNA “28938” se encontra na mesma região do pseudogene *CT030170.4*, e o lncRNA “41628” localiza-se em uma região intergênica no cromossomo 14, entre os pseudogenes Gm16439 (ENSMUSG00000072594) e Gm8256 (ENSMUSG00000091605). A relação desses novos lncRNAs com angiogênese ainda não é clara, sendo necessários estudos mais específicos para elucidar seus papéis nesse processo.

Para a sub-rede com os 20 RNAs mais centrais da rede “brown” (Figura 35) observa-se que os mRNAs dessa rede pertencem a genes que estão envolvidos com regulação da angiogênese, seja na regulação da proliferação, migração e/ou na diferenciação das células endoteliais, o que corrobora com o experimento executado,

onde o dia 17 é quando a neovascularização atinge o seu pico. Todos esses genes são potenciais alvos para o desenvolvimento de terapias anti-angiogênicas.

Ainda sobre a sub-rede do módulo “brown”, nós observamos 4 lncRNAs, 3 preditos *in silico* em nossas análises e um previamente anotado, o *FGF2OS*, também conhecido com *FGF2-AS*. Esse último é um lncRNA do tipo antisense, e se encontra na fita oposta ao gene *FGF2*. O *FGF2OS* é complementar a região 3' UTR do gene *FGF2*, e ao se ligar a essa região, ele impede a transcrição do gene *FGF2*, regulando assim sua expressão. A expressão elevada de *FGF2* em pacientes com câncer esofágico representa um prognóstico ruim, sendo associado com recorrência do tumor e tempo de sobrevivência reduzido, mas quando coexpresso com *FGF2OS* ocorre uma melhora desse prognóstico (BARCLAY et al., 2005; MACFARLANE et al., 2010).

Quanto aos novos lncRNAs preditos em nossos dados, o lncRNA “36289” se encontra em região intergênica entre os genes *NSD1* e *FGFR4*. O lncRNA “72430” é possivelmente uma variante não codificante do gene *EGFL7*, um fator angiogênico que atua na proliferação, migração e invasão de células endoteliais, expresso durante embriogênese e angiogênese fisiológica e patológica (NICHOL; STUHLMANN, 2012). O lncRNA “61969” se encontra na fita oposta próximo ao gene *RAB12*, que está associado com radioresistência de células de câncer cervical de pacientes HPV-positivo (HUANG et al., 2021). A relação desses novos lncRNAs com angiogênese ainda não é clara, sendo necessário estudos mais específicos para elucidar seus papéis nesse processo.

Três miRNAs apareceram entre os RNAs mais centrais da rede “Brown”, o miR-30e-5p, miR-30c-5p e o miR-1224-5p. A baixa expressão de miR-30e-5p em pacientes com carcinoma em células escamosas da cabeça e do pescoço (SCCHN) foram associados com prognóstico ruim (ZHANG et al., 2020a). O mesmo estudo observou em experimentos *in vitro* e *in vivo* que o miRNA inibiu a migração, invasão e metástase de SCCHN bloqueando a angiogênese por meio da inibição de fatores pró-angiogênicos como *VEGF*, *HGF*, *FGF1/2*, *CSF1/2/3*, *IL-8*, *IGF1/2* (ZHANG et al., 2020a). Da mesma família do miR-30e, o miR-30c-5p também é associado ao papel de supressor tumoral. Seu nível de expressão é um fator prognóstico para câncer de estômago, onde a baixa expressão desse miRNA é associada a um prognóstico ruim para a doença (HU et al., 2021). Em células endoteliais do fígado, foi observado que o miR-30c-5p inibe a angiogênese ao se ligar ao gene *DLL4*, reduzindo os efeitos da fibrose hepática (GU et al., 2021). O miR-1224 por outro lado pode atuar como um fator pró-angiogênico. Em ensaios *in vitro* com HUVEC foi observado que o aumento de expressão desse miRNA também aumentou a sinalização por *VEGF* e inibiu a via de sinalização do *NOTCH*, estimulando a formação de tubo (SAKAI et al., 2017). O

mesmo estudo observou que a expressão do miR-1224 é controlada pela expressão do *EPSIN2*, um fator anti-angiogênico. Quando a expressão do *EPSIN2* é inibida, a expressão do miR-1224 aumenta, e vice-versa (SAKAI et al., 2017). A literatura mostra, portanto, que os 3 miRNAs são reguladores da angiogênese, e por isso são bons candidatos para futuras validações e para estudos que visam desenvolver novas terapias anti-angiogênicas para tratamento de doenças dependentes de angiogênese.

## 6 Assinatura gênica de angiogênese para prognóstico de câncer de mama

O câncer está entre as principais causas de mortes no mundo; no Brasil, por exemplo, a doença ocupa a segunda posição (SUNG et al., 2021). Em 2020, o câncer de mama foi o tipo de câncer de maior incidência (24,5%) e a de maior mortalidade (15,5%) em mulheres (SUNG et al., 2021). Quadro semelhante ocorre em muitos países do mundo e, portanto, é uma doença de preocupação global.

O câncer de mama pode ser subdividido em 3 classes moleculares, com características clínicas importantes. Os do tipo luminal (subdivido em A e B) são aqueles que expressam receptores de estrogênio e progesterona e que não expressam ERBB2, sendo esses associados com um melhor prognóstico, por crescer mais lentamente. Os do tipo HER2 positivos, são aqueles que expressam ERBB2 mas não expressam os receptores hormonais; já os Triplo negativos (também conhecido como Basal), não expressam nem os receptores hormonais e nem a proteína HER2, sendo esse o tipo de pior prognóstico e sem medicamentos específicos disponíveis para o tratamento (HARBECK et al., 2019; LOIBL et al., 2021; WAKS; WINER, 2019). Mesmo com essa classificação molecular, o tumor de mama (como outros tipos tumorais) ainda é uma doença bastante heterogênea, e a decisão sobre o tratamento adequado para cada paciente baseada nessa classificação ainda é limitada. Por isso, a aplicação de assinaturas gênicas para classificar as pacientes em diferentes categorias de risco/sobrevivência, em conjunto com os dados clínico-patológicos, podem auxiliar na decisão do tratamento mais adequado ao paciente (ARRANZ et al., 2012; QIAN et al., 2021).

Desde 2002, quando a primeira assinatura gênica prognóstica para câncer de mama foi publicada, a AGENDIA MammaPrint (VAN DE VIJVER et al., 2002), novas assinaturas vêm sendo desenvolvidas para auxiliar no prognóstico de pacientes com câncer de mama (QIAN et al., 2021). Em 2019, nosso grupo publicou uma assinatura gênica de angiogênese com poder prognóstico para pacientes diagnosticadas com esse tipo de câncer (GUARISCHI-SOUSA et al., 2019), que foi capaz de distinguir significativamente ( $p = 2.56 \times 10^{-21}$ ) as pacientes em 3 grupos de risco: risco baixo, quando probabilidade de sobrevivência ( $P$ ) > 50% após 15 anos do diagnóstico da doença; risco médio: quando  $P > 50\%$  entre 7,5 e 15 anos; e risco alto nos demais casos. A assinatura foi construída a partir dos resultados de expressão diferencial, onde foram identificados 153 genes DE. Como nós geramos as redes de coexpressão (seção 5.2.4), abrangendo um conjunto de genes diferentes (em parte) dos 153 genes



DE, mas que também são relevantes para a regulação da angiogênese, nós desenvolvemos 3 novas assinaturas gênicas para prognóstico de câncer de mama utilizando os genes co-expressos presentes nessas redes.

## 6.1 Métodos

A partir dos genes validados dos módulos “darkred” (18 genes) e “brown” (220 genes) (Seção 5.2.3 ) nós desenvolvemos 3 modelos de assinatura gênica de angiogênese para prognóstico de câncer de mama: 1) utilizando apenas os genes do módulo “darkred”; 2) apenas os genes do módulo “brown”; e a união dos genes dos módulos “darkred” e “brown”. Para realizar esta tarefa, utilizamos o pacote `randomForestSRC` do R, que aplica um algoritmo de aprendizado de máquina conhecido como Survival Random Forest. O algoritmo gera diversas árvores de decisão com diferentes combinações aleatórias (com repetições) de *features* e de amostras (*bootstrap*). As árvores crescem conforme seus nós são divididos seguindo a regra “*log-rank test*”. Ao final, todas as árvores geradas são combinadas, fazendo-se uma média das estatísticas dos nós terminais de cada uma das árvores, e então uma função de sobrevivência e uma função de risco acumulado são estimadas para o modelo final.

Esse algoritmo foi utilizado para construir um modelo capaz de prever o tempo de sobrevivência das pacientes. Primeiro, nós ranqueamos, em ordem crescente, a lista de genes de interesse de acordo com o p-valor calculado na análise de regressão de Cox univariada (seção 5.2.3 ). Em seguida, os genes dessa lista foram incrementados um a um (*forward selection*) na fórmula da função “`rfsrc`” do pacote `randomForestSRC`. Cada vez que um gene era incrementado à fórmula, um novo modelo era gerado. E por fim, utilizando o *Harrell’s C-index* (HARRELL et al., 1982) para determinar a taxa de erro nas predições dos modelos, nós selecionamos como modelo final aquele que obteve a menor taxa de erro. Além dos valores contínuos de expressão dos genes, também foram utilizados valores discretos para cada gene (representado pelo nome do gene seguido do sufixo “\_\_B”) na construção dos modelos, sendo 1 para expresso e 0 para não-expresso. Um gene era considerado expresso, se o valor contínuo de expressão estivesse acima da mediana da expressão do mesmo gene em todas as amostras, e não-expresso, se esse valor estivesse abaixo da mediana (GUARISCHI-SOUSA et al., 2019). Também foram incluídos nos modelos os dados de idade do paciente quando recebeu o diagnóstico e o estágio do

tumor. Cada modelo foi gerado e testado com o conjunto de treinamento do METABRIC (Tabela 21).

Uma vez obtido o modelo final de cada assinatura, nós as aplicamos no conjunto de validação do METABRIC para gerar as predições. Para categorizar os pacientes, nós seguimos as mesmas regras aplicadas em (GUARISCHI-SOUSA et al., 2019): risco baixo, quando probabilidade de sobrevivência (P) > 50% após 15 anos do diagnóstico da doença; risco médio: quando P > 50% entre 7,5 e 15 anos; e risco alto nos demais casos. Para visualizar os resultados foram plotadas curvas de Kaplan-Meier geradas com a função *ggsurvplot* do pacote “*survminer*”.

Tabela 21: Dados clínicos das pacientes do banco de dados do METABRIC nas coortes de Treinamento e Validação. NA = Não disponível; \* = Mediana (Quartil inferior, Quartil superior).

	Treinamento	Validação
Número de pacientes	961	943
Idade no diagnóstico (Anos)*	61.26 (50,98, 70.22)	62.61 (51.90, 70.93)
Tempo de sobrevivência (Anos)*	9.70 (5.09, 15.43)	9.46 (5.03, 15.26)
Tamanho do tumor*	23 (18, 30)	23 (17, 30)
<b>Estágio do tumor</b>		
0	1	3
1	300	175
2	517	283
3	72	43
4	8	1
NA	63	438
<b>Estado ER</b>		
Positivo	772	687
Negativo	189	256
<b>Subtipos PAM50</b>		
Basal	69	130
HER2	77	143
LumA	436	243
LumB	255	206
Normal	37	103
NA	0	6

## 6.2 Resultados

### 6.2.1 Módulo “Darkred”

Para construção da assinatura referente ao módulo darkred, nós utilizamos os 18 genes (seção 5.2.3 ) do módulo que foram associados significativamente com o tempo de sobrevivência das pacientes com câncer de mama presentes no banco de dados METABRIC. A assinatura gerada é composta por 8 desses 18 genes (Figura 37). A Figura 38 apresenta a curva de Kaplan-Meier resultante do teste da assinatura com o conjunto de validação do METABRIC. É possível notar que a assinatura foi capaz de categorizar as pacientes nas diferentes categorias de risco, o que mostra que a assinatura possui um bom poder prognóstico para essa doença ( $p$ -valor =  $7e^{-15}$ , teste log-rank).

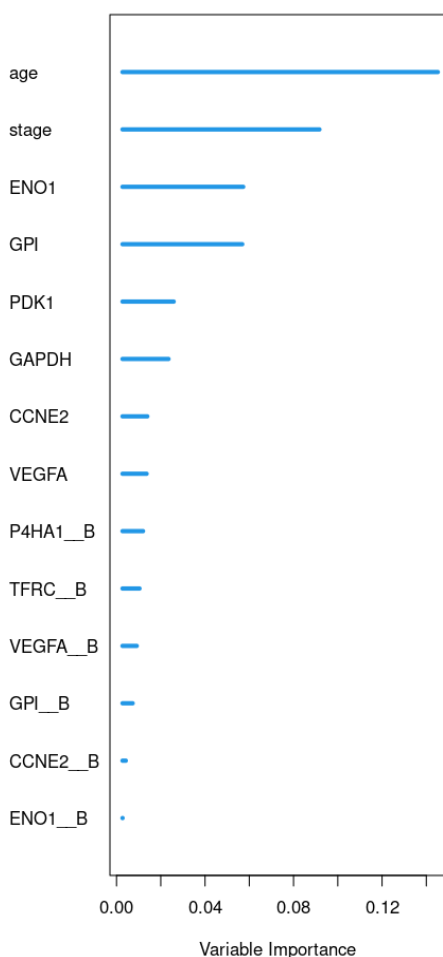


Figura 37: Atributos da assinatura gênica do módulo “Darkred” em ordem de importância para o modelo. O sufixo “\_\_B” representa o valor discreto da expressão do gene. A importância da variável (VIMP) foi calculada utilizando o método de Breiman-Cutler VIMP. Quanto maior o VIMP, maior é a capacidade preditiva da variável, enquanto VIMP negativo, ou igual a zero, indicam variáveis ruidosas (ISHWARAN et al., 2021).

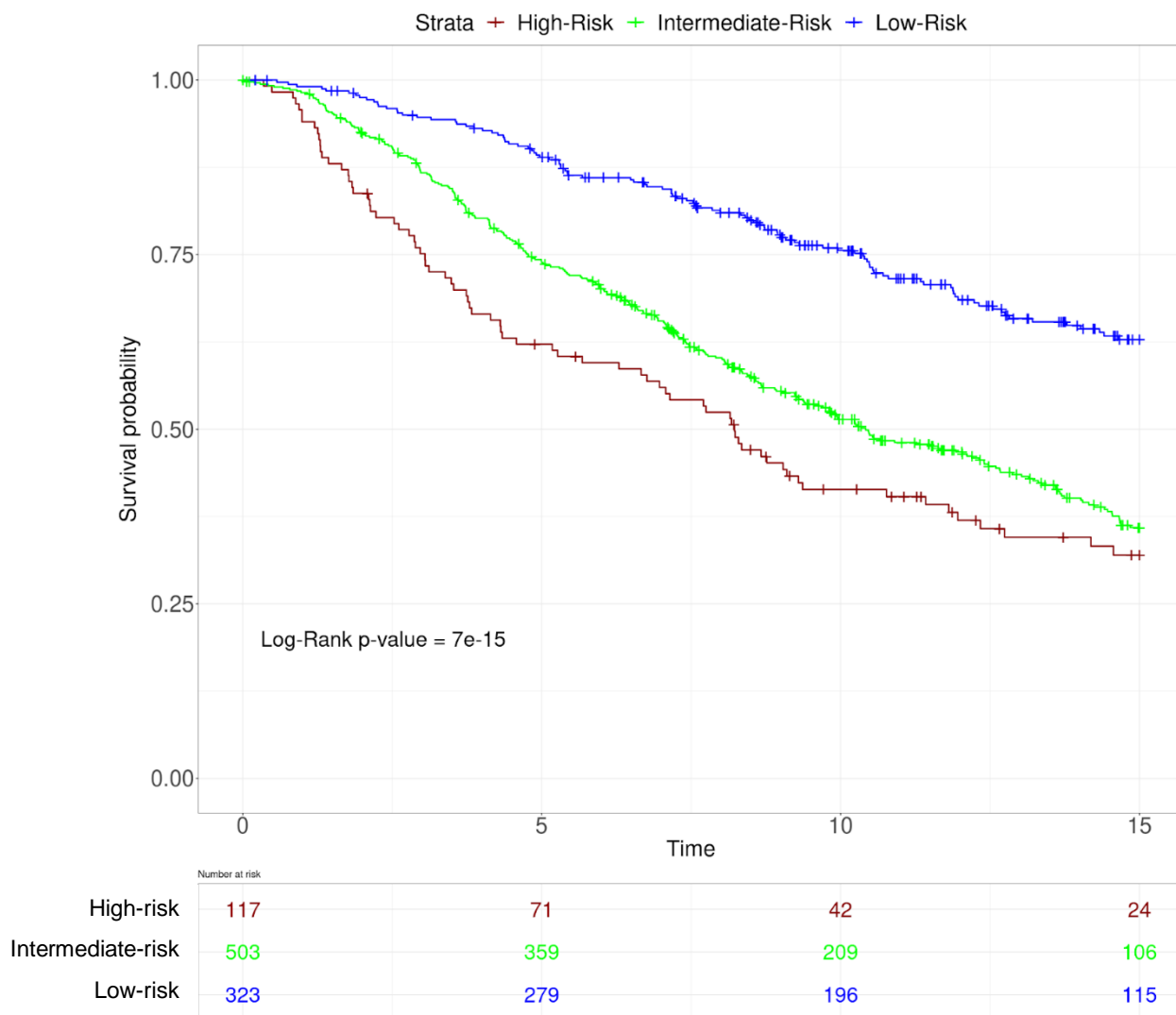


Figura 38: Curva de Kaplan-Meier da assinatura gênica do módulo “Darkred” aplicada no conjunto de validação do METABRIC. O eixo X representa o tempo desde o diagnóstico (em anos), e o eixo Y representa a probabilidade de sobrevivência. A tabela de risco (posicionada abaixo da curva de Kaplan-Meier) apresenta o número de pacientes vivos em cada um dos tempos. As categorias são: risco baixo, quando probabilidade de sobrevivência ( $P$ ) > 50% após 15 anos do diagnóstico da doença; risco médio: quando  $P > 50\%$  entre 7,5 e 15 anos; e risco alto nos demais casos. A teste de Log-Rank mostra que a assinatura foi capaz de distinguir as diferentes categorias de riscos das pacientes. Esta imagem, assim como as Figura 39, Figura 41, Figura 42, Figura 44, Figura 45, Figura 47, Figura 48, foram geradas pela função “*ggsurvplot*” do pacote “*survminer*”.

Também verificamos a performance da assinatura nos diferentes subtipos do câncer de mama (Figura 39). A assinatura foi eficiente em categorizar os pacientes nos subtipos *LumA* ( $p$ -valor  $\leq 2e^{-16}$ ) e *LumB* ( $p$ -valor =  $6e^{-3}$ ), mas para os subtipos *HER2* ( $p$ -valor = 0.8) e *Basal* ( $p$ -valor = 0.4) não foi eficiente.

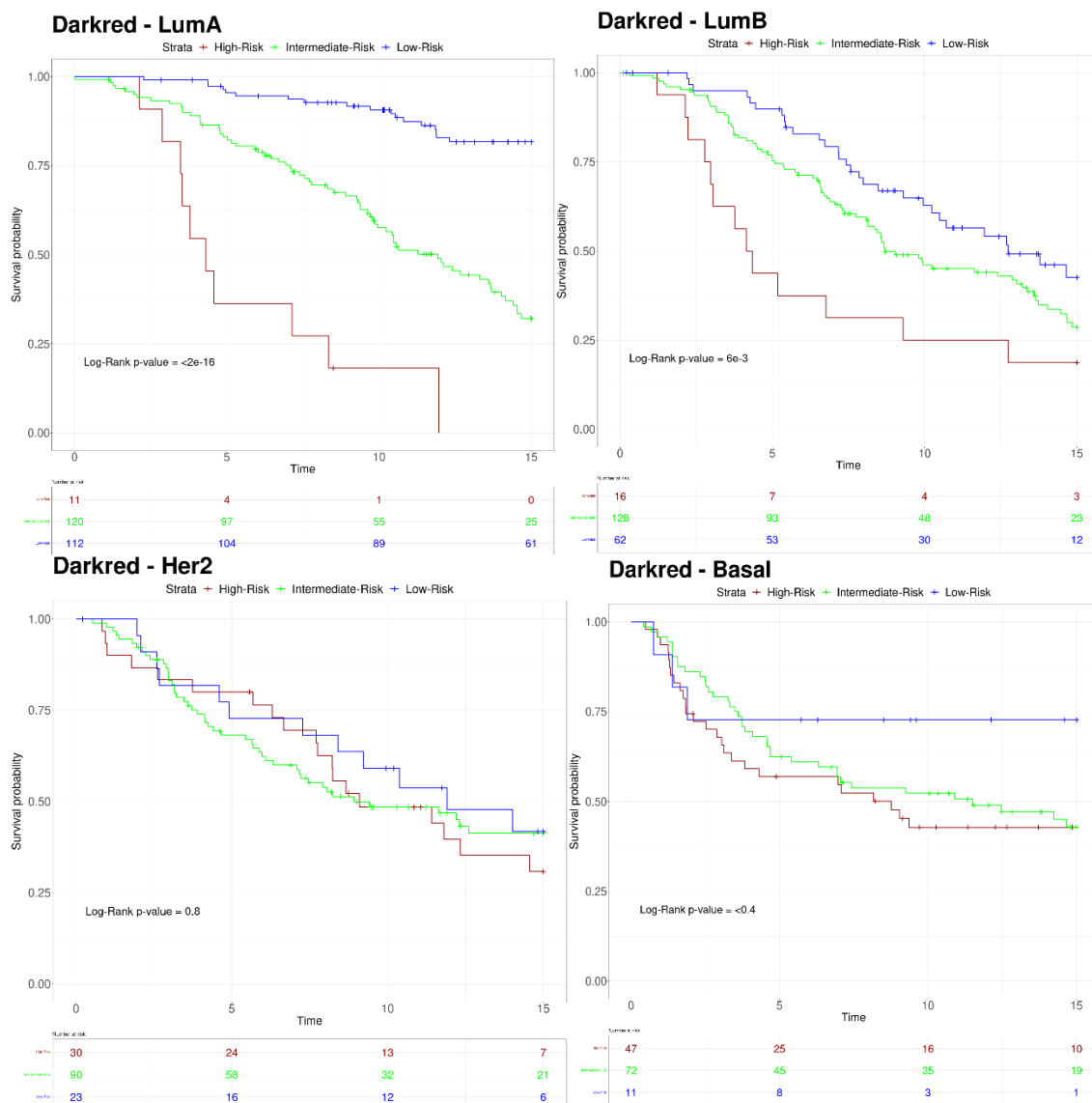


Figura 39: Curva de Kaplan-Meier da assinatura gênica do módulo “Darkred” aplicada para cada um dos subtipos PAM50 de câncer de mama utilizando o conjunto de validação do METABRIC. O eixo X representa o tempo desde o diagnóstico (em anos), e o eixo Y representa a probabilidade de sobrevivência. A tabela de risco (posicionada abaixo da curva de Kaplan-Meier) apresenta o número de pacientes vivas em cada um dos tempos. As categorias são: risco baixo, quando probabilidade de sobrevivência (P) > 50% após 15 anos do diagnóstico da doença; risco médio: quando P > 50% entre 7,5 e 15 anos; e risco alto nos demais casos. A teste de Log-Rank mostra que a assinatura apenas teve um bom desempenho em classificar as pacientes do subtipo Luminal (LumA e LumB) do câncer de mama, mas não foi eficaz nos subtipos mais agressivos (Her2 e Basal).

## 6.2.2 Módulo “Brown”

Para construção da assinatura referente ao módulo brown, nós utilizamos os 220 genes (seção 5.2.3) do módulo que foram associados significativamente com o tempo de sobrevivência das pacientes com câncer de mama presentes no banco de dados METABRIC. A assinatura gerada é composta por 29 desses 220 genes (Figura 40). A Figura 41 apresenta a curva de Kaplan-Meier resultante do teste da assinatura com o

conjunto de validação do METABRIC. Assim como o módulo “darkred”, a assinatura do módulo “brown” também obteve um bom poder prognóstico para essa doença (p-valor  $< 2e^{-16}$ , teste log-rank).

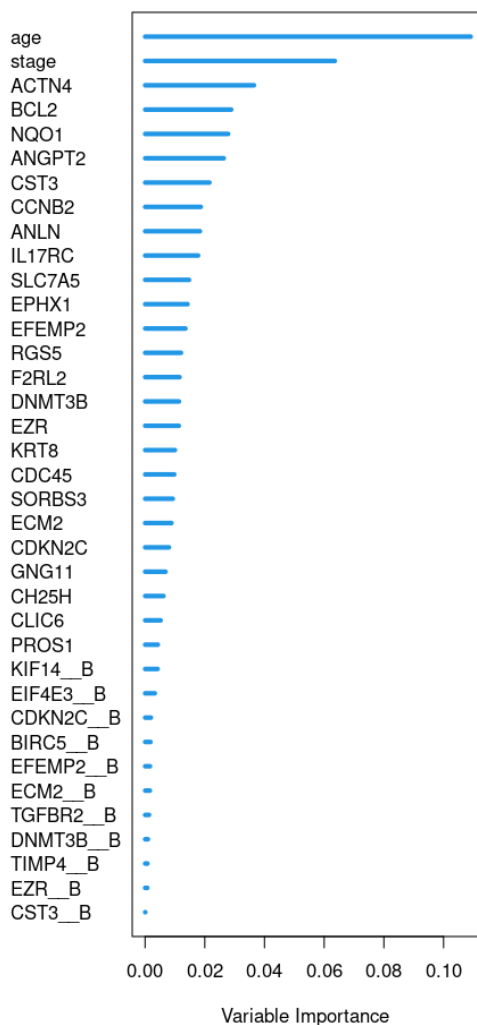


Figura 40: Atributos da assinatura gênica do módulo “Brown” em ordem de importância para o modelo. O sufixo “\_B” representa o valor discreto da expressão do gene. A importância da variável (VIMP) foi calculada utilizando o método de Breiman-Cutler VIMP. Quanto maior o VIMP, maior é a capacidade preditiva da variável, enquanto VIMP negativo, ou igual a zero, indicam variáveis ruidosas (ISHWARAN et al., 2021).

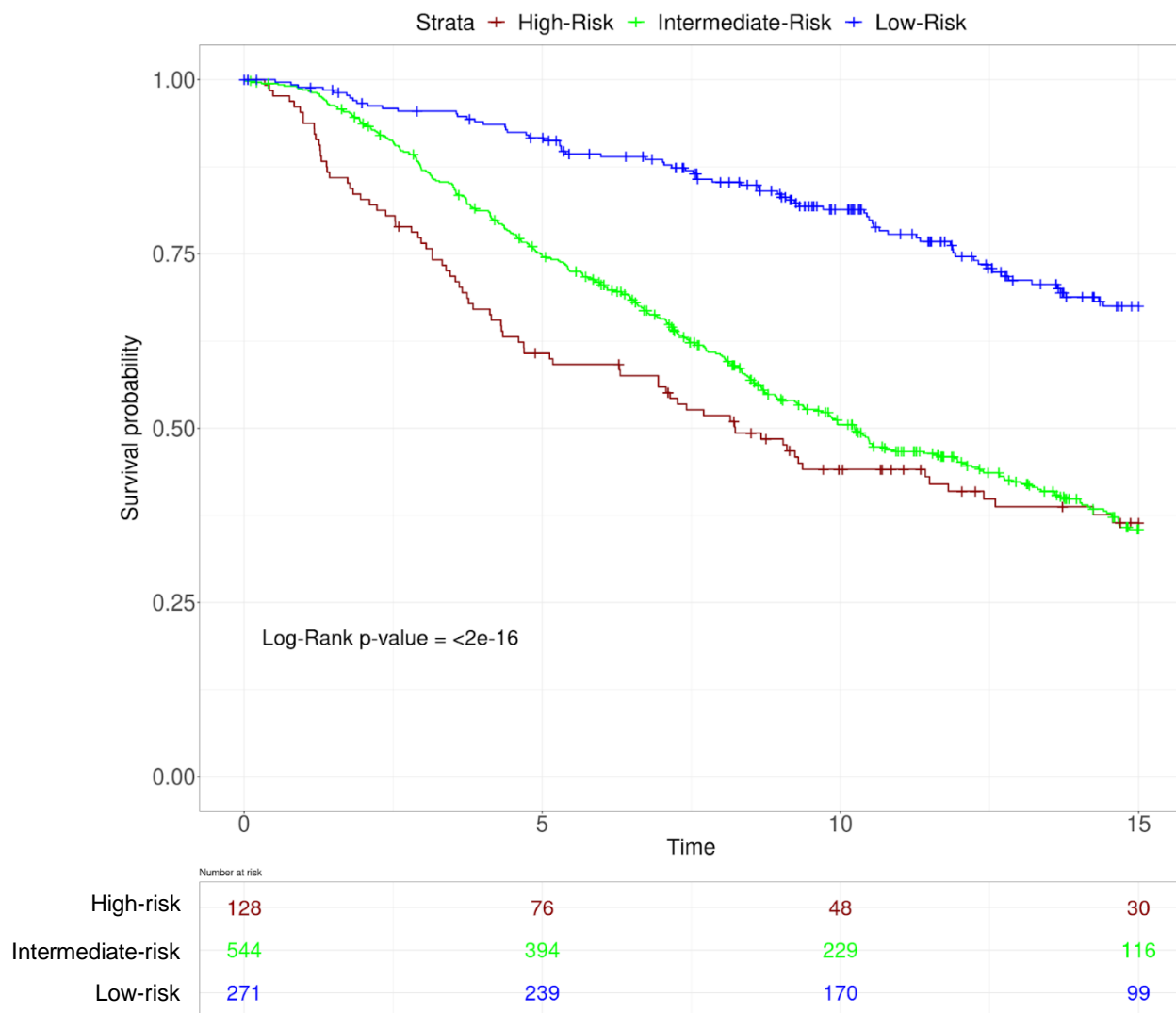


Figura 41: Curva de Kaplan-Meier da assinatura gênica do módulo “Brown” aplicada no conjunto de validação do METABRIC. O eixo X representa o tempo desde o diagnóstico (em anos), e o eixo Y representa a probabilidade de sobrevivência. A tabela de risco (posicionada abaixo da curva de Kaplan-Meier) apresenta o número de pacientes vivos em cada um dos tempos. As categorias são: risco baixo, quando probabilidade de sobrevivência ( $P$ )  $> 50\%$  após 15 anos do diagnóstico da doença; risco médio: quando  $P > 50\%$  entre 7,5 e 15 anos; e risco alto nos demais casos. A teste de Log-Rank mostra que a assinatura foi capaz de distinguir as diferentes categorias de riscos das pacientes. Porém, para pacientes das categorias de risco médio e alto que sobreviveram 15 ou mais, a assinatura não foi muito eficaz em separar essas categorias.

Também verificamos a performance da assinatura nos diferentes subtipos do câncer de mama (Figura 42). A assinatura foi eficiente em categorizar os pacientes nos subtipos *LumA* ( $p$ -valor =  $6e^{-9}$ ) e *LumB* ( $p$ -valor =  $2e^{-5}$ ), mas para os subtipos *HER2* ( $p$ -valor = 0.9) e *Basal* ( $p$ -valor = 0.5) não foi eficiente.

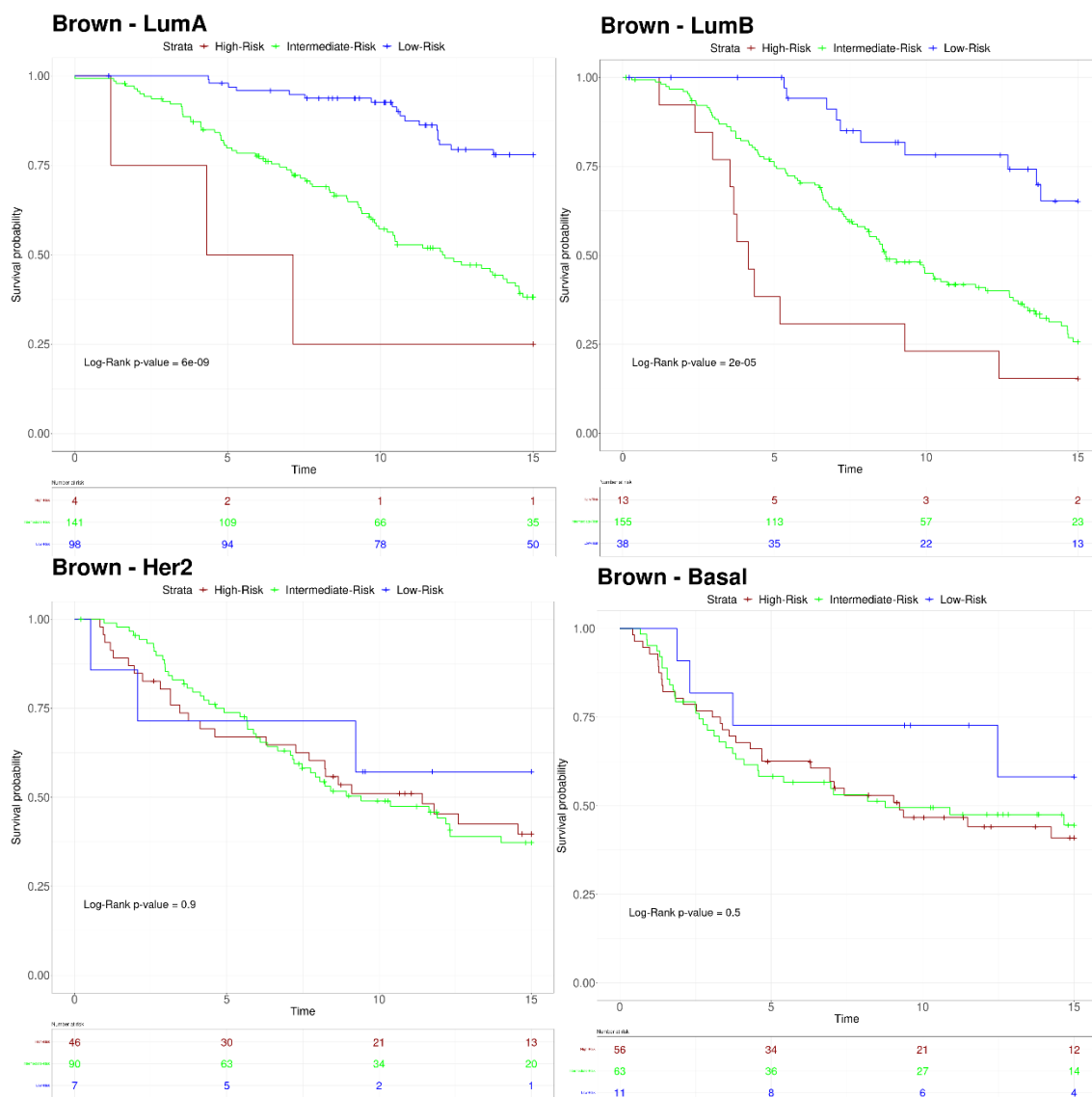


Figura 42: Curva de Kaplan-Meier da assinatura gênica do módulo “Brown” aplicada para cada um dos subtipos PAM50 de câncer de mama utilizando o conjunto de validação do METABRIC. O eixo X representa o tempo desde o diagnóstico (em anos), e o eixo Y representa a probabilidade de sobrevivência. A tabela de risco (posicionada abaixo da curva de Kaplan-Meier) apresenta o número de pacientes vivos em cada um dos tempos. As categorias são: risco baixo, quando probabilidade de sobrevivência ( $P$ ) > 50% após 15 anos do diagnóstico da doença; risco médio: quando  $P > 50\%$  entre 7,5 e 15 anos; e risco alto nos demais casos. A teste de Log-Rank mostra que a assinatura apenas teve um bom desempenho em classificar as pacientes do subtipo Luminal (LumA e LumB) do câncer de mama, mas não foi eficaz nos subtipos mais agressivos (Her2 e Basal).

### 6.2.3 Módulos “Darkred” + “Brown”

A assinatura gerada a partir da união dos genes dos módulos “darkred” e “brown” resultou em um modelo composto por 35 genes (Figura 43). A Figura 44 apresenta a curva de Kaplan-Meier resultante do teste da assinatura com o conjunto de validação do METABRIC.



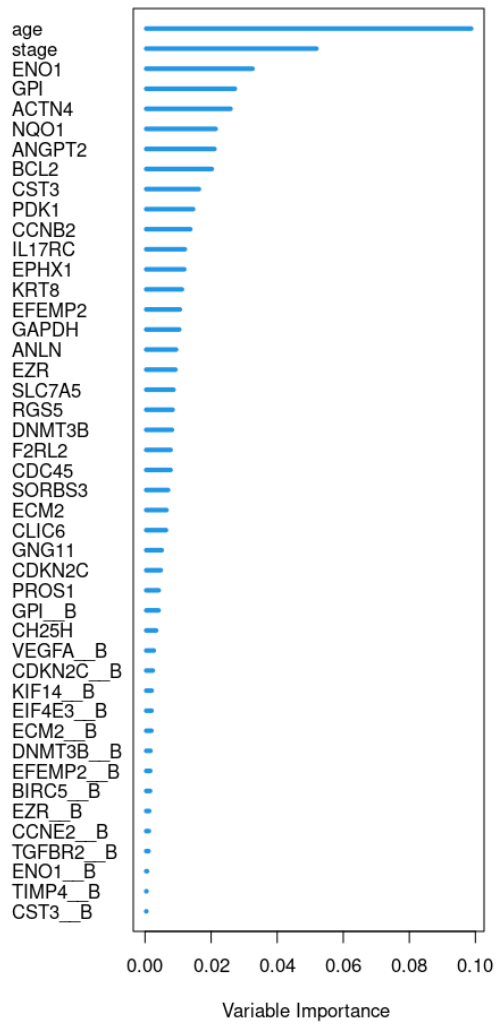


Figura 43: Atributos da assinatura gênica da união dos módulos “darkred” e “brown” em ordem de importância para o modelo. O sufixo “\_B” representa o valor discreto da expressão do gene. A importância da variável (VIMP) foi calculada utilizando o método de Breiman-Cutler VIMP. Quanto maior o VIMP, maior é a capacidade preditiva da variável, enquanto VIMP negativo, ou igual a zero, indicam variáveis ruidosas (ISHWARAN et al., 2021).

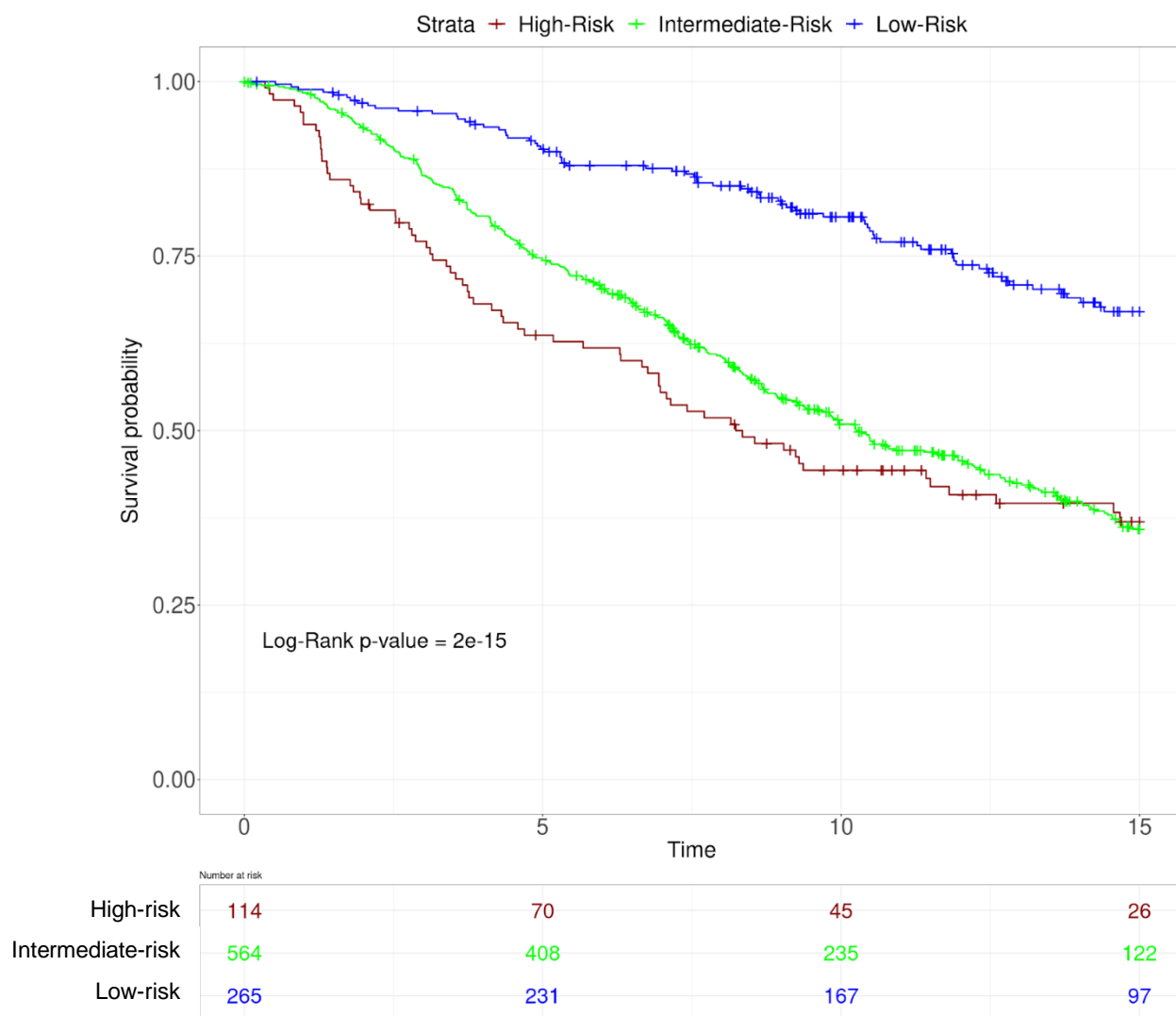


Figura 44: Curva de Kaplan-Meier da assinatura gênica da união dos módulos “Darkred” e “Brown” aplicada no conjunto de validação do METABRIC. O eixo X representa o tempo desde o diagnóstico (em anos), e o eixo Y representa a probabilidade de sobrevivência. A tabela de risco (posicionada abaixo da curva de Kaplan-Meier) apresenta o número de pacientes vivos em cada um dos tempos. As categorias são: risco baixo, quando probabilidade de sobrevivência ( $P$ ) > 50% após 15 anos do diagnóstico da doença; risco médio: quando  $P > 50\%$  entre 7,5 e 15 anos; e risco alto nos demais casos. A teste de Log-Rank mostra que a assinatura foi capaz de distinguir as diferentes categorias de riscos das pacientes. Porém, para pacientes das categorias de risco médio e alto que sobreviveram 15 ou mais, a assinatura não foi muito eficaz em separar essas categorias.

Também verificamos a performance da assinatura nos diferentes subtipos do câncer de mama (Figura 45). A assinatura foi eficiente em categorizar os pacientes nos subtipos *LumA* ( $p$ -valor =  $6e^{-10}$ ), *LumB* ( $p$ -valor =  $3e^{-7}$ ) e *Normal* ( $p$ -valor =  $6e^{-7}$ ), mas para os subtipos *HER2* ( $p$ -valor = 0.6) e *Basal* ( $p$ -valor = 0.4) não foi eficiente.

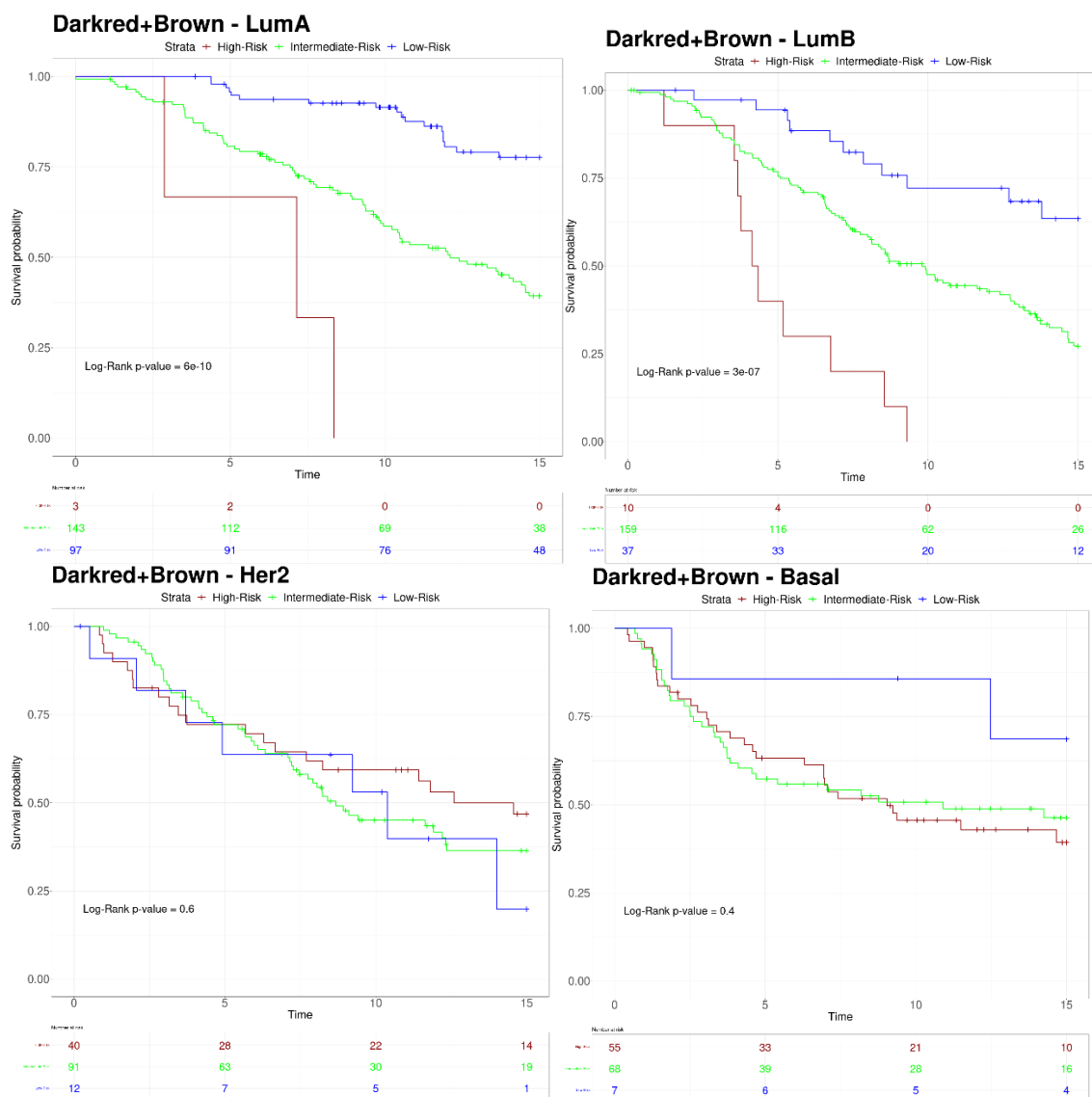


Figura 45: Curva de Kaplan-Meier da assinatura gênica da união dos módulos “Darkred e “Brown” aplicada para cada um dos subtipos PAM50 de câncer de mama utilizando o conjunto de validação do METABRIC. O eixo X representa o tempo desde o diagnóstico (em anos), e o eixo Y representa a probabilidade de sobrevivência. A tabela de risco (posicionada abaixo da curva de Kaplan-Meier) apresenta o número de pacientes vivos em cada um dos tempos. As categorias são: risco baixo, quando probabilidade de sobrevivência (P) > 50% após 15 anos do diagnóstico da doença; risco médio: quando P > 50% entre 7,5 e 15 anos; e risco alto nos demais casos. A teste de Log-Rank mostra que a assinatura apenas teve um bom desempenho em classificar as pacientes do subtipo Luminal (LumA e LumB) do câncer de mama, mas não foi eficaz nos subtipos mais agressivos (Her2 e Basal).

#### 6.2.4 Comparação com a assinatura gênica para angiogênese proposta em Guarischi-Sousa et al. (2019)

Decidimos comparar nossas assinaturas com a assinatura publicada em GUARISCHI-SOUSA et al. (2019) (Figura 46). Apesar de nossas assinaturas terem sido construídas a partir dos dados do METABRIC, esses dados foram coletados de fontes diferentes e em datas diferentes, podendo haver divergências entre os dados

utilizados no presente estudo e os dados utilizados por GUARISCHI-SOUSA et al. (2019). Por isso, nós seguimos os mesmos passos em GUARISCHI-SOUSA et al. (2019) para recriar seu modelo de assinatura utilizando os nossos dados (Figura 47 Figura 48).

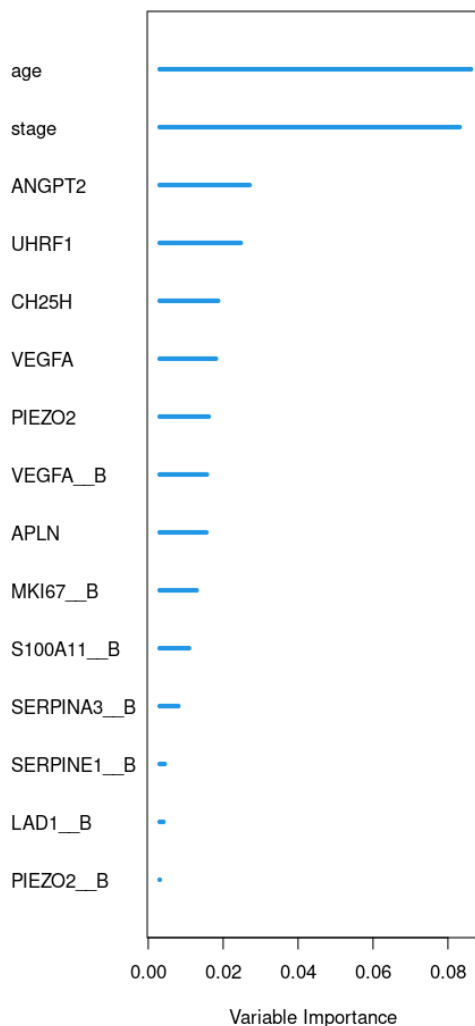


Figura 46: Atributos da assinatura gênica proposta por GUARISCHI-SOUSA et al. (2019) em ordem de importância para o modelo. O sufixo “\_\_B” representa o valor discreto da expressão do gene, sendo 0 para não-expresso e 1 para expresso. A importância da variável (VIMP) foi calculada utilizando o método de Breiman-Cutler VIMP. Quanto maior o VIMP, maior é a capacidade preditiva da variável, enquanto VIMP negativo, ou igual a zero, indicam variáveis ruidosas (ISHWARAN et al., 2021).

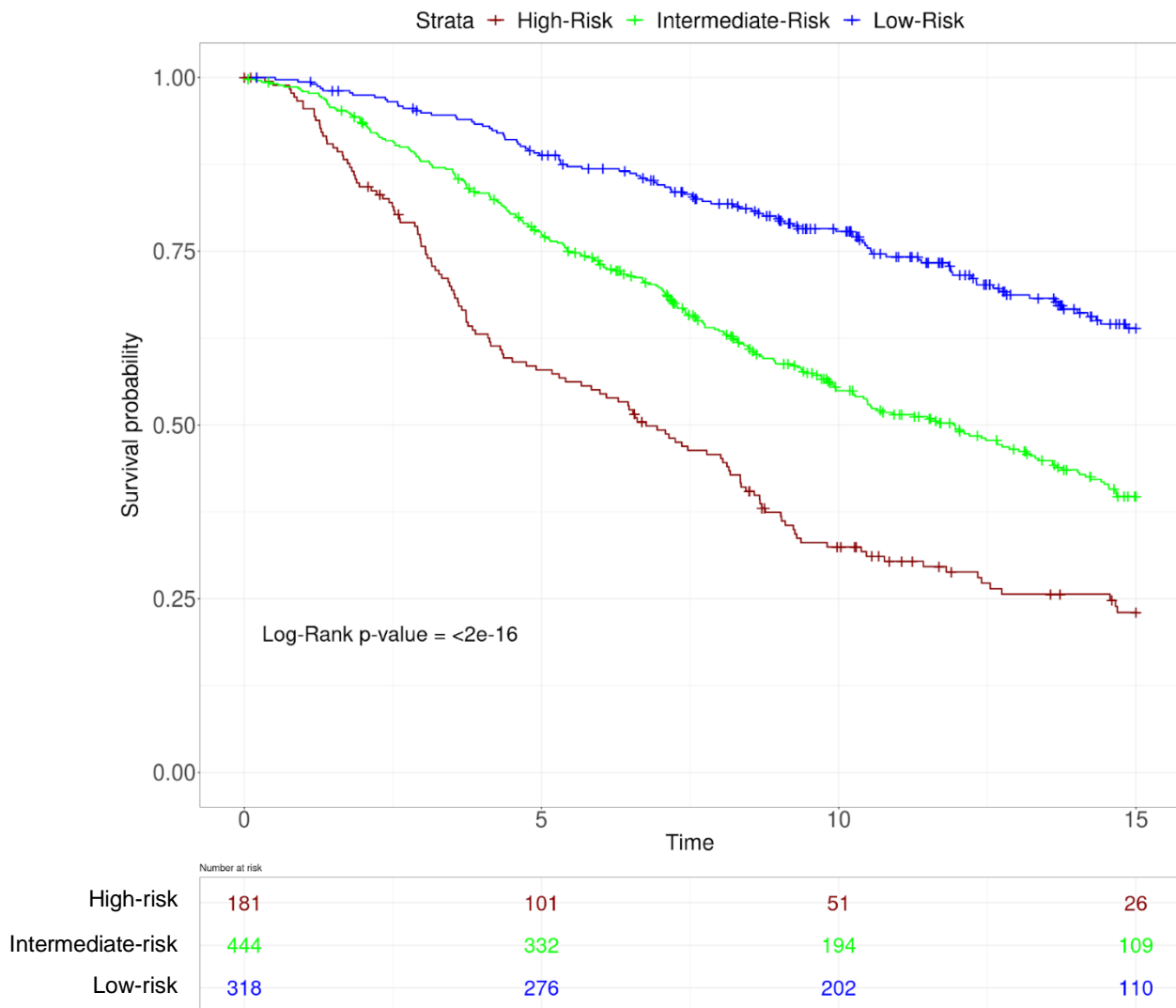


Figura 47: Curva de Kaplan-Meier da assinatura gênica proposta por GUARISCHI-SOUSA et al. (2019) aplicada no conjunto de validação do METABRIC. O eixo X representa o tempo desde o diagnóstico (em anos), e o eixo Y representa a probabilidade de sobrevivência. A tabela de risco (posicionada abaixo da curva de Kaplan-Meier) apresenta o número de pacientes vivos em cada um dos tempos. As categorias são: risco baixo, quando probabilidade de sobrevivência ( $P$ )  $> 50\%$  após 15 anos do diagnóstico da doença; risco médio: quando  $P > 50\%$  entre 7,5 e 15 anos; e risco alto nos demais casos. A teste de Log-Rank mostra que a assinatura foi capaz de distinguir as diferentes categorias de riscos das pacientes.

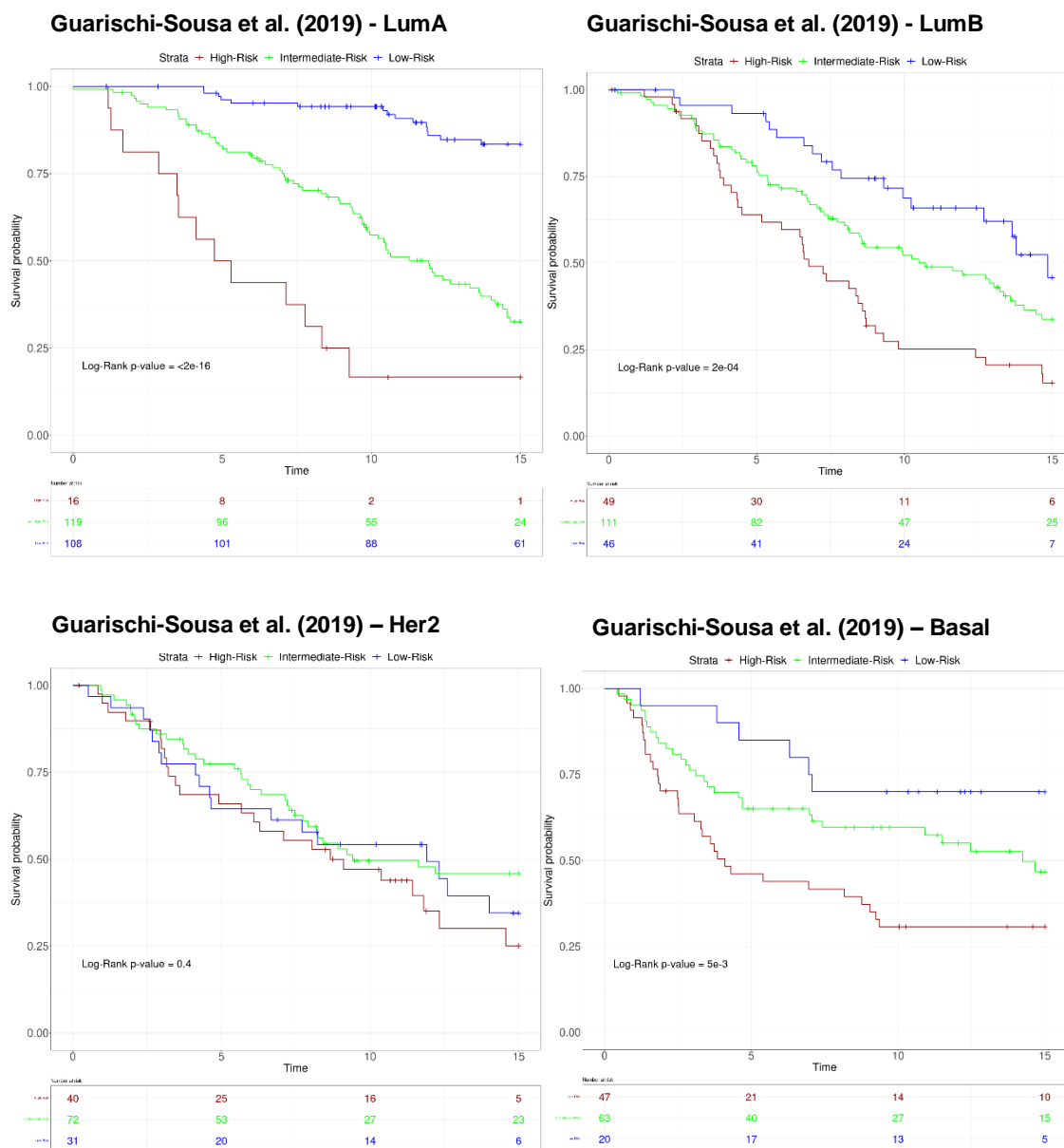


Figura 48: Curva de Kaplan-Meier da assinatura gênica proposta por GUARISCHI-SOUSA et al. (2019) aplicada para cada um dos subtipos PAM50 de câncer de mama utilizando o conjunto de validação do METABRIC. O eixo X representa o tempo desde o diagnóstico (em anos), e o eixo Y representa a probabilidade de sobrevivência. A tabela de risco (posicionada abaixo da curva de Kaplan-Meier) apresenta o número de pacientes vivos em cada um dos tempos. As categorias são: risco baixo, quando probabilidade de sobrevivência ( $P$ )  $> 50\%$  após 15 anos do diagnóstico da doença; risco médio: quando  $P > 50\%$  entre 7,5 e 15 anos; e risco alto nos demais casos. A teste de Log-Rank mostra que a assinatura teve um bom desempenho em classificar as pacientes do subtipo Luminal (LumA e LumB) e Basal, mas não foi eficaz para o subtipo Her2.

É possível verificar que a assinatura proposta em GUARISCHI-SOUSA et al. (2019) separa com mais eficiência os pacientes em suas categorias de risco, em comparação com as assinaturas desenvolvidas em nosso estudo, principalmente, para os subtipos de câncer de mama, em que a assinatura proposta por GUARISCHI-SOUSA et al. (2019) consegue distinguir significativamente os pacientes do subtipo Basal, enquanto as nossas assinaturas não conseguem. Porém, quando comparamos

a taxa erro das assinaturas, todas possuem uma performance semelhante, com uma pequena vantagem para a assinatura do módulo “Darkred” (Figura 49).

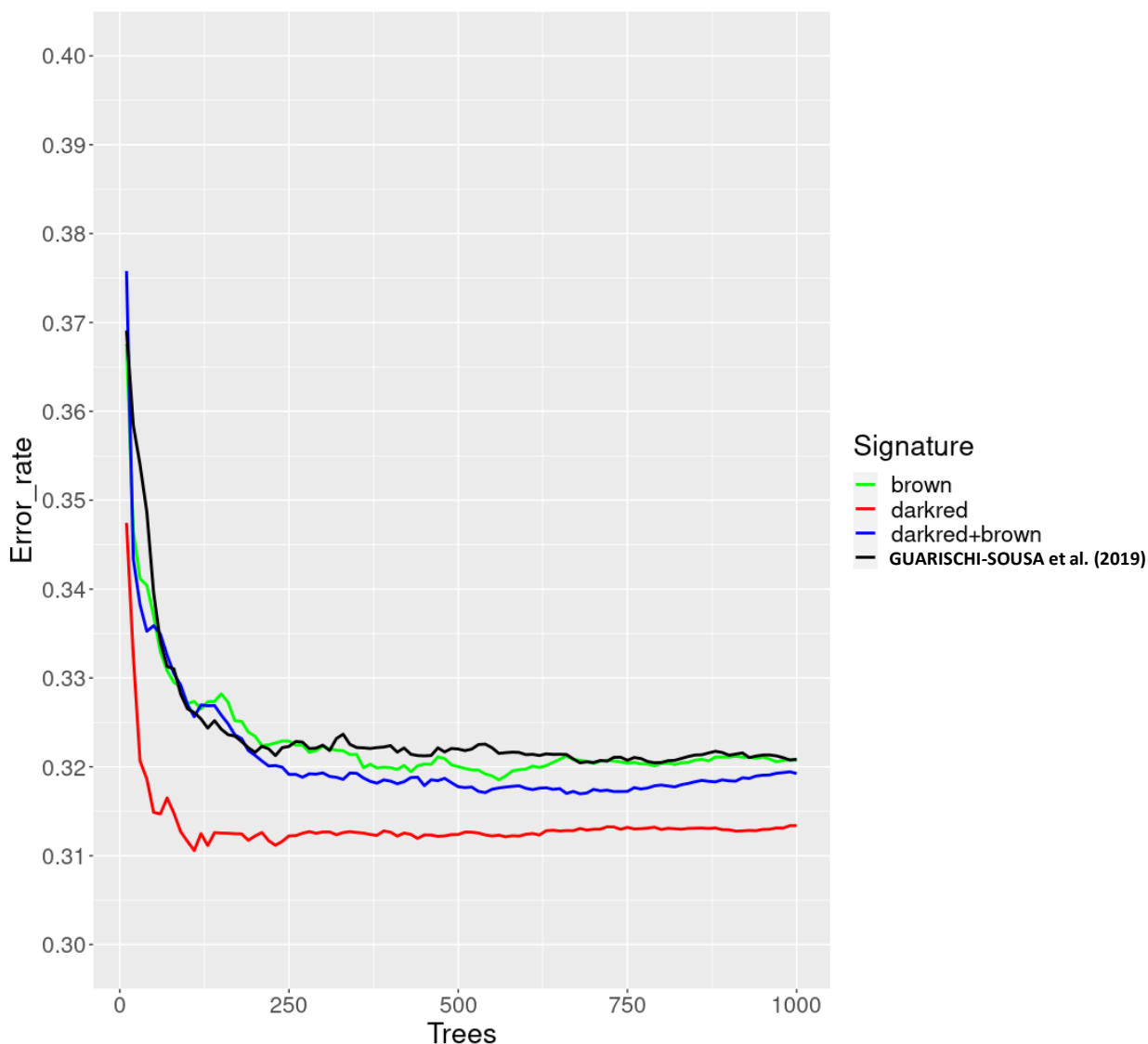


Figura 49: Taxa de erro de predição de cada uma das assinaturas apresentadas nesse estudo. Todas as assinaturas apresentam taxas de erro muito similares, com uma diferença inferior a 1%. Observa-se também que a taxa de erro atingiu uma estabilidade após 375 árvores geradas (eixo X), o que indica que o aumento do número de árvores geradas não traria nenhuma melhoria para as assinaturas.

### 6.3 Discussão

Nós geramos 3 modelos de assinatura gênica de angiogênese a partir dos genes coexpressos dos módulos “darkred” e “brown”, e todas as assinaturas foram capazes de discriminar significativamente as pacientes com câncer de mama nas 3 categorias de risco: baixo, intermediário e alto. No entanto, no caso das assinaturas que envolvem os genes do módulo “brown”, é possível observar que para os pacientes que foram classificados nos grupos de risco intermediário e alto, as curvas de

sobrevivência se cruzam quando se aproximam dos 15 anos (Figura 41 e Figura 44), isso mostra uma imprecisão do modelo para distinguir o grupo de risco desses pacientes, pois era esperado que os pacientes do grupo de alto risco tivessem uma probabilidade de sobrevivência significativamente menor que os pacientes do grupo intermediário. Uma hipótese seria que uma das possíveis causas desse comportamento pode estar relacionada com o número maior de variáveis (genes) que o módulo “brown” traz para as assinaturas, aumentando a complexidade do modelo, e assim, dificultando o algoritmo de Random Forest de identificar um padrão na expressão desses conjuntos de genes entre os pacientes. É necessário rever os filtros aplicados e talvez deixá-los mais restritivos, para tentar eliminar os genes que estão causando a imprecisão das assinaturas e assim tornar o modelo mais simples.

A assinatura gênica apresentada por GUARISCHI-SOUSA et al. (2019) foi construída a partir dos genes DE identificados no modelo OIR. Como nossas assinaturas foram construídas a partir de genes co-expressos, abrangendo um número maior de genes com funções relevantes para angiogênese, hipotizávamos que nossas assinaturas poderiam ser ainda mais eficientes no prognóstico de pacientes com câncer de mama, mas não foi o que observamos quando comparamos os resultados. A assinatura proposta por GUARISCHI-SOUSA et al. (2019) conseguiu classificar as pacientes com subtipo Basal nos diferentes grupos de risco, enquanto nossas assinaturas não foram capazes de fazer o mesmo. Também não sofreu com o problema observado para as assinaturas com os genes do módulo “brown”, as quais não foram tão eficientes em distinguir as pacientes entre os grupos de risco intermediário e alto. No entanto, não podemos descartar as contribuições que as novas assinaturas podem trazer, uma vez que essas assinaturas são compostas por um conjunto de genes diferentes da assinatura de GUARISCHI-SOUSA et al. (2019). Por exemplo, a assinatura do módulo “darkred” traz os genes da via glicolítica que contribuem para progressão da angiogênese em alguns tipos de câncer, como discutimos na seção 5.3 . Isso amplia o número de genes que podem ser alvos para desenvolvimento de novos fármacos para tratamento da doença, ou ainda como novos marcadores de doenças dependentes de angiogênese.



## 7 Conclusão

O objetivo geral deste trabalho foi conhecer melhor os mecanismos moleculares da angiogênese, identificando os principais genes, codificantes e não-codificantes, que atuam nesse processo e como eles se relacionam entre si. Nós propomos duas redes, uma apelidada de “darkred” e a outra de “brown”, devido às cores que foram atribuídas aos módulos de coexpressão gênica os quais são a base dessas redes. As análises de enriquecimento funcional nos mostraram que os genes que compõem essas redes não só estão relacionados com o processo de angiogênese, mas como também representam as fases do experimento executado (seção 5.2.4 ). Além disso, essas análises revelaram RNAs importantes para regulação da angiogênese que podem nortear novas pesquisas de desenvolvimento de fármacos para tratamento de doenças dependentes de angiogênese, ou ainda em pesquisas que buscam identificar novos marcadores para as mesmas doenças. Com base nisso, podemos afirmar que nosso objetivo foi alcançado.

Nossos resultados reforçam a capacidade do modelo OIR em reproduzir a angiogênese induzida por processos patológicos em humanos, pois conseguimos observar no nosso modelo algumas associações que já haviam sido feitas em células tumorais humanas, como por exemplo a dependência entre a via glicolítica e a de resposta a hipóxia presente na rede “darkred” (seção 5.3 ). Outro exemplo é o lncRNA Gm15283, que se trata de um transcrito não caracterizado em camundongo, mas que pelo contexto da rede em que ele se encontra e sua posição no genoma do camundongo, muito provavelmente se trata de um homólogo da família de lncRNAs *HIF1A-AS* descritos em humanos, cujos papéis na manutenção da angiogênese em alguns tipos de câncer já foram descritos em estudos anteriores (seção 5.3 ).

Houve algumas limitações em nosso estudo que não permitiram explorar mais a fundo o papel de alguns RNAs na angiogênese. A principal delas é referente a falta de um sequenciamento do tipo *paired-end*, que trouxe dificuldades na quantificação das variantes dos 153 genes DE (seção 3.3 ), e na identificação e quantificação dos circRNAs (seção 4.2.5 ). A outra foi a falta de uma metodologia capaz de enriquecer os circRNAs na amostra e assim ter um melhor aproveitamento do sequenciamento desse tipo de RNA. Por isso sugerimos que estudos futuros da expressão diferencial de variantes de mRNAs, ou que buscam por circRNAs, sejam realizados por sequenciamentos do tipo *paired-end* ou utilizando-se das novas plataformas de sequenciamento (PacBio e/ou Nanopore) que conseguem gerar reads mais longas, e, portanto, seriam suficientes para superar as limitações do nosso trabalho.

Também desenvolvemos novas assinaturas gênicas de angiogênese para prognósticos de pacientes com câncer de mama. Conforme descrito na seção 6.2 , obtivemos bons resultados nesta tarefa, já que todos os modelos de assinaturas proposta conseguiram categorizar as pacientes nas 3 distintas categorias de risco de forma significativa segundo os testes de log-rank, mas existe espaço para melhorar suas eficiências (seção 6.3 ). Para essas assinaturas nós enxergamos algumas investigações que podem ser interessantes de serem realizadas futuramente. Uma delas é verificar se existem mutações conhecidas para os genes que compõem as assinaturas, e se essas mutações possuem alguma associação com o prognóstico das pacientes com câncer de mama.

Em resumo, nosso trabalho apresentou quatro principais contribuições:

- 1) uma lista de variantes DEs que pertencem aos 153 genes DE identificados por GUARISCHI-SOUSA et al. (2019)
- 2) uma lista de ncRNAs DE também no modelo OIR
- 3) 2 redes de regulação gênica para angiogênese patológica que representam as fases do experimento do nosso modelo animal
- 4) 3 novas assinaturas gênicas com conjunto de genes diferentes do proposto por GUARISCHI-SOUSA et al. (2019).

## Referências

- ADAIR, T. H.; MONTANI, J.-P. **Angiogenesis**. San Rafael (CA): Morgan & Claypool Life Sciences, 2010.
- AGARWAL, V. et al. Predicting effective microRNA target sites in mammalian mRNAs. **eLife**, v. 4, n. AUGUST2015, 12 ago. 2015.
- AKSHAYAA, S. et al. Exploring pain insensitivity inducing gene ZFH2 by using deep convolutional neural network. **Proceedings of the 3rd International Conference on Computing Methodologies and Communication, ICCMC 2019**, p. 68–72, 1 mar. 2019.
- ALTSCHUL, S. F. et al. Basic local alignment search tool. **Journal of Molecular Biology**, v. 215, n. 3, p. 403–410, 5 out. 1990.
- ARRANZ, E. E. et al. Gene Signatures in Breast Cancer: Current and Future Uses. **Translational Oncology**, v. 5, n. 6, p. 398, 2012.
- ASHBURNER, M. et al. Gene Ontology: tool for the unification of biology. **Nature Genetics**, v. 25, n. 1, p. 25–29, maio 2000.
- BARABÁSI, A. L.; OLTVAI, Z. N. Network biology: understanding the cell's functional organization. **Nature Reviews Genetics 2004 5:2**, v. 5, n. 2, p. 101–113, fev. 2004.
- BARCLAY, C. et al. Basic fibroblast growth factor (FGF-2) overexpression is a risk factor for esophageal cancer recurrence and reduced survival, which is ameliorated by coexpression of the FGF-2 antisense gene. **Clinical Cancer Research**, v. 11, n. 21, p. 7683–7691, 1 nov. 2005.
- BERNABEU, C.; LOPEZ-NOVOA, J. M.; QUINTANILLA, M. The emerging role of TGF-beta superfamily coreceptors in cancer. **Biochimica et biophysica acta**, v. 1792, n. 10, p. 954–973, out. 2009.
- BERNAL, A.; ARRANZ, L. Nestin-expressing progenitor cells: function, identity and therapeutic implications. **Cellular and molecular life sciences : CMLS**, v. 75, n. 12, p. 2177–2195, 1 jun. 2018.
- BETZ, C. et al. Cell behaviors and dynamics during angiogenesis. **Development (Cambridge, England)**, v. 143, n. 13, p. 2249–60, 1 jul. 2016.
- BIELAJEW, B. J.; HU, J. C.; ATHANASIOU, K. A. Collagen: quantification, biomechanics and role of minor subtypes in cartilage. **Nature Reviews Materials 2020 5:10**, v. 5, n. 10, p. 730–747, 20 jul. 2020.
- BLANQUART, S. et al. **Assisted transcriptome reconstruction and splicing orthology**. BMC Genomics. **Anais...BioMed Central Ltd.**, 11 nov. 2016. Disponível em: <<http://bmcgenomics.biomedcentral.com/articles/10.1186/s12864-016-3103-6>>. Acesso em: 1 nov. 2020
- BRIDGETT, S.; DELLETT, M.; SIMPSON, D. A. RNA-Sequencing data supports the existence of novel VEGFA splicing events but not of VEGFAxxx isoforms. **Scientific reports**, v. 7, n. 1, p. 58, 2017.
- BULT, C. J. et al. **The alliance of genome resources: Building a modern data ecosystem for model organism databases**. **Genetics** Genetics Society of America, , 1 dez. 2019. Disponível em: <<https://doi.org/10.1534/genetics.119.302523>>. Acesso em: 30 out. 2020
- CAMARGO, A. P. et al. RNAsamba: neural network-based assessment of the protein-coding potential of RNA sequences. **NAR Genomics and Bioinformatics**, v. 2, n. 1, 1 mar. 2020.
- CERAMI, E. et al. The cBio Cancer Genomics Portal: An Open Platform for Exploring Multidimensional Cancer Genomics Data. **Cancer Discovery**, v. 2, n. 5, p. 401–404, 1 maio 2012.
- CHEN, C.; LIU, J.; XU, G. Overexpression of PIWI proteins in human stage III epithelial ovarian cancer with lymph node metastasis. **Cancer biomarkers : section A of Disease markers**, v. 13, n. 5, p. 315–321, 2013.

- CHEN, C. Y. et al. Lengthening of 3'UTR increases with morphological complexity in animal evolution. **Bioinformatics (Oxford, England)**, v. 28, n. 24, p. 3178–3181, dez. 2012.
- CHEN, H. et al. **DIFFUSE: Predicting isoform functions from sequences and expression profiles via deep learning**. *Bioinformatics*. Anais...Oxford University Press, 15 jul. 2019. Disponível em: <<https://github.com/haochenucr/DIFFUSE>>. Acesso em: 28 nov. 2020
- CHEN, L. C. et al. A novel role for TNFAIP2: its correlation with invasion and metastasis in nasopharyngeal carcinoma. **Modern Pathology** 2011 **24:2**, v. 24, n. 2, p. 175–184, 5 nov. 2010.
- CHEN, L.-L. The biogenesis and emerging roles of circular RNAs. **Nature Reviews Molecular Cell Biology**, v. 17, n. 4, p. 205–211, 24 abr. 2016.
- CHEN, X. et al. Characterization of microRNAs in serum: a novel class of biomarkers for diagnosis of cancer and other diseases. **Cell Research**, v. 18, n. 10, p. 997–1006, 2 out. 2008.
- CHEN, Y. et al. **edgeR: differential analysis of sequence read count data User's Guide**. [s.l.: s.n.]. . Acesso em: 30 out. 2020.
- CHUNHAROJRITH, P. et al. Tumor suppression by MEG3 lncRNA in a human pituitary tumor derived cell line. **Molecular and Cellular Endocrinology**, v. 416, p. 27–35, 15 nov. 2015.
- CRAWFORD, T. N. et al. Diabetic retinopathy and angiogenesis. **Current diabetes reviews**, v. 5, n. 1, p. 8–13, fev. 2009.
- CUI, H. et al. GPR126 Protein Regulates Developmental and Pathological Angiogenesis through Modulation of VEGFR2 Receptor Signaling. **The Journal of Biological Chemistry**, v. 289, n. 50, p. 34871, 12 dez. 2014.
- CUI, H. et al. GPR126 regulates colorectal cancer cell proliferation by mediating HDAC2 and GLI2 expression. **Cancer Science**, v. 112, n. 5, p. 1798, 1 maio 2021.
- CUNNINGHAM, F. et al. Ensembl 2019. **Nucleic Acids Research**, v. 47, n. D1, p. D745–D751, 8 jan. 2019.
- CURTIS, C. et al. The genomic and transcriptomic architecture of 2,000 breast tumours reveals novel subgroups. **Nature**, v. 486, n. 7403, p. 346, 21 jun. 2012.
- DELGADO, F. M.; GÓMEZ-VELA, F. **Computational methods for Gene Regulatory Networks reconstruction and analysis: A review**. *Artificial Intelligence in Medicine* Elsevier B.V., , 1 abr. 2019. . Acesso em: 8 dez. 2020
- DHANOJA, J. K. et al. Long non-coding RNA: its evolutionary relics and biological implications in mammals: a review. **Journal of animal science and technology**, v. 60, p. 25, 2018.
- DISCHER, D. E. et al. Matrix Mechanosensing: From Scaling Concepts in 'Omics Data to Mechanisms in the Nucleus, Regeneration, and Cancer. **Annual review of biophysics**, v. 46, p. 295–315, 22 maio 2017.
- DRAOUI, N.; DE ZEEUW, P.; CARMELIET, P. Angiogenesis revisited from a metabolic perspective: role and therapeutic implications of endothelial cell metabolism. **Open biology**, v. 7, n. 12, 2017.
- DU, W. et al. Endothelial Cell Glucose Metabolism and Angiogenesis. **Biomedicines**, v. 9, n. 2, p. 1–17, 1 fev. 2021.
- DUSART, P. et al. A systems-approach reveals human nestin is an endothelial-enriched, angiogenesis-independent intermediate filament protein. **Scientific Reports** 2018 **8:1**, v. 8, n. 1, p. 1–15, 2 out. 2018.
- EKSI, R. et al. Systematically Differentiating Functions for Alternatively Spliced Isoforms through Integrating RNA-seq Data. **PLoS Computational Biology**, v. 9, n. 11, p. e1003314, 7 nov. 2013.
- ENUKA, Y. et al. Circular RNAs are long-lived and display only minimal early alterations in response to a growth factor. **Nucleic Acids Research**, v. 44, n. 3, p. 1370, 2 fev. 2016.

- ESWARAPPA, S. M. et al. Programmed translational readthrough generates antiangiogenic VEGF-Ax. **Cell**, v. 157, n. 7, p. 1605–1618, 19 jun. 2014.
- FARHADIAN, M. et al. Weighted gene co-expression network analysis identifies modules and functionally enriched pathways in the lactation process. **Scientific Reports** **2021** **11:1**, v. 11, n. 1, p. 1–15, 27 jan. 2021.
- FASANARO, P. et al. MicroRNA-210 modulates endothelial cell response to hypoxia and inhibits the receptor tyrosine kinase ligand Ephrin-A3. **The Journal of biological chemistry**, v. 283, n. 23, p. 15878–83, 6 jun. 2008.
- FINN, R. D.; CLEMENTS, J.; EDDY, S. R. HMMER web server: interactive sequence similarity searching. **Nucleic acids research**, v. 39, n. Web Server issue, p. W29–37, jul. 2011.
- FISH, J. E. et al. miR-126 regulates angiogenic signaling and vascular integrity. **Developmental cell**, v. 15, n. 2, p. 272–84, ago. 2008.
- FOLKMAN, J. Angiogenesis: an organizing principle for drug discovery? **Nature Reviews Drug Discovery** **2007** **6:4**, v. 6, n. 4, p. 273–286, abr. 2007.
- FRANKISH, A. et al. GENCODE reference annotation for the human and mouse genomes. **Nucleic Acids Research**, v. 47, n. D1, p. D766–D773, 8 jan. 2019.
- FRIEDLÄNDER, M. R. et al. miRDeep2 accurately identifies known and hundreds of novel microRNA genes in seven animal clades. **Nucleic acids research**, v. 40, n. 1, p. 37–52, jan. 2012.
- FU, W. et al. Effect of microRNA-144-5p on the proliferation, invasion and migration of human umbilical vein endothelial cells by targeting SMAD1. **Experimental and Therapeutic Medicine**, v. 19, n. 1, p. 165–171, 1 jan. 2020.
- FU, W.-M. et al. Long noncoding RNA Hotair mediated angiogenesis in nasopharyngeal carcinoma by direct and indirect signaling pathways. **Oncotarget**, v. 7, n. 4, p. 4712–23, 26 jan. 2016.
- GAD, A. A.; BALENGA, N. The Emerging Role of Adhesion GPCRs in Cancer. **ACS Pharmacology & Translational Science**, v. 3, n. 1, p. 29, 2 fev. 2020.
- GALLARDO-VARA, E. et al. Soluble endoglin regulates expression of angiogenesis-related proteins and induction of arteriovenous malformations in a mouse model of hereditary hemorrhagic telangiectasia. **Disease models & mechanisms**, v. 11, n. 9, 1 set. 2018.
- GANAPATHI, M. K. et al. Expression profile of COL2A1 and the pseudogene SLC6A10P predicts tumor recurrence in high-grade serous ovarian cancer. **International journal of cancer**, v. 138, n. 3, p. 679–688, 1 fev. 2016.
- GAO, J. et al. Integrative analysis of complex cancer genomics and clinical profiles using the cBioPortal. **Science signaling**, v. 6, n. 269, 2 abr. 2013.
- GAO, Y.; WANG, J.; ZHAO, F. CIRI: an efficient and unbiased algorithm for de novo circular RNA identification. **Genome Biology**, v. 16, n. 1, p. 4, 13 jan. 2015.
- GAO, Y.; ZHANG, J.; ZHAO, F. Circular RNA identification based on multiple seed matching. **Briefings in bioinformatics**, v. 19, n. 5, p. 803–810, 28 set. 2018.
- GATTER, T.; STADLER, P. F. Ryūtō: Network-flow based transcriptome reconstruction. **BMC Bioinformatics**, v. 20, n. 1, p. 190, 16 abr. 2019.
- GIORDANO, R. J.; PASQUALINI, R.; ARAP, W. Mecanismos de angiogênese e metástase. Em: FERREIRA, C. G.; CASALI, J. C. (Eds.). **Oncologia Molecular**. 1. ed. São Paulo: Atheneu, 2010. p. 139–149.
- GLAŽAR, P.; PAPAVALASILEIOU, P.; RAJEWSKY, N. CircBase: A database for circular RNAs. **RNA**, v. 20, n. 11, p. 1666–1670, 1 nov. 2014.
- GORDON, F. E. et al. Increased Expression of Angiogenic Genes in the Brains of Mouse *Meg3*-Null Embryos. **Endocrinology**, v. 151, n. 6, p. 2443–2452, jun. 2010.
- GRIFFITHS-JONES, S. et al. Rfam: an RNA family database. **Nucleic Acids Research**, v. 31, n. 1, p. 439–441, 1 jan. 2003.
- GU, T. et al. miR-30c inhibits angiogenesis by targeting delta-like ligand 4 in liver sinusoidal endothelial cell to attenuate liver fibrosis. **FASEB Journal**, v. 35, n. 5, 1 maio 2021.

- GUARISCHI-SOUSA, R. et al. A transcriptome-based signature of pathological angiogenesis predicts breast cancer patient survival. **PLOS Genetics**, v. 15, n. 12, p. e1008482, 17 dez. 2019.
- HA, M.; KIM, V. N. Regulation of microRNA biogenesis. **Nature Reviews Molecular Cell Biology**, v. 15, n. 8, p. 509–524, 16 ago. 2014.
- HABIB, A. M. et al. A novel human pain insensitivity disorder caused by a point mutation in ZFH2. **Brain**, v. 141, n. 2, p. 365, 1 fev. 2018.
- HARBECK, N. et al. Breast cancer. **Nature Reviews Disease Primers 2019 5:1**, v. 5, n. 1, p. 1–31, 23 set. 2019.
- HARRELL, F. E. et al. Evaluating the Yield of Medical Tests. **JAMA**, v. 247, n. 18, p. 2543–2546, 14 maio 1982.
- HARROW, J. et al. GENCODE: the reference human genome annotation for The ENCODE Project. **Genome research**, v. 22, n. 9, p. 1760–1774, set. 2012.
- HARTNETT, M. E.; PENN, J. S. Mechanisms and management of retinopathy of prematurity. **The New England journal of medicine**, v. 367, n. 26, p. 2515–26, 27 dez. 2012.
- HE, Y. et al. PGK1 contributes to tumorigenesis and sorafenib resistance of renal clear cell carcinoma via activating CXCR4/ERK signaling pathway and accelerating glycolysis. **Cell Death & Disease 2022 13:2**, v. 13, n. 2, p. 1–15, 4 fev. 2022.
- HENAOUI, I. S. et al. PIWI-interacting RNAs as novel regulators of pancreatic beta cell function. **Diabetologia**, v. 60, n. 10, p. 1977–1986, 1 out. 2017.
- HU, S. et al. Effects and prognostic values of miR-30c-5p target genes in gastric cancer via a comprehensive analysis using bioinformatics. **Scientific Reports 2021 11:1**, v. 11, n. 1, p. 1–13, 18 out. 2021.
- HUANG, Y. et al. Rab12 Promotes Radioresistance of HPV-Positive Cervical Cancer Cells by Increasing G2/M Arrest. **Frontiers in Oncology**, v. 11, 25 fev. 2021.
- ILIEV, R. et al. Decreased expression levels of PIWIL1, PIWIL2, and PIWIL4 are associated with worse survival in renal cell carcinoma patients. **Oncotargets and therapy**, v. 9, p. 217–222, 8 jan. 2016.
- ISHIWATA, T.; MATSUDA, Y.; NAITO, Z. Nestin in gastrointestinal and other cancers: Effects on cells and tumor angiogenesis. **World Journal of Gastroenterology: WJG**, v. 17, n. 4, p. 409, 1 jan. 2011.
- ISHWARAN, H. et al. **randomForestSRC: random survival forests vignette**. , 2021. Disponível em: <<http://randomforestsrc.org/articles/survival.html>>
- ITATANI, Y. et al. **Resistance to anti-angiogenic therapy in cancer-alterations to anti-VEGF pathway**. **International Journal of Molecular Sciences** MDPI AG, , 18 abr. 2018. Disponível em: <[pmc/articles/PMC5979390/?report=abstract](https://pubmed.ncbi.nlm.nih.gov/32811111/)>. Acesso em: 28 nov. 2020
- KAMBA, T.; MCDONALD, D. M. **Mechanisms of adverse effects of anti-VEGF therapy for cancer**. **British Journal of Cancer** Nature Publishing Group, , 18 jun. 2007. Disponível em: <[pmc/articles/PMC2359962/?report=abstract](https://pubmed.ncbi.nlm.nih.gov/16811111/)>. Acesso em: 28 nov. 2020
- KANEHISA, M.; GOTO, S. KEGG: Kyoto Encyclopedia of Genes and Genomes. **Nucleic Acids Research**, v. 28, n. 1, p. 27–30, 1 jan. 2000.
- KEANE, T. M. et al. Mouse genomic variation and its effect on phenotypes and gene regulation. **Nature**, v. 477, n. 7364, p. 289–94, 14 set. 2011.
- KIERANS, S. J.; TAYLOR, C. T. Regulation of glycolysis by the hypoxia-inducible factor (HIF): implications for cellular physiology. **The Journal of physiology**, v. 599, n. 1, p. 23–37, 1 jan. 2021.
- KOMINE, Y. et al. Behavioral Abnormalities Observed in Zfhx2-Deficient Mice. **PLOS ONE**, v. 7, n. 12, p. e53114, 31 dez. 2012.
- KOZAK, M. An analysis of 5'-noncoding sequences from 699 vertebrate messenger RNAs. **Nucleic Acids Research**, v. 15, n. 20, p. 8125, 10 out. 1987.
- KRISTENSEN, L. S. et al. The biogenesis, biology and characterization of circular RNAs. **Nature Reviews Genetics**, v. 20, n. 11, p. 675–691, 2019.

- KUMAR, M. M.; GOYAL, R. LncRNA as a Therapeutic Target for Angiogenesis. **Current topics in medicinal chemistry**, v. 17, n. 15, p. 1750–1757, 2017.
- LANDSKRONER-EIGER, S.; MONEKE, I.; SESSA, W. C. miRNAs as modulators of angiogenesis. **Cold Spring Harbor perspectives in medicine**, v. 3, n. 2, p. a006643, 1 fev. 2013.
- LANGFELDER, P.; HORVATH, S. WGCNA: An R package for weighted correlation network analysis. **BMC Bioinformatics**, v. 9, n. 1, p. 559, 29 dez. 2008.
- LANGFELDER, P.; ZHANG, B.; HORVATH, S. Defining clusters from a hierarchical cluster tree: the Dynamic Tree Cut package for R. **Bioinformatics**, v. 24, n. 5, p. 719–720, 1 mar. 2008.
- LAROCHELLE, S. Protein isoforms: more than meets the eye. **Nature Methods**, v. 13, n. 4, p. 291, 2016.
- LI, B.; DEWEY, C. N. RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome. **BMC Bioinformatics**, v. 12, n. 1, p. 323, 4 dez. 2011.
- LI, C. et al. PLASMA LEVELS OF SOLUBLE CD105 CORRELATE WITH METASTASIS IN PATIENTS WITH BREAST CANCER. **Int. J. Cancer (Pred. Oncol.)**, v. 89, p. 122–126, 2000.
- LI, C.-Y.; MA, L.; YU, B. Circular RNA hsa\_circ\_0003575 regulates oxLDL induced vascular endothelial cells proliferation and angiogenesis. **Biomedicine & Pharmacotherapy**, v. 95, p. 1514–1519, 1 nov. 2017.
- LI, H. et al. Long noncoding RNA MEG3 plays a promoting role in the proliferation, invasion, and angiogenesis of lung adenocarcinoma cells through the AKT pathway. **Journal of Cellular Biochemistry**, v. 120, n. 9, p. 16143–16152, 1 set. 2019.
- LI, H.; DURBIN, R. Fast and accurate short read alignment with Burrows-Wheeler transform. **Bioinformatics**, v. 25, n. 14, p. 1754–1760, 15 jul. 2009.
- LI, H.; KO, H. P.; WHITLOCK, J. P. Induction of phosphoglycerate kinase 1 gene expression by hypoxia. Roles of Arnt and HIF1alpha. **The Journal of biological chemistry**, v. 271, n. 35, p. 21262–21267, 1996.
- LI, L. et al. lncRNAs HIF1A-AS2 facilitates the up-regulation of HIF-1 $\alpha$  by sponging to miR-153-3p, whereby promoting angiogenesis in HUVECs in hypoxia. **Biomedicine & pharmacotherapy = Biomedecine & pharmacotherapie**, v. 96, p. 165–172, 1 dez. 2017.
- LIN, Z.; LI, W. H. Evolution of 5' Untranslated Region Length and Gene Expression Reprogramming in Yeasts. **Molecular Biology and Evolution**, v. 29, n. 1, p. 81, 2012.
- LIU, C. et al. Silencing Of Circular RNA-ZNF609 Ameliorates Vascular Endothelial Dysfunction. **Theranostics**, v. 7, n. 11, p. 2863–2877, 2017a.
- LIU, W. et al. Microarray Expression Profile and Functional Analysis of Circular RNAs in Osteosarcoma. **Cellular physiology and biochemistry: international journal of experimental cellular physiology, biochemistry, and pharmacology**, v. 43, n. 3, p. 969–985, 2017b.
- LIU, Z. et al. Detection of circular RNA expression and related quantitative trait loci in the human dorsolateral prefrontal cortex. **Genome Biology**, v. 20, n. 1, p. 1–16, 20 maio 2019.
- LOIBL, S. et al. Breast cancer. **The Lancet**, v. 397, n. 10286, p. 1750–1769, 8 maio 2021.
- LOVE, M. I.; HUBER, W.; ANDERS, S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. **Genome Biology**, v. 15, n. 12, p. 1–21, 5 dez. 2014.
- MACFARLANE, L. A. et al. Regulation of Fibroblast Growth Factor-2 by an Endogenous Antisense RNA and by Argonaute-2. **Molecular Endocrinology**, v. 24, n. 4, p. 800–812, 1 abr. 2010.
- MADEIRA, F. et al. The EMBL-EBI search and sequence analysis tools APIs in 2019. **Nucleic Acids Research**, v. 47, n. W1, p. W636–W641, 2 jul. 2019.

- MATSUDA, Y.; HAGIO, M.; ISHIWATA, T. Nestin: A novel angiogenesis marker and possible target for tumor angiogenesis. **World Journal of Gastroenterology: WJG**, v. 19, n. 1, p. 42, 1 jan. 2013.
- MATSUSHITA, R. et al. Tumour-suppressive microRNA-144-5p directly targets CCNE1/2 as potential prognostic markers in bladder cancer. **British Journal of Cancer**, v. 113, n. 2, p. 282, 7 jul. 2015.
- MCCARTHY, D. J.; CHEN, Y.; SMYTH, G. K. Differential expression analysis of multifactor RNA-Seq experiments with respect to biological variation. **Nucleic Acids Research**, v. 40, n. 10, p. 4288–4297, 1 maio 2012.
- MCCREIGHT, J. C. et al. Evolution of microRNA in primates. **PLOS ONE**, v. 12, n. 6, p. e0176596, 22 jun. 2017.
- MEURER, S. K. et al. Endoglin in liver fibrogenesis: Bridging basic science and clinical practice. **World journal of biological chemistry**, v. 5, n. 2, p. 180–203, 26 maio 2014.
- MICHALIK, K. M. et al. Long Noncoding RNA MALAT1 Regulates Endothelial Cell Function and Vessel Growth. **Circulation Research**, v. 114, n. 9, p. 1389–1397, 25 abr. 2014.
- MINEO, M. et al. The Long Non-coding RNA HIF1A-AS2 Facilitates the Maintenance of Mesenchymal Glioblastoma Stem-like Cells in Hypoxic Niches. **Cell reports**, v. 15, n. 11, p. 2500–2509, 14 jun. 2016.
- MISTRY, J. et al. Pfam: The protein families database in 2021. **Nucleic Acids Research**, v. 49, n. D1, p. D412–D419, 8 jan. 2021.
- NAWROCKI, E. P.; EDDY, S. R. Infernal 1.1: 100-fold faster RNA homology searches. **Bioinformatics (Oxford, England)**, v. 29, n. 22, p. 2933–5, 15 nov. 2013.
- NCBI. Disponível em: <<https://www.ncbi.nlm.nih.gov/gene/55149>>. Acesso em: 1 ago. 2022.
- NICHOL, D.; STUHLMANN, H. EGFL7: a unique angiogenic signaling factor in vascular development and disease. **Blood**, v. 119, n. 6, p. 1345, 2 fev. 2012.
- NIU, X. et al. Weighted Gene Co-Expression Network Analysis Identifies Critical Genes in the Development of Heart Failure After Acute Myocardial Infarction. **Frontiers in Genetics**, v. 10, p. 1214, 26 nov. 2019.
- NOWAK, A. et al. Role of nestin expression in angiogenesis and breast cancer progression. **International Journal of Oncology**, v. 52, n. 2, p. 527–535, 1 fev. 2018.
- O'BRIEN, J. et al. Overview of MicroRNA Biogenesis, Mechanisms of Actions, and Circulation. **Frontiers in endocrinology**, v. 9, p. 402, 2018.
- PAVESI, G. et al. Exalign: A new method for comparative analysis of exon-intron gene structures. **Nucleic Acids Research**, v. 36, n. 8, p. e47, maio 2008.
- PEACH, C. J. et al. **Molecular pharmacology of VEGF-A isoforms: Binding and signalling at VEGFR2**. **International Journal of Molecular Sciences** MDPI AG, , 23 abr. 2018. Disponível em: <[/pmc/articles/PMC5979509/?report=abstract](https://pubmed.ncbi.nlm.nih.gov/30000000/)>. Acesso em: 29 nov. 2020
- PEREIRA, B. et al. The somatic mutation profiles of 2,433 breast cancers refines their genomic and transcriptomic landscapes. **Nature communications**, v. 7, 10 maio 2016.
- PÉREZ-GÓMEZ, E. et al. Characterization of murine S-endoglin isoform and its effects on tumor development. **Oncogene 2005 24:27**, v. 24, n. 27, p. 4450–4461, 4 abr. 2005.
- PERTEA, M. et al. StringTie enables improved reconstruction of a transcriptome from RNA-seq reads. **Nature Biotechnology**, v. 33, n. 3, p. 290–295, 18 mar. 2015.
- POTENTE, M.; GERHARDT, H.; CARMELIET, P. Basic and therapeutic aspects of angiogenesis. **Cell**, v. 146, n. 6, p. 873–87, 16 set. 2011.
- QIAN, Y. et al. Prognostic Cancer Gene Expression Signatures: Current Status and Challenges. **Cells**, v. 10, n. 3, p. 1–17, 1 mar. 2021.



- RICARD-BLUM, S. The Collagen Family. **Cold Spring Harbor Perspectives in Biology**, v. 3, n. 1, p. 1–19, jan. 2011.
- ROBEY, I. F. et al. Hypoxia-Inducible Factor-1 $\alpha$  and the Glycolytic Phenotype in Tumors. **Neoplasia (New York, N.Y.)**, v. 7, n. 4, p. 324, 2005.
- ROBINSON, M. D.; MCCARTHY, D. J.; SMYTH, G. K. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. **Bioinformatics**, v. 26, n. 1, p. 139–140, 1 jan. 2010.
- RODRIGUEZ, J. M. et al. APPRIS 2017: Principal isoforms for multiple gene sets. **Nucleic Acids Research**, v. 46, n. D1, p. D213–D217, 1 jan. 2018.
- RUAN, W. et al. Knockdown of long noncoding RNA MEG3 impairs VEGF-stimulated endothelial sprouting angiogenesis via modulating VEGFR2 expression in human umbilical vein endothelial cells. **Gene**, v. 649, p. 32–39, 5 abr. 2018.
- SAKAI, E. et al. A mammalian mirtron miR-1224 promotes tube-formation of human primary endothelial cells by targeting anti-angiogenic factor epsin2. **Scientific Reports 2017 7:1**, v. 7, n. 1, p. 1–12, 17 jul. 2017.
- SARMA, V. et al. Cloning of a novel tumor necrosis factor-alpha-inducible primary response gene that is differentially expressed in development and capillary tube-like formation in vitro. **Journal of immunology (Baltimore, Md. : 1950)**, v. 148, n. 10, p. 3302–12, 15 maio 1992.
- SEAMAN, S. et al. Genes that Distinguish Physiological and Pathological Angiogenesis. **Cancer cell**, v. 11, n. 6, p. 539, 12 jun. 2007.
- SHAN, K. et al. Circular Noncoding RNA HIPK3 Mediates Retinal Vascular Dysfunction in Diabetes Mellitus. **Circulation**, v. 136, n. 17, p. 1629–1642, 24 out. 2017.
- SHANNON, P. et al. Cytoscape: A Software Environment for Integrated Models of Biomolecular Interaction Networks. **Genome Research**, v. 13, n. 11, p. 2498–2504, 1 nov. 2003.
- SHAO, M.; KINGSFORD, C. Accurate assembly of transcripts through phase-preserving graph decomposition. **Nature biotechnology**, v. 35, n. 12, p. 1167, 2017.
- SHEN, T. et al. LncRNA Meg3 knockdown reduces corneal neovascularization and VEGF-induced vascular endothelial angiogenesis via SDF-1/CXCR4 and Smad2/3 pathway. **Experimental Eye Research**, v. 222, p. 109166, 1 set. 2022.
- SHERMAN, B. T. et al. DAVID: a web server for functional enrichment analysis and functional annotation of gene lists (2021 update). **Nucleic acids research**, v. 50, n. W1, 5 jul. 2022.
- SIER, V. Q. et al. Endoglin/CD105-Based Imaging of Cancer and Cardiovascular Diseases: A Systematic Review. **International journal of molecular sciences**, v. 22, n. 9, 1 maio 2021.
- SIMON, M. M. et al. A comparative phenotypic and genomic analysis of C57BL/6J and C57BL/6N mouse strains. **Genome biology**, v. 14, n. 7, p. R82, 31 jul. 2013.
- STAPOR, P. et al. Angiogenesis revisited - role and therapeutic potential of targeting endothelial metabolism. **Journal of cell science**, v. 127, n. Pt 20, p. 4331–4341, 2014.
- SU, C. et al. PIWIL4 regulates cervical cancer cell line growth and is involved in down-regulating the expression of p14ARF and p53. **FEBS letters**, v. 586, n. 9, p. 1356–1362, 7 maio 2012.
- SUÁREZ, Y.; SESSA, W. C. MicroRNAs as novel regulators of angiogenesis. **Circulation research**, v. 104, n. 4, p. 442–54, 27 fev. 2009.
- SUNG, H. et al. Global Cancer Statistics 2020: GLOBOCAN Estimates of Incidence and Mortality Worldwide for 36 Cancers in 185 Countries. **CA: A Cancer Journal for Clinicians**, v. 71, n. 3, p. 209–249, 1 maio 2021.
- SZKLARCZYK, D. et al. The STRING database in 2021: customizable protein-protein networks, and functional characterization of user-uploaded gene/measurement sets. **Nucleic acids research**, v. 49, n. D1, p. D605–D612, 8 jan. 2021.

- TALLURI, B. et al. COL2A1 Is a Novel Biomarker of Melanoma Tumor Repopulating Cells. **Biomedicines**, v. 8, n. 9, 1 set. 2020.
- TIWARI, A.; MUKHERJEE, B.; DIXIT, M. MicroRNA Key to Angiogenesis Regulation: MiRNA Biology and Therapy. **Current Cancer Drug Targets**, v. 18, n. 3, p. 266–277, 19 fev. 2018.
- TRIMARCHI, T. et al. Genome-wide mapping and characterization of Notch-regulated long noncoding RNAs in acute leukemia. **Cell**, v. 158, n. 3, p. 593–606, 31 jul. 2014.
- VAN DE VIJVER, M. J. et al. A gene-expression signature as a predictor of survival in breast cancer. **The New England journal of medicine**, v. 347, n. 25, p. 1999–2009, 19 dez. 2002.
- VAREY, A. H. R. et al. VEGF165b, an antiangiogenic VEGF-A isoform, binds and inhibits bevacizumab treatment in experimental colorectal carcinoma: Balance of pro- and antiangiogenic VEGF-A isoforms has implications for therapy. **British Journal of Cancer**, v. 98, n. 8, p. 1366–1379, 22 abr. 2008.
- VIEIRA, A. S. et al. Role of non-coding RNAs in non-aging-related neurological disorders. **Brazilian Journal of Medical and Biological Research**, v. 51, n. 8, 11 jun. 2018.
- WAKS, A. G.; WINER, E. P. Breast Cancer Treatment: A Review. **JAMA**, v. 321, n. 3, p. 288–300, 22 jan. 2019.
- WANG, L. et al. CPAT: Coding-Potential Assessment Tool using an alignment-free logistic regression model. **Nucleic acids research**, v. 41, n. 6, p. e74, 1 abr. 2013a.
- WANG, R. et al. MicroRNA-195 Suppresses Angiogenesis and Metastasis of Hepatocellular Carcinoma by Inhibiting the Expression of VEGF, VAV2, and CDC42. **Hepatology**, v. 58, n. 2, p. 642–653, ago. 2013b.
- WANG, W. et al. Evolutionary and functional implications of 3' untranslated region length of mRNAs by comprehensive investigation among four taxonomically diverse metazoan species. **Genes and Genomics**, v. 41, n. 7, p. 747–755, 21 mar. 2019.
- WANG, Y. et al. The PIWI protein acts as a predictive marker for human gastric cancer. **International Journal of Clinical and Experimental Pathology**, v. 5, n. 4, p. 315, 2012.
- WANG, Z. et al. The Role of PIWIL4, an Argonaute Family Protein, in Breast Cancer. **The Journal of Biological Chemistry**, v. 291, n. 20, p. 10646, 5 maio 2016.
- WILUSZ, J. E. Long noncoding RNAs: Re-writing dogmas of RNA processing and stability. **Biochimica et biophysica acta**, v. 1859, n. 1, p. 128, 1 jan. 2016.
- WU, D. et al. lncRNA HIF1A Antisense RNA 2 Modulates Trophoblast Cell Invasion and Proliferation through Upregulating PHLDA1 Expression. **Molecular therapy. Nucleic acids**, v. 16, p. 605–615, 7 jun. 2019a.
- WU, H. et al. Upregulated miR-20a-5p expression promotes proliferation and invasion of head and neck squamous cell carcinoma cells by targeting of TNFRSF21. **Oncology reports**, v. 40, n. 2, p. 1138–1146, 1 ago. 2018.
- WU, S. et al. Whole-genome sequencing identifies ADGRG6 enhancer mutations and FRS2 duplications as angiogenesis-related drivers in bladder cancer. **Nature Communications 2019 10:1**, v. 10, n. 1, p. 1–12, 12 fev. 2019b.
- XIE, Y.; WANG, B. Downregulation of TNFAIP2 suppresses proliferation and metastasis in esophageal squamous cell carcinoma through activation of the Wnt/ $\beta$ -catenin signaling pathway. **Oncology Reports**, v. 37, n. 5, p. 2920–2928, 1 maio 2017.
- XU, F. et al. LncRNA HIF1A-AS1 promotes gemcitabine resistance of pancreatic cancer by enhancing glycolysis through modulating the AKT/YB1/HIF1a pathway. **Cancer Research**, v. 81, n. 22, p. 5678–5691, 15 nov. 2021.
- XU, S. et al. The role of collagen in cancer: from bench to bedside. **Journal of Translational Medicine 2019 17:1**, v. 17, n. 1, p. 1–22, 14 set. 2019.
- YAMADA, Y. et al. Regulation of antitumor miR-144-5p targets oncogenes: Direct regulation of syndecan-3 and its clinical significance. **Cancer Science**, v. 109, n. 9, p. 2919, 1 set. 2018.

- YAMAUCHI, M. et al. The fibrotic tumor stroma. **The Journal of clinical investigation**, v. 128, n. 1, p. 16–25, 2 jan. 2018.
- YANG, H.; CHEN, Z.; LIU, Z. MiR-20a-5p Negatively Regulates NR4A3 to Promote Metastasis in Bladder Cancer. **Journal of oncology**, v. 2021, 2021.
- YETKIN-ARIK, B. et al. The role of glycolysis and mitochondrial respiration in the formation and functioning of endothelial tip cells during angiogenesis. **Scientific Reports 2019 9:1**, v. 9, n. 1, p. 1–14, 30 ago. 2019.
- YIP, A. M.; HORVATH, S. Gene network interconnectedness and the generalized topological overlap measure. **BMC Bioinformatics**, v. 8, n. 1, p. 1–14, 24 jan. 2007.
- YU, Y. et al. MiR-20a-5p suppresses tumor proliferation by targeting autophagy-related gene 7 in neuroblastoma. **Cancer Cell International**, v. 18, n. 1, p. 1–10, 4 jan. 2018.
- ZAMBELLI, F. et al. Assessment of orthologous splicing isoforms in human and mouse orthologous genes. **BMC Genomics**, v. 11, n. 1, 1 out. 2010.
- ZHANG, K. et al. Long non-coding RNA HIF1A-AS2 modulates the proliferation, migration, and phenotypic switch of aortic smooth muscle cells in aortic dissection via sponging microRNA-33b. **Bioengineered**, v. 13, n. 3, p. 6383, 2022.
- ZHANG, S. et al. miR-30e-5p represses angiogenesis and metastasis by directly targeting AEG-1 in squamous cell carcinoma of the head and neck. **Cancer Science**, v. 111, n. 2, p. 356, 1 fev. 2020a.
- ZHANG, W. et al. Long Noncoding RNA HIF1A-AS2 Promotes Non-Small Cell Lung Cancer Progression by the miR-153-5p/S100A14 Axis. **OncoTargets and therapy**, v. 13, p. 8715–8722, 2020b.
- ZHANG, Y. et al. Activation of PGK1 under hypoxic conditions promotes glycolysis and increases stem cell-like properties and the epithelial-mesenchymal transition in oral squamous cell carcinoma cells via the AKT signalling pathway. **International journal of oncology**, v. 57, n. 3, p. 743–755, 1 set. 2020c.
- ZHANG, Y.; YANG, L.; CHEN, L. L. Life without A tail: new formats of long noncoding RNAs. **The international journal of biochemistry & cell biology**, v. 54, p. 338–349, 2014.
- ZHAO, Y.; SINGH, R. P. **The role of anti-vascular endothelial growth factor (anti-VEGF) in the management of proliferative diabetic retinopathy**. **Drugs in Context**Bioexcel Publishing LTD, , 13 ago. 2018. Disponível em: </pmc/articles/PMC6113746/?report=abstract>. Acesso em: 28 nov. 2020
- ZHENG, F. et al. The HIF-1 $\alpha$  antisense long non-coding RNA drives a positive feedback loop of HIF-1 $\alpha$  mediated transactivation and glycolysis. **Nature Communications 2021 12:1**, v. 12, n. 1, p. 1–20, 26 fev. 2021.
- ZHOU, H. et al. Oxygen-induced circRNA profiles and coregulatory networks in a retinopathy of prematurity mouse model. **Experimental and Therapeutic Medicine**, 26 jul. 2019.
- ZHU, W. et al. The roles of endoglin gene in cerebrovascular diseases. **Neuroimmunology and neuroinflammation**, v. 4, n. 10, p. 199, 17 out. 2017.
- ZHU, Y. et al. lncRNAHIF1A-AS2 Promotes Renal Carcinoma Cell Proliferation and Migration via miR-130a-5p/ERBB2 Pathway. **OncoTargets and therapy**, v. 13, p. 9807–9820, 2020.

## Apêndice I: Artigos publicados ao longo do doutorado

Aqui estão listados os artigos que participei como co-autor ao longo do meu doutorado, mas que não se relacionam com a presente Tese. O primeiro artigo da lista se refere a tese de doutorado do Rodrigo Guarischi-Souza, cujo trabalho antecedeu o estudo apresentado nesta tese. Os artigos 2, 3, e 4 são frutos de colaborações do orientador e co-orientador do meu doutorado. E os artigos 5 e 6 são frutos de colaborações do meu co-orientador de mestrado. O artigo referente a esta Tese se encontra em preparo.

- 1) Guarischi-Souza, R.; Monteiro, J.S.; Alecrim, L.C.; Michaloski, J.S.; Cardeal, L.B.; Ferreira, E.N.; Carraro, D.M.; Nunes, D.N.; Dias-Neto, E.; Boutros, P.C.; Setubal, J.C.; Giordano, R.J. (2019). A transcriptome-based signature of pathological angiogenesis predicts breast cancer patient survival. *PLoS Genetics*. <http://dx.doi.org/10.1371/journal.pgen.1008482>
- 2) de Pedro-Jové, R.; Puigvert, M.; Sebastià, P.; Macho, A.P.; Monteiro, J.S.; Coll, N.S.; Setubal, J.C.; Valls, M. (2021). Dynamic expression of *Ralstonia solanacearum* virulence factors and metabolism-controlling genes during plant infection. *BMC Genomics*. 22(1):170. doi: 10.1186/s12864-021-07457-w.
- 3) Fabro, A.T.; Machado-Rugolo, J.; Baldavira, C.M.; Prieto, T.G.; Farhat, C.; ManGone, F.R.R.; Batah, S.S.; Cruvinel, H.R.; Aldá, M.A.; Monteiro, J.S.; Pádua, A.I.; Morais, S.S.; de Oliveira, R.A.; Santos, M.K; Baddini-Martinez, J.A.; Setubal, J.C.; Rainho, C.A.; Yoo, H.H.B.; Silva, P.L.; Nagai, M.A.; Capelozzi, V.L. (2021). Circulating Plasma miRNA and Clinical/Hemodynamic Characteristics Provide Additional Predictive Information About Acute Pulmonary Thromboembolism, Chronic Thromboembolic Pulmonary Hypertension and Idiopathic Pulmonary Hypertension. *Front Pharmacol*. 2021;12. doi:10.3389/fphar.2021.648769
- 4) Reis Teixeira, A.A.; Carnero, L.R.; Kuramoto, A.; Fen Tang, F.H.; Gomes, C.H.; Pereira, N.B.; Campos de Oliveira, L.; Garrini, R.; Monteiro, J.S.; Setubal, J.C.; Sabino, E.C.; Pasqualini, R.; Colli, W.; Arap, W.; Manso Alves, M.J.; Cunha-Neto, E.; Giordano, R.J.; A refined genome phage display methodology delineates the human antibody response in Chagas disease patients. *iScience* (2021), doi: <https://doi.org/10.1016/j.isci.2021.102540>.
- 5) Lima, D.; Mattos, J.J.; Piazza, R.S.; Righetti, B.P.H.; Monteiro, J.S.; Grott, S.C.; Alves, T.C.; Taniguchi, S.; Bicego, M.C.; Almeida, E.A.; Bebianno, M.J.; Medeiros, I.D.; Bairy, A.C.D. (2019). Stress responses in *Crassostrea gasar* exposed to combined effects of acute pH changes and phenanthrene. *Science of the total environment*. <http://dx.doi.org/10.1016/j.scitotenv.2019.04.450>
- 6) Guerreiro, A.S.; Monteiro, J.S.; Medeiros, I.D.; Sandrini, J.Z. (2020). First evidence of transcriptional modulation by chlorothalonil in mussels *Perna perna*. *Chemosphere*. <http://dx.doi.org/10.1016/j.chemosphere.2020.126947>

## Apêndice II: Lista de variantes diferencialmente expressas (DE)

Gene ID	Símbolo	Nome	Transcrito ID	Codificador de proteína?	LFC	DIA DE
ENSMUSG00000039116	Adgrg6	Adhesion G protein-coupled receptor G6	ENSMUST00000041168	Sim	1.15	12_5
ENSMUSG00000039167	Adgrl4	adhesion G protein-coupled receptor L4	ENSMUST00000046977	Sim	-1.20	12_5
ENSMUSG00000030790	Adm	Adrenomedullin	ENSMUST00000033054	Sim	1.60	12_5
ENSMUSG00000030790	Adm	Adrenomedullin	ENSMUST00000033054	Sim	02.03	17
ENSMUSG00000030790	Adm	Adrenomedullin	ENSMUST00000033054	Sim	2.49	15
ENSMUSG00000053175	Bcl3	B cell leukemia/lymphoma 3	ENSMUST00000120537	Sim	3.22	17
ENSMUSG00000053175	Bcl3	B cell leukemia/lymphoma 3	ENSMUST00000135609	Sim	2.37	17
ENSMUSG00000078566	Bnip3	BCL2/adenovirus E1B interacting protein 3	ENSMUST00000106112	Sim	1.46	12_5
ENSMUSG00000035385	Ccl2	chemokine (C-C motif) ligand 2	ENSMUST00000000193	Sim	04.06	17
ENSMUSG00000035385	Ccl2	chemokine (C-C motif) ligand 2	ENSMUST00000124479	Não (lncRNA)	6.99	17
ENSMUSG00000031871	Cdh5	cadherin 5	ENSMUST00000034339	Sim	-1.44	12_5
ENSMUSG00000071637	Cebpd	CCAAT/enhancer binding protein (C/EBP), delta	ENSMUST00000096232	Sim	1.29	12_5
ENSMUSG00000071637	Cebpd	CCAAT/enhancer binding protein (C/EBP), delta	ENSMUST00000096232	Sim	1.31	17
ENSMUSG00000022483	Col2a1	collagen, type II, alpha 1	ENSMUST00000088355	Sim	-1.02	17
ENSMUSG00000045382	Cxcr4	chemokine (C-X-C motif) receptor 4	ENSMUST00000052172	Sim	-1.29	12_5
ENSMUSG00000031616	Ednra	endothelin receptor type A / B	ENSMUST00000034029	Sim	-2.05	12_5
ENSMUSG00000037868	Egr2	early growth response 2	ENSMUST00000048289	Sim	10.69	17
ENSMUSG00000026814	Eng	endoglin	ENSMUST00000113272	Sim	-1.19	12_5
ENSMUSG00000032035	Ets1	E26 avian leukemia oncogene 1, 5' domain	ENSMUST00000034534	Sim	-1.17	12_5
ENSMUSG00000089984	Fbxo24	F-box protein 24	ENSMUST00000136028	NMD	4.20	17
ENSMUSG00000026193	Fn1	fibronectin 1	ENSMUST00000055226	Sim	1.98	17
ENSMUSG00000026193	Fn1	fibronectin 1	ENSMUST00000189821	Sim	-9.82	12_5
ENSMUSG00000026193	Fn1	fibronectin 1	ENSMUST00000190780	Sim	11.86	12_5
ENSMUSG00000021250	Fos	FBJ osteosarcoma oncogene	ENSMUST00000021674	Sim	1.75	12_5
ENSMUSG00000021250	Fos	FBJ osteosarcoma oncogene	ENSMUST00000136122	Não (lncRNA)	02.03	12_5
ENSMUSG00000030406	Gipr	gastric inhibitory polypeptide receptor	ENSMUST00000094790	Sim	3.68	12_5
ENSMUSG00000030406	Gipr	gastric inhibitory polypeptide receptor	ENSMUST00000206857	Não (lncRNA)	1.89	15
ENSMUSG00000030406	Gipr	gastric inhibitory polypeptide receptor	ENSMUST00000206857	Não (lncRNA)	1.91	17
ENSMUSG00000030406	Gipr	gastric inhibitory polypeptide receptor	ENSMUST00000206857	Não (lncRNA)	04.01	12_5
ENSMUSG00000030406	Gipr	gastric inhibitory polypeptide receptor	ENSMUST00000206971	NMD	4.18	12_5
ENSMUSG00000042284	Itga1	integrin alpha 1	ENSMUST00000061673	Sim	-1.14	12_5
ENSMUSG00000041782	Lad1	ladinin 1	ENSMUST00000038760	Sim	2.57	12_5
ENSMUSG00000041782	Lad1	ladinin 1	ENSMUST00000038760	Sim	3.34	17
ENSMUSG00000019846	Lama4	laminin, alpha 4	ENSMUST00000019992	Sim	-1.00	12_5
ENSMUSG00000021998	Lcp1	lymphocyte cytosolic protein 1	ENSMUST00000124608	Não (lncRNA)	-3.25	12_5
ENSMUSG00000042622	Maff	v-maf musculoaponeurotic fibrosarcoma oncogene family, protein F (avian)	ENSMUST00000096350	Sim	2.96	12_5
ENSMUSG00000031004	Mki67	antigen identified by monoclonal antibody Ki 67	ENSMUST00000033310	Sim	02.03	17

ENSMUSG00000025515	Muc2	mucin 2	ENSMUST00000187945	Sim	02.04	12_5
ENSMUSG00000040280	Ndufa4l2	Ndufa4, mitochondrial complex associated like 2	ENSMUST00000035735	Sim	1.58	12_5
ENSMUSG00000040280	Ndufa4l2	Ndufa4, mitochondrial complex associated like 2	ENSMUST00000035735	Sim	2.18	15
ENSMUSG00000040280	Ndufa4l2	Ndufa4, mitochondrial complex associated like 2	ENSMUST00000035735	Sim	2.25	17
ENSMUSG00000040280	Ndufa4l2	Ndufa4, mitochondrial complex associated like 2	ENSMUST00000218302	Sim	3.45	12_5
ENSMUSG00000040280	Ndufa4l2	Ndufa4, mitochondrial complex associated like 2	ENSMUST00000218492	Sim	1.74	12_5
ENSMUSG00000004891	Nes	nestin	ENSMUST00000160694	Sim	1.33	12_5
ENSMUSG00000022146	Osmr	oncostatin M receptor	ENSMUST00000022746	Sim	1.10	12_5
ENSMUSG00000024440	Pcdh12	protocadherin 12	ENSMUST00000025311	Sim	-1.56	12_5
ENSMUSG00000024440	Pcdh12	protocadherin 12	ENSMUST00000195747	Não (lncRNA)	-1.52	12_5
ENSMUSG00000040204	Pclaf	PCNA clamp associated factor	ENSMUST00000045802	Sim	2.99	17
ENSMUSG00000029718	Pcolce	procollagen C-endopeptidase enhancer protein	ENSMUST00000128755	Não (lncRNA)	2.20	17
ENSMUSG00000041482	Piezo2	piezo-type mechanosensitive ion channel component 2	ENSMUST00000047480	Sim	1.88	17
ENSMUSG00000036912	Piwil4	piwi-like RNA-mediated gene silencing 4	ENSMUST00000115644	Sim	2.80	17
ENSMUSG00000036912	Piwil4	piwi-like RNA-mediated gene silencing 4	ENSMUST00000115644	Sim	4.22	12_5
ENSMUSG00000036912	Piwil4	piwi-like RNA-mediated gene silencing 4	ENSMUST00000136399	Sim	2.77	12_5
ENSMUSG00000028036	Ptgfr	prostaglandin F receptor	ENSMUST00000029670	Sim	1.73	17
ENSMUSG00000028036	Ptgfr	prostaglandin F receptor	ENSMUST00000029670	Sim	2.78	12_5
ENSMUSG00000026678	Rgs5	regulator of G-protein signaling 5	ENSMUST00000027997	Sim	-1.59	12_5
ENSMUSG00000045667	Smtnl2	smoothelin-like 2	ENSMUST00000050226	Sim	2.00	12_5
ENSMUSG00000053113	Socs3	suppressor of cytokine signaling 3	ENSMUST00000054002	Sim	1.49	17
ENSMUSG00000020303	Stc2	stanniocalcin 2	ENSMUST00000020546	Sim	1.63	12_5
ENSMUSG00000037820	Tgm2	transglutaminase 2, C polypeptide	ENSMUST00000103122	Sim	1.32	12_5
ENSMUSG00000001131	Timp1	tissue inhibitor of metalloproteinase 1	ENSMUST00000009530	Sim	2.11	17
ENSMUSG00000021281	Tnfaip2	tumor necrosis factor, alpha-induced protein 2	ENSMUST00000102745	Sim	02.09	17
ENSMUSG00000023905	Tnfrsf12a	tumor necrosis factor receptor superfamily, member 12a	ENSMUST00000024698	Sim	2.38	12_5
ENSMUSG00000026414	Tnnt2	troponin T2, cardiac	ENSMUST00000186225	Não (lncRNA)	2.84	17
ENSMUSG00000026414	Tnnt2	troponin T2, cardiac	ENSMUST00000191055	Não (retained intron)	4.30	17
ENSMUSG00000020914	Top2a	topoisomerase (DNA) II alpha	ENSMUST00000068031	Sim	2.14	17
ENSMUSG00000032554	Trf	Transferrin	ENSMUST00000035158	Sim	-1.39	15
ENSMUSG00000023951	Vegfa	vascular endothelial growth factor A	ENSMUST00000142321	Não (lncRNA)	-1.37	17
ENSMUSG00000001930	Vwf	Von Willebrand factor	ENSMUST00000112254	Sim	-1.51	12_5
ENSMUSG00000001930	Vwf	Von Willebrand factor	ENSMUST00000147101	NMD	-1.62	12_5

## Apêndice III: Lista de candidatos homólogos para as variantes diferencialmente expressas

Variante camundongo	Símbolo	Variante humana	BLASTN	BLASTP	BLANQUART et al., 2016	ZABELLI et al., 2010	Exalign	Categoria evidência
ENSMUST00000055226	Fn1	ENST00000354785.10	x	x	x	x	x	Forte
ENSMUST00000120537	Bcl3	ENST00000164227.10		x	x	x	x	Moderado
ENSMUST00000088355	Col2a1	ENST00000337299.7	x	x	x		x	Forte
ENSMUST00000189821	Fn1	ENST00000357867.8	x	x		x	x	Forte
ENSMUST00000061673	Itga1	ENST00000282588.7		x	x	x	x	Moderado
ENSMUST00000019992	Lama4	ENST00000424408.6	x	x	x		x	Forte
ENSMUST00000103122	Tgm2	ENST00000361475.7		x	x	x	x	Moderado
ENSMUST00000112254	Vwf	ENST00000261405.9	x	x	x		x	Forte
ENSMUST00000041168	Adgrg6	ENST00000367608		x	x		x	Moderado
ENSMUST00000034029	Ednra	ENST00000651419		x		x	x	Moderado
ENSMUST00000048289	Egr2	ENST00000242480	x	x		x		Forte
ENSMUST00000034534	Ets1	ENST00000330871.3		x	x		x	Moderado
ENSMUST00000021674	Fos	ENST00000303562.9		x	x	x		Moderado
ENSMUST00000094790	Gipr	ENST00000590918.6		x	x		x	Moderado
ENSMUST00000035735	Ndufa4l2	ENST00000554503		x	x	x		Moderado
ENSMUST00000025311	Pcdh12	ENST00000231484.4		x	x	x		Moderado
ENSMUST00000045802	Pclaf	ENST00000300035		x	x	x		Moderado
ENSMUST00000047480	Piezo2	ENST00000503781.7		x	x		x	Moderado
ENSMUST00000027997	Rgs5	ENST00000313961.10		x	x	x		Moderado
ENSMUST00000020546	Stc2	ENST00000265087.8		x	x	x		Moderado
ENSMUST00000068031	Top2a	ENST00000423485.6		x	x		x	Moderado
ENSMUST00000034339	Cdh5	NM_001795 / CCDS10804			x	x		Moderado
ENSMUST00000038760	Lad1	NM_005558 / CCDS1410			x	x		Moderado
ENSMUST00000050226	Smtnl2	ENST00000389313.9		x			x	Moderado
ENSMUST00000102745	Tnfaip2	NM_006291 / CCDS9979			x	x		Moderado
ENSMUST00000024698	Tnfrsf12a	NM_016639 / CCDS10489			x	x		Moderado
ENSMUST00000046977	Adgrl4	NM_022159			x			Fraca
ENSMUST00000033054	Adm	CCDS7801				x		Fraca
ENSMUST00000106112	Bnip3	NM_004052			x			Fraca
ENSMUST00000096232	Cebpd	ENST00000408965.4		x				Fraca
ENSMUST00000113272	Eng	NM_000118			x			Fraca
ENSMUST00000033310	Mki67	NM_002417			x			Fraca
ENSMUST00000022746	Osmr	NM_003999			x			Fraca
ENSMUST00000054002	Socs3	ENST00000330871.3		x				Fraca
ENSMUST00000009530	Timp1	CCDS14281				x		Fraca
ENSMUST00000035158	Trf	NM_001063			x			Fraca
ENSMUST00000142321	Vegfa	NM_003376			x			Fraca

## Apêndice IV: Lista de miRNAs diferencialmente expressos (DE)

miRDeep2_ID	miRBase_ID	LFC	FDR	Dia_DE
X_22195	mmu-miR-676-5p	1.248916	0.00621	12
X_22300	mmu-miR-98-5p	1.489884	0.005511	17
X_22300	mmu-miR-98-5p	-1.53636	0.00082	12
X_22369	mmu-miR-532-5p	1.090021	0.018067	12
X_22466	mmu-miR-92a-2-5p	-1.49183	0.000756	17
X_22466	mmu-miR-92a-2-5p	1.493899	0.000596	12
X_22481	mmu-miR-351-5p	-2.23091	0.000513	12
X_22481	mmu-miR-351-5p	4.74876	6.77E-06	17
X_22485	mmu-miR-322-5p	-1.84439	0.025552	12.5
X_22485	mmu-miR-322-5p	-2.85873	0.00629	12
X_22660	-	1.739138	0.000181	12
1_1018	mmu-miR-375-5p	-1.35726	0.011699	17
1_1463	mmu-miR-205-5p	1.290072	0.025552	12.5
1_1463	mmu-miR-205-5p	-8.57009	7.33E-05	12
1_224	mmu-miR-26b-5p	-1.71368	5.84E-05	12
1_224	mmu-miR-26b-5p	1.978905	5.67E-06	17
1_37	mmu-miR-206-5p	1.422562	0.001787	12.5
1_37	mmu-miR-206-5p	-2.83151	2.19E-08	12
1_37	mmu-miR-206-5p	2.956154	3.24E-06	17
1_436	mmu-miR-135b-5p	-6.24507	0.049774	12.5
1_45	mmu-miR-30c-5p	2.073231	3.73E-06	12
1_45	mmu-miR-30c-5p	-2.22868	2.57E-07	17
1_492	mmu-miR-181b-5p	1.234713	0.003754	12
1_595	mmu-miR-199a-5p	-2.22406	0.03011	12
2_11130	mmu-miR-669c-5p	1.301414	0.017322	12.5
2_11198	mmu-miR-126a-5p	-1.62477	0.000826	12.5
2_11198	mmu-miR-126a-5p	-1.72966	0.000581	12
2_11253	mmu-miR-199b-5p	-2.22406	0.03011	12
2_11320	mmu-miR-181b-5p	1.230511	0.003754	12
2_11424	mmu-miR-10b-5p	-1.12394	0.016482	12
2_11424	mmu-miR-10b-5p	1.955968	7.43E-06	12.5
2_11987	mmu-miR-1a-1-5p	-2.28796	0.000254	12
2_11987	mmu-miR-1a-1-5p	3.343804	7.66E-05	17
2_12128	mmu-miR-219a-5p	-5.43711	2.17E-07	15
2_12383	mmu-miR-129-5p	1.943415	4.65E-06	12
2_12383	mmu-miR-129-5p	1.990336	6.11E-06	12.5
2_12383	mmu-miR-129-5p	-2.9403	1.64E-12	17
2_12724	mmu-miR-296-5p	-1.29076	0.005185	17
2_12724	mmu-miR-296-5p	1.743718	6.69E-05	12



3_13062	mmu-miR-9-5p	1.6029	0.000127	17
3_13479	-	-2.2977	0.012888	17
4_14783	mmu-miR-873a-5p	-1.38896	0.011699	17
4_15000	mmu-miR-101a-5p	1.438645	0.000736	12
4_15110	mmu-miR-30e-5p	1.577853	0.000369	12.5
4_15110	mmu-miR-30e-5p	2.746851	4.63E-11	12
4_15110	mmu-miR-30e-5p	-2.79393	1.32E-11	17
5_15432	mmu-miR-879-5p	-2.11903	0.048147	12
5_15432	mmu-miR-879-5p	-2.51807	0.004532	12.5
5_16372	-	-1.56318	0.044729	17
5_16372	-	2.314681	0.000925	12
5_16651	mmu-miR-93-5p	-1.84397	0.000192	12.5
5_16651	mmu-miR-93-5p	-2.14814	3.84E-05	12
5_16654	mmu-miR-106b-5p	1.305057	0.003754	12
5_16654	mmu-miR-106b-5p	1.307463	0.005465	12.5
6_16808	mmu-miR-129-5p	1.943415	4.65E-06	12
6_16808	mmu-miR-129-5p	1.990875	6.11E-06	12.5
6_16808	mmu-miR-129-5p	-2.93934	1.64E-12	17
6_17413	mmu-miR-182-5p	1.589173	0.000343	12.5
6_17413	mmu-miR-182-5p	2.209169	1.35E-07	12
6_17413	mmu-miR-182-5p	-2.3806	1.02E-08	17
6_17417	mmu-miR-183-5p	1.289695	0.004532	12.5
6_17417	mmu-miR-183-5p	-1.74463	2.93E-05	17
6_17417	mmu-miR-183-5p	2.201891	1.35E-07	12
6_17516	mmu-miR-148a-5p	1.515399	0.000355	12
7_18320	mmu-miR-7a-5p	-1.14936	0.017797	12
7_18320	mmu-miR-7a-5p	1.732823	0.00295	17
7_18330	mmu-miR-9-5p	1.6029	0.000127	17
7_18350	mmu-miR-1839-5p	1.036238	0.022404	12
7_18417	mmu-miR-708-5p	1.10164	0.038296	12.5
7_18417	mmu-miR-708-5p	1.971152	2.11E-05	12
7_18417	mmu-miR-708-5p	-2.35828	1.25E-07	17
7_19414	mmu-miR-210-5p	-1.11221	0.020618	12
7_19414	mmu-miR-210-5p	1.866344	4.70E-05	17
8_19785	mmu-miR-24-2-5p	-1.42907	0.003754	12
8_19785	mmu-miR-24-2-5p	-1.43549	0.003312	12.5
8_19786	mmu-miR-24-3p	-1.39419	0.00444	12.5
8_19786	mmu-miR-24-3p	-1.42814	0.003396	12
8_20105	mmu-miR-486a-5p	0.975963	0.045591	17
8_20312	mmu-miR-7068-5p	1.512166	0.048147	12
8_20349	mmu-miR-181d-5p	1.283868	0.004606	12.5
8_20349	mmu-miR-181d-5p	-1.33706	0.001923	17
8_20349	mmu-miR-181d-5p	1.719879	4.71E-05	12
8_20351	mmu-miR-181c-5p	-1.16378	0.012958	12.5
9_20818	mmu-miR-100-5p	1.380642	0.007044	12.5

9_20818	mmu-miR-100-5p	1.598088	0.001449	12
9_20820	mmu-let-7a-5p	1.209294	0.005511	17
9_21132	mmu-let-7g-5p	1.060003	0.019763	17
9_21348	mmu-miR-199a-5p	-2.22406	0.03011	12
9_21672	mmu-miR-184-5p	-1.38435	0.001334	17
9_21672	mmu-miR-184-5p	3.070232	2.68E-13	12
9_21672	mmu-miR-184-5p	3.622344	6.42E-18	12.5
10_1861	-	1.785679	0.003754	12
11_2744	mmu-miR-340-5p	1.20918	0.011699	17
11_2997	mmu-miR-22-5p	1.140736	0.012888	17
11_3014	-	-1.46102	0.021376	17
11_3014	-	2.243449	0.001865	12
11_3014	-	2.714677	7.43E-06	12.5
11_3016	mmu-miR-144-5p	6.93943	0.021965	17
11_3101	mmu-miR-301a-5p	-1.39484	0.047576	12.5
11_3101	mmu-miR-301a-5p	-2.0987	0.014396	12
11_3174	mmu-miR-10a-5p	1.412372	0.012627	12.5
11_3633	mmu-miR-146a-5p	2.972129	1.45E-05	17
11_3634	mmu-miR-146a-5p	3.376223	9.41E-06	17
11_3961	mmu-miR-21a-5p	1.993921	5.67E-06	17
12_4340	-	1.725462	0.000213	12
12_4387	mmu-miR-5099	1.396044	0.005188	12
12_4701	mmu-miR-770-5p	1.28604	0.008509	12.5
12_4701	mmu-miR-770-5p	1.398929	0.003127	12
12_4701	mmu-miR-770-5p	-1.44216	0.001829	17
12_4703	mmu-miR-673-5p	-1.33661	0.004567	17
12_4703	mmu-miR-673-5p	2.019043	8.35E-06	12.5
12_4703	mmu-miR-673-5p	2.212605	1.32E-06	12
12_4725	mmu-miR-341-5p	1.415256	0.001627	12
12_4725	mmu-miR-341-5p	-1.68353	0.000118	17
12_4729	mmu-miR-370-5p	1.246112	0.004236	12
12_4741	mmu-miR-379-5p	-1.24029	0.005511	17
12_4741	mmu-miR-379-5p	1.291646	0.003396	12
12_4741	mmu-miR-379-5p	1.63477	0.00029	12.5
12_4742	mmu-miR-379-5p	1.258325	0.003754	12
12_4742	mmu-miR-379-5p	-1.28601	0.003945	17
12_4742	mmu-miR-379-5p	1.531025	0.000702	12.5
12_4745	mmu-miR-299a-5p	1.484252	0.017322	12.5
12_4775	mmu-miR-381-5p	1.14249	0.012112	12
12_4784	mmu-miR-382-5p	-1.11247	0.012997	17
12_4784	mmu-miR-382-5p	1.182706	0.00629	12
12_4784	mmu-miR-382-5p	1.430982	0.001493	12.5
13_5604	mmu-miR-9-5p	1.6029	0.000127	17
13_5636	mmu-miR-5624-5p	-4.52795	0.019763	17
13_5711	mmu-miR-582-5p	1.470191	0.001037	12

13_5711	mmu-miR-582-5p	-2.21049	2.47E-06	17
13_5907	mmu-let-7d-5p	0.996861	0.037391	17
13_5909	mmu-let-7f-5p	1.055022	0.018661	17
13_5949	mmu-miR-7a-5p	-1.15085	0.017797	12
13_5949	mmu-miR-7a-5p	1.658992	0.004276	17
13_6049	mmu-miR-5624-5p	-4.52795	0.019763	17
14_6459	mmu-miR-598-5p	1.47119	0.001091	12.5
14_6660	mmu-miR-20a-5p	-7.5556	0.00621	12
14_6929	mmu-miR-15a-5p	-2.60977	0.000689	12
15_7281	-	-1.42029	0.038446	17
15_7281	-	2.314211	0.001841	12
15_7709	mmu-miR-30b-5p	1.771733	0.006567	17
15_7711	mmu-miR-30d-5p	1.112773	0.013172	17
16_8078	mmu-miR-1224-5p	1.624475	0.000946	12.5
16_8078	mmu-miR-1224-5p	-1.83569	7.66E-05	17
16_8078	mmu-miR-1224-5p	1.864006	0.00012	12
16_8443	mmu-miR-185-5p	-6.73554	0.048147	12
16_8559	-	2.615902	0.008809	12
16_8559	-	-4.3891	0.001346	17
17_8782	mmu-let-7e-5p	1.274565	0.003959	17
17_9110	mmu-miR-7b-5p	1.393185	0.021637	17
17_9410	mmu-miR-3082-5p	1.483543	0.045289	12.5
17_9410	mmu-miR-3082-5p	2.539207	3.46E-05	12
18_10156	mmu-miR-1a-2-5p	-2.28796	0.000254	12
18_10156	mmu-miR-1a-2-5p	3.343804	7.66E-05	17
18_10338	mmu-miR-378a-5p	1.248528	0.003739	12
18_10348	mmu-miR-143-5p	1.173084	0.00629	12
18_9861	mmu-miR-12191-5p	-1.36885	0.029834	17
18_9861	mmu-miR-12191-5p	1.727113	0.011454	12.5
19_10604	mmu-miR-204-5p	1.382601	0.001346	17
19_10718	mmu-miR-146b-5p	2.011257	1.87E-05	17

## Apêndice V: Lista de lncRNAs diferencialmente expressos (DE)

lncRNA_ID	Símbolo	LFC	FDR	Dia DE
ENSMUST00000084713	Gm10475-201	3.075844	9.99E-05	17
ENSMUST00000111270	4631405J19Rik-202	-1.48655	0.0176	17
ENSMUST00000111906	Kcnp2-204	-0.78123	0.028014	17
ENSMUST00000122926	4930470G03Rik-201	1.881172	5.41E-07	17
ENSMUST00000122926	4930470G03Rik-201	2.142584	4.15E-06	15
ENSMUST00000125308	Gm15283-201	2.062447	1.25E-07	17
ENSMUST00000125308	Gm15283-201	2.470027	5.16E-11	12.5
ENSMUST00000125308	Gm15283-201	2.599457	1.36E-05	15
ENSMUST00000128165	Rapgef4os2-201	2.54441	0.010923	17
ENSMUST00000137373	Gm16174-201	3.646438	2.35E-08	17
ENSMUST00000137629	Gm14342-201	1.350382	0.010407	12.5
ENSMUST00000140000	Gm15983-201	2.667521	0.000437	17
ENSMUST00000140000	Gm15983-201	-2.96827	0.017007	15
ENSMUST00000140000	Gm15983-201	3.054865	0.001368	12.5
ENSMUST00000145510	Gm12802-201	1.613747	0.001815	15
ENSMUST00000145510	Gm12802-201	2.150608	6.18E-07	17
ENSMUST00000148585	2610203C22Rik-204	2.173408	0.036267	17
ENSMUST00000153753	C230035I16Rik-202	1.86307	0.033717	17
ENSMUST00000153785	Fgf2os-201	2.157875	0.000146	12.5
ENSMUST00000153785	Fgf2os-201	3.762125	3.36E-10	15
ENSMUST00000153785	Fgf2os-201	6.305324	1.32E-43	17
ENSMUST00000154848	Parvaos-201	-2.0027	0.00452	17
ENSMUST00000155331	Sept9-210	1.02557	0.04606	17
ENSMUST00000160385	4833422C13Rik-202	1.530777	2.47E-05	17
ENSMUST00000163836	Tmem250-ps-202	6.765659	0.000291	17
ENSMUST00000163892	4930523C07Rik-203	1.512152	0.000298	17
ENSMUST00000173672	Neat1-201	1.601282	0.009119	12.5
ENSMUST00000174728	BC023719-201	4.001501	0.01613	12.5
ENSMUST00000174728	BC023719-201	6.539766	0.021633	17
ENSMUST00000176742	Gm20633-201	1.565979	0.000673	17
ENSMUST00000176760	Gm20632-201	1.465588	0.014656	17
ENSMUST00000180497	Gm6634-201	7.308509	0.001289	17
ENSMUST00000180519	F630040K05Rik-201	1.720982	0.000146	12.5
ENSMUST00000180569	9330179D12Rik-201	3.185637	0.002174	17
ENSMUST00000181222	Gm16754-201	1.578619	1.36E-05	17
ENSMUST00000181222	Gm16754-201	2.03952	1.32E-08	12.5
ENSMUST00000181222	Gm16754-201	2.054039	7.87E-06	15
ENSMUST00000181362	Gm26704-201	2.214993	0.003661	17
ENSMUST00000181502	Gm17586-201	1.252096	0.008816	17

ENSMUST00000181502	Gm17586-201	-4.27545	9.93E-05	15
ENSMUST00000185662	Mir155hg-202	2.660932	0.01247	17
ENSMUST00000186132	Gm29611-201	2.689308	2.23E-05	12.5
ENSMUST00000190320	B130024G19Rik-203	1.664639	0.009167	17
ENSMUST00000190488	BC055402-201	-2.22247	2.85E-05	15
ENSMUST00000190488	BC055402-201	-2.50307	5.17E-08	17
ENSMUST00000191603	2900052N01Rik-203	0.762203	0.001965	17
ENSMUST00000192691	Gm38250-201	5.168278	0.003827	17
ENSMUST00000194018	Gm37799-201	6.961893	0.00043	17
ENSMUST00000197321	Gm10030-203	1.40701	0.000917	17
ENSMUST00000197905	Gm32051-202	4.118818	0.003191	17
ENSMUST00000199631	Gm15903-202	-2.06812	0.00043	17
ENSMUST00000199941	AI506816-205	1.27184	0.002534	17
ENSMUST00000201551	Gm19590-201	2.968668	0.008816	17
ENSMUST00000204911	4930417O13Rik-203	-2.67557	0.027071	17
ENSMUST00000206358	AU020206-202	1.095811	0.013521	17
ENSMUST00000206829	Gm35040-201	1.884389	0.005551	12.5
ENSMUST00000206829	Gm35040-201	2.823751	5.01E-05	15
ENSMUST00000206829	Gm35040-201	2.878883	9.32E-08	17
ENSMUST00000207965	Gm44658-201	3.184172	4.04E-08	17
ENSMUST00000208024	Gm15635-204	-1.82387	3.88E-05	17
ENSMUST00000208302	Gm44751-201	1.416987	0.002179	17
ENSMUST00000209340	Gm45552-201	-2.88226	0.006761	17
ENSMUST00000211251	Gm6145-206	-1.5911	0.001395	17
ENSMUST00000211483	Gm45601-201	3.207652	0.045617	15
ENSMUST00000212253	Gm39822-201	2.586263	0.010602	17
ENSMUST00000216170	AC160562.1-201	1.715515	1.85E-05	17
ENSMUST00000216217	Gm47887-201	1.605184	0.000111	17
ENSMUST00000217082	A730065G17Rik-201	-2.12533	0.035874	17
ENSMUST00000220166	Lilrb4-211	3.800689	0.002308	17
ENSMUST00000223092	Gm48236-201	1.789483	1.85E-05	17
ENSMUST00000224993	Gm14005-209	2.450585	0.01247	17
ENSMUST00000227651	AC112142.1-207	1.302448	0.037962	12.5
ENSMUST00000227651	AC112142.1-207	2.165923	0.000203	15
ENSMUST00000227651	AC112142.1-207	3.112551	1.78E-15	17
1467	-	1.42649	0.007927	17
5283	-	-1.43571	0.000381	17
5283	-	-1.77644	0.015995	12
5998	-	3.776921	9.02E-06	17
7807	-	-9.25526	3.56E-06	15
7810	-	3.480742	0.000153	17
9166	-	7.818163	0.000995	15
9243	-	1.009896	0.000917	17
9243	-	1.24311	4.01E-05	12.5
10751	-	1.177056	0.018734	15

10755	-	9.537931	1.30E-07	15
14112	-	0.680605	0.024023	17
15895	-	1.965859	0.008801	12.5
15895	-	3.316449	3.27E-08	17
15895	-	3.96205	1.33E-06	15
17081	-	9.680042	4.91E-10	15
18064	-	1.839809	0.013907	17
19170	-	1.005118	0.023276	15
19170	-	1.479097	6.34E-06	12.5
21404	-	-2.84611	0.015995	12
24186	-	9.855245	0.029958	15
26803	-	7.686738	0.014202	15
27101	-	-4.55707	2.58E-06	15
27113	-	3.415825	0.034011	15
27117	-	6.869515	0.002875	15
27271	-	0.803662	0.0176	17
27272	-	-8.4708	8.37E-07	17
27272	-	-8.5159	8.53E-06	12
27273	-	1.41518	0.0021	17
27281	-	-1.10467	0.015067	17
28256	-	7.814485	0.021653	15
28637	-	-6.68521	0.000578	15
28859	-	3.094329	3.85E-14	17
29322	-	2.276419	0.040309	15
29511	-	1.433168	1.31E-06	17
31374	-	-1.09795	0.004487	17
31374	-	-1.28369	0.001428	12
31831	-	3.045044	4.72E-10	17
32492	-	8.680915	0.028686	15
32952	-	-17.411	0.049903	17
32961	-	2.695573	0.045094	15
33014	-	2.769416	0.013894	15
33525	-	1.849268	1.05E-07	17
36289	-	1.67278	4.64E-11	17
37449	-	-4.3905	0.006153	15
38018	-	7.870933	0.000318	15
38029	-	9.726906	0.00417	17
38030	-	-8.82424	0.013508	17
38601	-	-3.98594	2.13E-05	15
39137	-	0.896626	0.018663	17
39286	-	1.733414	0.00021	17
39301	-	-9.15103	8.01E-05	12
39303	-	8.101224	0.002375	15
41628	-	1.114516	0.006808	15
41628	-	1.84513	1.38E-09	12.5

42839	-	1.072206	0.010407	12.5
43799	-	1.633208	0.030387	17
43803	-	1.887515	0.029998	17
44936	-	2.765112	3.85E-14	17
45100	-	-1.36249	0.027915	17
45454	-	1.004195	0.0002	17
45454	-	1.139596	5.48E-05	15
45467	-	1.04531	0.006153	15
45467	-	1.483846	4.28E-09	17
47193	-	-1.67597	0.049095	17
49471	-	5.183678	0.011202	15
51818	-	1.578257	0.008141	15
51818	-	2.091204	6.92E-07	17
51819	-	1.717729	0.000151	17
51820	-	1.947013	0.019948	17
51825	-	1.880932	0.002592	17
51826	-	1.595459	1.24E-05	15
51826	-	2.080278	2.26E-12	17
53165	-	1.526361	0.008696	17
54361	-	2.664269	0.013067	17
54506	-	0.964789	0.011118	17
55958	-	-1.0743	0.000793	17
57074	-	6.786382	0.008311	15
59025	-	7.897421	0.000265	15
59032	-	-8.21495	4.71E-10	15
60629	-	1.389711	0.017914	12.5
60629	-	-1.80821	0.001045	12
60632	-	1.280518	0.000699	12.5
60634	-	2.510835	0.009119	12.5
60637	-	-11.5619	1.87E-13	17
61677	-	2.443567	0.000266	12.5
61677	-	2.462272	0.049087	15
61677	-	4.014949	2.61E-14	17
61879	-	2.461184	3.85E-14	17
61951	-	0.82981	0.00043	17
61964	-	2.92552	3.27E-08	17
61965	-	2.641764	0.000197	17
61965	-	3.359914	0.000513	15
61969	-	2.189977	1.71E-05	15
61969	-	3.165145	1.03E-15	17
62942	-	10.33531	0.000403	15
62942	-	-9.89117	0.004384	12
64114	-	-8.09213	1.63E-06	17
64131	-	1.643543	0.014202	15
65936	-	-8.63392	0.000523	12

65938	-	-8.38605	0.045094	15
66828	-	-4.29787	0.000521	15
67807	-	-6.6288	0.014202	15
68204	-	-1.00616	0.008241	12
69429	-	-2.58624	0.012163	17
69892	-	8.746632	0.008454	17
69893	-	-4.41595	1.13E-05	15
70256	-	-3.81863	0.010257	15
70827	-	1.624611	0.010036	12.5
70827	-	2.27809	7.94E-09	17
72414	-	1.496002	0.001607	17
72430	-	1.374254	2.42E-05	17
75417	-	7.704708	0.000513	15
75558	-	3.68236	0.000125	17
75559	-	1.309772	1.10E-07	17
75915	-	-7.27198	0.033012	15
75915	-	-9.4314	0.000171	12.5
75916	-	7.346592	0.039759	15
75920	-	7.902322	0.000187	15
76072	-	0.928787	0.030387	17
76255	-	-3.21862	0.008259	15
77486	-	2.638191	5.60E-08	17
78730	-	-8.3767	0.002787	15
79180	-	-8.2449	0.021314	12
79423	-	-0.65499	0.044287	17
82513	-	1.490024	0.000251	17
83625	-	1.490454	0.005401	17
83650	-	-7.35011	0.000675	15
83926	-	1.883216	3.43E-05	17
84536	-	2.6435	0.048779	12.5
86102	-	1.019526	0.0084	17
88633	-	0.679354	0.049095	17
89424	-	1.147006	0.00016	17
91637	-	-1.77813	0.02559	12.5
93334	-	2.906735	8.04E-05	17
93362	-	9.30005	1.93E-09	12
93362	-	9.992823	4.82E-12	12.5
94556	-	1.841791	5.40E-07	15
94556	-	2.313548	8.89E-12	17
94559	-	1.840906	0.044287	17
94566	-	1.71832	0.019699	17
95387	-	4.293256	0.006153	15
97798	-	1.156842	0.021297	17
99280	-	-8.40407	2.01E-07	15
99683	-	-1.76073	0.000124	17



100111	-	1.153534	0.023836	15
100521	-	0.973692	0.002134	17
101122	-	-7.77421	3.53E-06	15
101776	-	3.966194	0.000124	17
103070	-	-8.70141	2.58E-06	15
104674	-	1.180456	0.020932	17
106162	-	5.373748	0.007669	12
106162	-	6.093145	0.000146	12.5
107091	-	-2.08447	0.023758	17
111302	-	-7.17396	0.001598	15
111542	-	-1.29403	0.000298	17
111542	-	-1.34049	0.011555	15
113151	-	1.199977	0.000577	17
113229	-	8.165679	1.98E-05	15
113234	-	-10.3031	2.82E-10	12.5
113517	-	1.496131	1.43E-05	17
114485	-	1.05146	0.005196	17
114812	-	8.255679	0.000577	15
115731	-	-8.16513	0.007221	17
118615	-	-8.58306	4.30E-14	15
121281	-	-7.46129	3.53E-06	15
122796	-	1.269104	7.16E-05	12.5
126451	-	-9.64661	0.009841	15
127074	-	0.856045	0.029958	15
127074	-	0.995566	0.000156	17
133068	-	-1.79511	2.23E-05	17
134166	-	0.736227	0.018663	17
138370	-	2.253605	6.18E-06	17
138929	-	-1.64877	0.001395	17
138931	-	-1.46064	0.03905	15
138931	-	-1.69171	1.36E-05	17
139788	-	3.227904	3.35E-07	15
139791	-	2.034004	0.000566	12.5
141779	-	9.588655	0.002943	17
143462	-	-2.39199	0.011422	17
144762	-	-1.52956	0.010602	17
145002	-	1.797441	0.000904	17
147687	-	2.58312	0.004053	17
148409	-	-0.74038	0.042404	17
150535	-	-1.0206	0.036267	17
153082	-	-3.45253	0.036819	15
153446	-	0.866498	0.019948	17

## Apêndice VI: Lista de RNAs que pertencem a rede “darkred”

ID	Tipo RNA	ID	Tipo RNA
11_3014	miRNA	16:32614545 32615373	circRNA
16:32618197 32619155	circRNA	16_8559	miRNA
19170	lncRNA	28938	lncRNA
41628	lncRNA	42839	lncRNA
5_16372	miRNA	Arrb1	mRNA
Bag3	mRNA	Eno1	mRNA
Gapdh	mRNA	Gm15283	lncRNA
Gpi1	mRNA	P4ha1	mRNA
P4ha2	mRNA	P4hb	mRNA
Pdk1	mRNA	Pdk3	mRNA
Pgk1	mRNA	Ptgs2	mRNA
Tfrc	mRNA	Vegfa	mRNA
X:93800124 93814077	circRNA	mmu-miR-1224-5p	miRNA
mmu-miR-126a-5p	miRNA	mmu-miR-135b-5p	miRNA
mmu-miR-143-5p	miRNA	mmu-miR-148a-5p	miRNA
mmu-miR-182-5p	miRNA	mmu-miR-183-5p	miRNA
mmu-miR-24-2-5p	miRNA	mmu-miR-24-3p	miRNA
mmu-miR-301a-5p	miRNA	mmu-miR-3082-5p	miRNA
mmu-miR-30c-5p	miRNA	mmu-miR-30e-5p	miRNA
mmu-miR-322-5p	miRNA	mmu-miR-370-5p	miRNA
mmu-miR-378a-5p	miRNA	mmu-miR-379-5p	miRNA
mmu-miR-532-5p	miRNA	mmu-miR-582-5p	miRNA
mmu-miR-676-5p	miRNA	mmu-miR-7068-5p	miRNA
mmu-miR-708-5p	miRNA	mmu-miR-770-5p	miRNA
mmu-miR-879-5p	miRNA	mmu-miR-92a-2-5p	miRNA
mmu-miR-93-5p	miRNA	rno-miR-1224	miRNA

## Apêndice VII: Lista de RNAs que pertencem a rede “brown”

ID	Tipo RNA	ID	Tipo RNA
100521	lncRNA	10:59166436 59178789	circRNA
10:59192098 59192747	circRNA	113151	lncRNA
11_3014	miRNA	135610	lncRNA
141440	lncRNA	1500015a07rik	mRNA
153446	lncRNA	15_7281	miRNA
16_8559	miRNA	18:35221976 35223486	circRNA
18:5037065 5049451	circRNA	18:5046546 5049451	circRNA
19:24112744 24120901	circRNA	19:24130853 24138830	circRNA
1:52479115 52490632	circRNA	28859	lncRNA
2:164424646 164424936	circRNA	36289	lncRNA
3:122747816 122748404	circRNA	3:122781494 122817999	circRNA
44936	lncRNA	45454	lncRNA
4833422c13rik	mRNA	4930470g03rik	mRNA
4930523c07rik	mRNA	4:16151885 16158398	circRNA
5_16372	miRNA	61677	lncRNA
61964	lncRNA	61969	lncRNA
6:139992585 140028220	circRNA	70827	lncRNA
72430	lncRNA	75558	lncRNA
75559	lncRNA	7:134724030 134876012	circRNA
7:134782048 134788984	circRNA	7:134847839 134876012	circRNA
7:19696268 19696940	circRNA	82513	lncRNA
Ac112142	mRNA	Ac160562	mRNA
Ackr2	mRNA	Actn4	mRNA
Acvr1	mRNA	Ai662270	mRNA
Amotl1	mRNA	Ano1	mRNA
Aox1	mRNA	Apoe	mRNA
Arhgap11a	mRNA	Arhgef5	mRNA
Asf1b	mRNA	Aspm	mRNA
Au020206	mRNA	Axl	mRNA
Bcl2	mRNA	Birc5	mRNA
Ccl7	mRNA	Ccnb2	mRNA
Ccnd2	mRNA	Cd248	mRNA
Cd34	mRNA	Cd44	mRNA
Cd47	mRNA	Cd69	mRNA
Cdc45	mRNA	Cdc6	mRNA
Cdk1	mRNA	Cdkn3	mRNA
Cebpb	mRNA	Ch25h	mRNA
Clca2	mRNA	Clc6	mRNA
Cntf	mRNA	Col4a1	mRNA
Col4a2	mRNA	Ctnna1	mRNA
Cyp39a1	mRNA	Dll4	mRNA

Dock1	mRNA	Dpp4	mRNA
Edn1	mRNA	Edn2	mRNA
Edn3	mRNA	Ednra	mRNA
Egr1	mRNA	Egr2	mRNA
Eif2ak2	mRNA	Emp3	mRNA
Entpd1	mRNA	Entpd5	mRNA
Ets1	mRNA	Ezr	mRNA
F2rl2	mRNA	F3	mRNA
Fbln1	mRNA	Fbxo5	mRNA
Fes	mRNA	Fgf23	mRNA
Fgf2os	lncRNA	Fgfr1	mRNA
Fn1	mRNA	Fos	mRNA
Fosb	mRNA	Fxyd1	mRNA
Gata2	mRNA	Ggt5	mRNA
Gins2	mRNA	Gja1	mRNA
Gm10825	lncRNA	Gm15513	lncRNA
Gm31812	lncRNA	Gm44751	lncRNA
Gng11	mRNA	Gng12	mRNA
Gpx3	mRNA	Hck	mRNA
Helz2	mRNA	Hmox1	mRNA
Icam1	mRNA	Ifnlr1	mRNA
Il6st	mRNA	Irf7	mRNA
Irs2	mRNA	Isg15	mRNA
Itga5	mRNA	Itgb3	mRNA
Jak3	mRNA	Kif14	mRNA
Kif18a	mRNA	Kifc1	mRNA
Krt18	mRNA	Krt8	mRNA
Lama5	mRNA	Lamb1	mRNA
Lrrk1	mRNA	Lyn	mRNA
Mafk	mRNA	Map3k8	mRNA
Mcm10	mRNA	Mcm4	mRNA
Mgp	mRNA	Micall2	mRNA
Myh11	mRNA	Myh9	mRNA
Nab1	mRNA	Nlrp3	mRNA
Nod2	mRNA	Nostrin	mRNA
Notch3	mRNA	Ntrk2	mRNA
P2rx7	mRNA	P2ry1	mRNA
P2ry2	mRNA	Pde5a	mRNA
Pdgfb	mRNA	Pdgfrb	mRNA
Pdlim4	mRNA	Plcg2	mRNA
Plcz1	mRNA	Plekhg5	mRNA
Plekhg6	mRNA	Plxnd1	mRNA
Pmaip1	mRNA	Pmp22	mRNA
Prkcd	mRNA	Pros1	mRNA
Rdh5	mRNA	Rel	mRNA

Rela	mRNA	Rgr	mRNA
Ripk2	mRNA	Rsad2	mRNA
Sdc4	mRNA	Sele	mRNA
Sema3b	mRNA	Sema3g	mRNA
Sept1	mRNA	Sept10	mRNA
Serpine1	mRNA	Skap2	mRNA
Slc3a2	mRNA	Slc7a5	mRNA
Smtn	mRNA	Sorbs3	mRNA
Sox13	mRNA	Stat1	mRNA
Stat3	mRNA	Stat6	mRNA
Svil	mRNA	Tac1	mRNA
Tcf7	mRNA	Tfpi	mRNA
Tgfbr2	mRNA	Tgfbr3	mRNA
Tjp2	mRNA	Tnfrsf1a	mRNA
Tnip2	mRNA	Trip10	mRNA
Tuba1c	mRNA	Wasf2	mRNA
mmu-miR-101a-5p	miRNA	mmu-miR-10b-5p	miRNA
mmu-miR-12191-5p	miRNA	mmu-miR-1224-5p	miRNA
mmu-miR-126a-5p	miRNA	mmu-miR-129-5p	miRNA
mmu-miR-135b-5p	miRNA	mmu-miR-143-5p	miRNA
mmu-miR-148a-5p	miRNA	mmu-miR-181b-5p	miRNA
mmu-miR-181d-5p	miRNA	mmu-miR-182-5p	miRNA
mmu-miR-183-5p	miRNA	mmu-miR-1839-5p	miRNA
mmu-miR-184-5p	miRNA	mmu-miR-199a-5p	miRNA
mmu-miR-199b-5p	miRNA	mmu-miR-1a-1-5p	miRNA
mmu-miR-1a-2-5p	miRNA	mmu-miR-204-5p	miRNA
mmu-miR-205-5p	miRNA	mmu-miR-206-5p	miRNA
mmu-miR-20a-5p	miRNA	mmu-miR-219a-5p	miRNA
mmu-miR-22-5p	miRNA	mmu-miR-24-2-5p	miRNA
mmu-miR-24-3p	miRNA	mmu-miR-296-5p	miRNA
mmu-miR-301a-5p	miRNA	mmu-miR-3082-5p	miRNA
mmu-miR-30c-5p	miRNA	mmu-miR-30e-5p	miRNA
mmu-miR-322-5p	miRNA	mmu-miR-340-5p	miRNA
mmu-miR-370-5p	miRNA	mmu-miR-375-5p	miRNA
mmu-miR-379-5p	miRNA	mmu-miR-381-5p	miRNA
mmu-miR-382-5p	miRNA	mmu-miR-5099	miRNA
mmu-miR-532-5p	miRNA	mmu-miR-5624-5p	miRNA
mmu-miR-582-5p	miRNA	mmu-miR-673-5p	miRNA
mmu-miR-708-5p	miRNA	mmu-miR-770-5p	miRNA
mmu-miR-7a-5p	miRNA	mmu-miR-873a-5p	miRNA
mmu-miR-879-5p	miRNA	mmu-miR-9-5p	miRNA
mmu-miR-92a-2-5p	miRNA	mmu-miR-93-5p	miRNA
rno-miR-1224	miRNA	rno-miR-181c-5p	miRNA