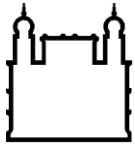


MINISTÉRIO DA SAÚDE  
FUNDAÇÃO OSWALDO CRUZ  
INSTITUTO OSWALDO CRUZ  
Doutorado em Programa de Pós-Graduação Biologia Computacional e Sistemas

**UTILIZAÇÃO DE DADOS DE SEQUENCIAMENTO DE ALTA VAZÃO DE  
PEQUENOS RNAs NÃO CODIFICADORES NA DISTINÇÃO DE AMOSTRAS  
HUMANAS**

**NATASHA ANDRESSA NOGUEIRA JORGE**

Rio de Janeiro  
Setembro de 2017



Ministério da Saúde

FIOCRUZ

Fundação Oswaldo Cruz

## **INSTITUTO OSWALDO CRUZ**

**Programa de Pós-Graduação em Biologia Computacional e Sistemas**

***NATASHA ANDRESSA NOGUEIRA JORGE***

*UTILIZAÇÃO DE RNAS NÃO CODIFICADORES PEQUENOS NA DISTINÇÃO DE AMOSTRAS HUMANAS DE CÂNCER DE PULMÃO E DE CÉLULAS DO SANGUE*

Tese apresentada ao Instituto Oswaldo Cruz  
como parte dos requisitos para obtenção do  
título de Doutor em Ciências

**Orientador:** Dr. Fabio Passetti

**RIO DE JANEIRO**

Setembro de 2017

Jorge, Natasha Andressa Nogueira.

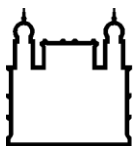
UTILIZAÇÃO DE DADOS DE SEQUENCIAMENTO DE ALTA VAZÃO DE PEQUENOS RNAs NÃO CODIFICADORES NA DISTINÇÃO DE AMOSTRAS HUMANAS / Natasha Andressa Nogueira Jorge. - Rio de Janeiro, 2017.  
157 f.

Tese (Doutorado) - Instituto Oswaldo Cruz, Pós-Graduação em Biologia Computacional e Sistemas, 2017.

Orientador: Fabio Passetti.

Bibliografia: f. 82-90

1. Pequenos RNAs Não Codificadores. 2. piRNAs. 3. snoRNA. 4. Tabagismo. 5. Hematopoiese. I. Título.



Ministério da Saúde

FIOCRUZ

Fundação Oswaldo Cruz

## **INSTITUTO OSWALDO CRUZ**

**Programa de Pós-Graduação em Biologia Computacional e Sistemas**

***AUTOR: NATASHA ANDRESSA NOGUEIRA JORGE***

**UTILIZAÇÃO DE RNAs NÃO CODIFICADORES PEQUENOS NA DISTINÇÃO DE AMOSTRAS HUMANAS DE CÂNCER DE PULMÃO E DE CÉLULAS DO SANGUE**

**ORIENTADOR: Dr. Fabio Passetti**

**Aprovada em: 28/09/2017**

### **EXAMINADORES:**

**Prof. Dr. Thiago Estevam Parente Martins - Presidente**  
Instituto Oswaldo Cruz, Fiocruz

**Prof. Dr. Nicole de Miranda Scherer**  
Instituto Nacional de Câncer, Rio de Janeiro

**Prof. Dr. Pedro Alexandre Favoretto Galante**  
Instituto de Ensino e Pesquisa / Hospital Sírio-Libanês, São Paulo

Rio de Janeiro, 28 de setembro de 2017



Ministério da Saúde

Fundação Oswaldo Cruz  
Instituto Oswaldo Cruz

Ata da defesa de tese de doutorado em Biologia Computacional e Sistemas de **Natasha Andressa Nogueira Jorge**, sob orientação do Dr. Fabio Passetti. Ao vigésimo oitavo dia do mês de setembro de dois mil e dezessete, realizou-se às treze horas, no Auditório Maria Deane/FIOCRUZ, o exame da tese de doutorado intitulada: "**Utilização de Dados de Sequenciamento de alta Vazão de Pequenos RNAs Não Codificadores na Distingção de Amostras Humanas**" No programa de Pós-graduação em Biologia Computacional e Sistemas do Instituto Oswaldo Cruz, como parte dos requisitos para obtenção do título de Doutora em Ciências - área de concentração: Genômica Funcional, Evolução e Filogenômica, na linha de pesquisa: Genoma, transcriptoma, proteoma e metaboloma. A banca examinadora foi constituída pelos Professores: Dr. Thiago Estevam Parente Martins - ENSP/FIOCRUZ (Presidente), Dr<sup>a</sup>. Nicole de Miranda Scherer - INCA/RJ, Dr. Pedro Alexandre Favoretto Galante - HSL/SP e como suplentes: Dr. Antonio Basilio de Miranda – IOC/FIOCRUZ e Dr. Diogo Antonio Tschoeke – UFRJ/RJ. Após arguir a candidata e considerando que a mesma demonstrou capacidade no trato do tema escolhido e sistematização da apresentação dos dados, a banca examinadora pronunciou-se pela APROVAÇÃO da defesa da tese de doutorado. De acordo com o regulamento do Curso de Pós-Graduação em Biologia Computacional e Sistemas do Instituto Oswaldo Cruz, a outorga do título de Doutora em Ciências está condicionada à emissão de documento comprobatório de conclusão do curso. Uma vez encerrado o exame, o Coordenador do Programa, Dr. Ernesto Raúl Caffarena, assinou a presente ata tomando ciência da decisão dos membros da banca examinadora. Rio de Janeiro, 28 de setembro de 2017

Dr. Thiago Estevam Parente Martins (Presidente da Banca):

Dr<sup>a</sup>. Nicole de Miranda Scherer (Membro da Banca):

Dr. Pedro Alexandre Favoretto Galante (Membro da Banca):

Dr. Ernesto Raúl Caffarena (Coordenador do Programa):

**Ao trio parada dura.**

## **AGRADECIMENTOS**

Ao trio parada dura, mais conhecidos como minha mãe Regina, meu pai José e minha tia Márcia, por todo suporte, carinho, preocupação e oração por mim ao longo de toda a minha vida. Por traduzirem dialetos, enfrentarem o Tâmba e puxarem minha orelha. Esses últimos quatro anos seriam impossíveis sem vocês.

Ao meu orientador Fabio Passeti pelos últimos 10 anos de trabalho juntos, desde a iniciação científica até o doutorado. Pelos conselhos profissionais e pessoais, por toda ajuda e orientação, por ser um exemplo a ser seguido.

Aos meus colegas de laboratório Raphael Tavares e Gabriel Wajnberg por todo o suporte e amizade, desde a iniciação científica e mestrado até hoje e pelo futuro. Por todo companheirismo e apoio.

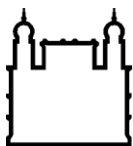
A todos os colegas do laboratório de Genômica Funcional e Bioinformática do IOC e do Laboratório de Bioinformática e Biologia Computacional do INCA por toda ajuda e momentos de descontração.

Ao Dr. Ernest Turro da Cambridge University que aceitou me orientar durante o período sanduíche, a todos os colegas que me receberam no Haematology Laboratory e à Laurentya Olga. Por uma experiência que mudou minha visão de ciência, aprendi muito com todos vocês.

Às agências de fomento CAPES, CNPq e FAPERJ pelo auxílio financeiro.

Mar calmo nunca fez bom marinheiro.  
(Provérbio inglês)





Ministério da Saúde

FIOCRUZ

Fundação Oswaldo Cruz

## **INSTITUTO OSWALDO CRUZ**

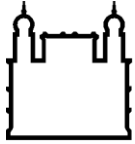
### **UTILIZAÇÃO DE RNAs NÃO CODIFICADORES PEQUENOS NA DISTINÇÃO DE AMOSTRAS HUMANAS DE CÂNCER DE PULMÃO E DE CÉLULAS DO SANGUE**

#### **RESUMO**

#### **TESE DE DOUTORADO EM BIOLOGIA COMPUTACIONAL E SISTEMAS**

**Natasha Andressa Nogueira Jorge**

Os RNAs não codificadores pequenos, como piRNAs, snoRNAs e miRNAs, atuam em diversas etapas do metabolismo celular, participando de processos como controle da expressão gênica, metilação e pseudouridilação de rRNAs e diferenciação celular. Neste trabalho, utilizamos dados públicos de sequenciamento de alta vazão de RNAs pequenos para avaliar a importância de RNAs não codificadores pequenos na biologia das células saudáveis e de neoplasia maligna. Na primeira análise, utilizamos dados de amostras normais e tumorais de pacientes de adenocarcinoma de pulmão obtidos no banco de dados do TCGA. Estas amostras foram utilizadas para avaliar o perfil de expressão de snoRNAs e piRNAs. Nesta análise, revelamos que o fumo inibe diversos snoRNAs relacionados à manutenção celular e promove a expressão de outros snoRNAs também encontrados mais expressos em diversos tumores, alterando, assim, o perfil de expressão de células normais para um que se assemelha mais a tumores. Também mostramos que os tumores de pulmão de não fumantes e fumantes são bastante heterogêneos e aconselhamos que sejam estudados separadamente. Nesta mesma análise, apontamos também snoRNAs cuja expressão não se altera em nenhuma das condições avaliadas, podendo assim serem usados como parâmetro para normalização de experimentos de biologia molecular. A segunda análise envolve a identificação de miRNAs diferencialmente expressos entre linfócitos T CD4 virgens, megacariócitos, eritroblastos, neutrófilos, monócitos e macrófagos M1 e M2 do projeto Blueprint. Nesta etapa, além de identificar os miRNAs, utilizamos dados de sequenciamento de alta vazão de mRNAs e um banco de dados de alvos validados para identificar as vias metabólicas inibidas por miRNAs durante o processo de diferenciação destes tipos celulares. Esta análise corrobora dados da literatura do papel de miRNAs no processo de diferenciação e ativação de diversas células da linhagem hematopoética. Em conjunto, as duas análises reforçam o papel dos RNAs não codificadores na manutenção do metabolismo celular.



Ministério da Saúde

FIOCRUZ

Fundação Oswaldo Cruz

## **INSTITUTO OSWALDO CRUZ**

### **USING SMALL NON CODING RNA TO DISTINGUISH HUMAN LUNG CANCER AND BLOOD CELL SAMPLES**

#### **ABSTRACT**

#### **PHD THESIS IN SYSTEMS AND COMPUTATIONAL BIOLOGY**

**Natasha Andressa Nogueira Jorge**

The small non-coding RNAs, such as piRNAs, snoRNAs and miRNAs act at several stages of cellular metabolism, participating in processes such as gene expression control, molecule modification and cell differentiation. In this work, we used public high-throughput sequencing data of small RNAs to evaluate the importance of small non-coding RNAs in healthy cell and malignant neoplasm biology. In the first analysis, we used data from normal and tumor samples from lung adenocarcinoma patients obtained from the TCGA database. These samples were used to evaluate the expression profile of snoRNAs and piRNAs. In this analysis, we have shown that smoking inhibits several snoRNAs related to cell maintenance and promotes the expression of other snoRNAs also found to be more expressed in several tumors, thereby altering the expression profile of normal cells towards one that resembles more tumors cells. We have also demonstrated that lung tumors of nonsmokers and smokers are quite heterogeneous and should be studied separately. In this same analysis, we also point out snoRNAs whose expression does not change in any of the cellular types evaluated, and can thus be used as a parameter for quantification in molecular biology experiments. The second analysis involves the identification of differentially expressed miRNAs between lymphocytes T CD4 naive, megakaryocytes, erythrocytes, neutrophils, monocytes, and macrophages M1 and M2 from the Blueprint project. In this step, in addition to identifying altered miRNAs, we used high-throughput sequencing data from mRNAs and a database of validated targets to identify metabolic pathways inhibited by miRNAs during the differentiation process of these cell types. This analysis corroborates data from the literature on the role of miRNAs in the process of differentiation and activation of several cells of the hematopoietic lineage. Together, the two analyzes reinforce the role of non-coding RNAs in the maintenance of cellular metabolism.

# ÍNDICE

<b>RESUMO</b> .....	<b>VIII</b>
<b>ABSTRACT</b> .....	<b>IX</b>
<b>1. INTRODUÇÃO</b> .....	<b>1</b>
1.1. O PROJETO GENOMA HUMANO.....	1
1.2. O PROJETO ENCODE.....	2
1.2.1. <i>Os RNAs não codificadores</i> .....	2
1.3. SEQUENCIAMENTO DE ALTA VAZÃO (HTS).....	13
1.4. CONSÓRCIOS DE SEQUENCIAMENTO.....	14
1.4.1. <i>O Projeto TCGA</i> .....	15
1.4.2. <i>O Projeto BLUEPRINT</i> .....	17
1.5. REUSO DE DADOS PÚBLICOS DE HTS.....	22
<b>2. JUSTIFICATIVA</b> .....	<b>24</b>
<b>3. OBJETIVOS</b> .....	<b>25</b>
3.1. OBJETIVO GERAL.....	25
3.2. OBJETIVOS ESPECÍFICOS.....	25
3.2.1. <i>Perfil de expressão de RNAs não codificadores pequenos em pacientes fumantes e não fumantes de adenocarcinoma de pulmão.</i> ....	25
3.2.2. <i>Perfil de expressão de miRNAs em células sanguíneas</i> .....	25
<b>4. MATERIAL E MÉTODOS</b> .....	<b>26</b>
4.1. PERFIL DE EXPRESSÃO DE RNAs NÃO CODIFICADORES PEQUENOS EM PACIENTES FUMANTES E NÃO FUMANTES DE ADENOCARCINOMA DE PULMÃO. ....	26
4.1.1. <i>Dados de sequenciamento</i> .....	26
4.1.2. <i>Pré-processamento e análise</i> .....	27
4.2. COMPARAÇÃO DO PERFIL DE MI RNAs EM CÉLULAS SANGUÍNEAS.....	29
<b>5. RESULTADOS</b> .....	<b>33</b>
5.1. PERFIL DE EXPRESSÃO DE RNAs NÃO CODIFICADORES PEQUENOS EM PACIENTES FUMANTES E NÃO FUMANTES DE ADENOCARCINOMA DE PULMÃO. ....	33
5.1.1. <i>Validação da metodologia</i> .....	33
5.1.2. <i>Comparação do perfil de expressão de piRNAs e snoRNAs em amostras normais</i> .....	35
5.1.3. <i>Comparação do perfil de expressão de piRNAs e snoRNAs em amostras tumorais</i> .....	40
5.1.4. <i>sncRNAs constitutivos em não fumantes</i> .....	45
5.1.5. <i>sncRNAs constitutivos em Fumantes</i> .....	46

5.2.	ANÁLISE DE MIRNAS DIFERENCIALMENTE EXPRESSOS ENTRE DIFERENTES TIPOS SANGUÍNEOS CELULARES DO PROJETO BLUEPRINT.....	48
<b>6.</b>	<b>DISCUSSÃO .....</b>	<b>68</b>
6.1.	SNCRNAS DIFERENCIALMENTE EXPRESSOS ENTRE NÃO FUMANTES E FUMANTES COM ADENOCARCINOMA DE PULMÃO. 68	
6.2.	MIRNAS DIFERENCIALMENTE EXPRESSOS NA ÁRVORE HEMATOPOIÉTICA.....	73
<b>7.</b>	<b>CONCLUSÕES .....</b>	<b>80</b>
7.1.	SNORNAS E PIRNAS DIFERENCIALMENTE EXPRESSOS ENTRE NÃO FUMANTES E FUMANTES COM ADENOCARCINOMA 80	
7.2.	MIRNAS DIFERENCIALMENTE EXPRESSOS NA ÁRVORE HEMATOPOIÉTICA.....	80
	<b>REFERÊNCIAS BIBLIOGRÁFICAS .....</b>	<b>82</b>
	<b>ANEXOS .....</b>	<b>91</b>
	<b>ANEXO 1: ARTIGO PUBLICADO .....</b>	<b>92</b>
	ANEXO 2: SNORNAS DIFERENCIALMENTE EXPRESSOS EM AMOSTRAS NORMAIS.....	110
	ANEXO 3: SNORNAS E PIRNAS DIFERENCIALMENTE EXPRESSOS EM AMOSTRAS TUMORAIS .....	112
	ANEXO 4: SNORNAS CONSTITUTIVAMENTE EXPRESSOS EM NÃO FUMANTES .....	114
	ANEXO 5: SNORNAS CONSTITUTIVAMENTE EXPRESSOS EM FUMANTES.....	118
	ANEXO 6: MIRNAS DIFERENCIALMENTE EXPRESSOS EM LINFÓCITOS.....	119
	ANEXO 7: MIRNAS DIFERENCIALMENTE EXPRESSOS EM MEGACARIÓCITOS E ERITROBLATOS.....	121
	ANEXO 8: MIRNAS DIFERENCIALMENTE EXPRESSOS EM NEUTRÓFILOS.....	122
	ANEXO 9: MIRNAS DIFERENCIALMENTE EXPRESSOS EM MONÓCITOS .....	126
	ANEXO 10: MIRNAS DIFERENCIALMENTE EXPRESSOS EM MAIS DE UMA COMPARAÇÃO.....	127
	ANEXO 11. VIAS METABÓLICAS ALTERADAS EM LINFÓCITOS .....	129
	ANEXO 12 VIAS METABÓLICAS ALTERADAS EM MEGACARIÓCITOS E ERITROBLATOS .....	131
	ANEXO 13. VIAS METABÓLICAS ALTERADAS EM NEUTRÓFILOS .....	133
	ANEXO 14. VIAS METABÓLICAS ALTERADAS EM MONÓCITOS .....	137

## Índice de Figuras

Figura 1.1. Famílias de ncRNAs. ....	3
Figura 1.2. Região “seed” do miRNA. ....	4
Figura 1.3. Biossíntese de miRNA. ....	5
Figura 1.4. Mecanismos de atuação dos miRNAs. A) Tradução normal. B) Inibição do início da tradução. C) Término prematuro da tradução. D) Recrutamento de enzimas que removem o 5' CAP o a cauda poli A. ....	6
Figura 1.5. Motivos de snoRNA. ....	8
Figura 1.6. Biossíntese snoRNAs. ....	10
Figura 1.7. Biogênese de piRNAs. ....	13
Figura 1.8. Árvore Hematopoiética. ....	18
Figura 4.1. Pipeline de processamento das amostras de Adenocarcinoma de pulmão. ....	27
Figura 4.2. Árvore Hematopoiética. HSC: célula tronco hematopoiética; CMP: precursor comum mieloide; CLP: precursor comum linfóide. ....	31
Figura 5.1. Diagrama de Venn entre Kim e colaboradores e o trabalho atual. O círculo azul representa a quantidade de miRNAs diferencialmente expressos encontrados por Kim e colaboradores e o círculo vermelho a quantidade encontrada no trabalho atual, entre parêntesis, a quantidade de miRNAs que não passou pelo nosso filtro. ....	35
Figura 5.2. Agrupamento das amostras normais. A) Agrupamento hierárquico. As amostras que começam com T pertencem a pacientes fumantes e as que começam com K, a pacientes não fumantes. B) Análise de componente principal. NonS representam amostras de pacientes não fumantes e S, de pacientes fumantes. ....	36
Figura 5.3. “Vulcano” plot dos genes diferencialmente expressos em amostras normais. Os genes diferencialmente expressos estão em vermelho. ....	38
Figura 5.4. “Heatmap” dos genes diferencialmente expressos em amostras normais. A barra azul indica as amostras de pacientes não fumantes e a barra cinza indica as amostras de pacientes fumantes. ....	39
Figura 5.5. Agrupamento das amostras tumorais. A) Agrupamento hierárquico. As amostras que começam com T pertencem a pacientes fumantes e as que começam com K, a pacientes não fumantes. B) Análise de componente	

principal. NonS representam amostras de pacientes não fumantes e S, de pacientes fumantes.....	40
Figura 5.6. “Volcano” plot dos genes diferencialmente expressos em amostras tumorais. Os genes diferencialmente expressos estão em vermelho.....	42
Figura 5.7. Heatmap dos genes diferencialmente expressos em amostras tumorais. A barra azul indica as amostras de pacientes não fumantes e a barra cinza indica as amostras de pacientes fumantes.....	43
Figura 5.8. Gráfico de correlação entre os logFC dos genes diferencialmente expressos. Os genes mais expressos e menos expressos formam dos grupos distintos na parte superior direita e inferior esquerda, respectivamente. ....	45
Figura 5.9. Comparação realizada para identificação de genes constitutivamente expressos. A) Comparação entre as amostras de não fumantes. B) Comparação entre as amostras de fumantes. ....	45
Figura 5.10. Perfil de expressão de miRNA por tipo celular do BLUEPRINT. ....	49
Figura 5.11. miRNAs diferencialmente expressos entre mielócitos e linfócitos. ....	51
Figura 5.12. miRNAs diferencialmente expressos entre megacariócitos e eritroblatos e os outros tipos celulares avaliados.....	52
Figura 5.13. miRNAs diferencialmente expressos em neutrófilos.....	54
Figura 5.14. miRNAs diferencialmente expressos em monócitos.....	55
Figura 5.15. miRNAs alterados em macrófagos M1. ....	57
Figura 5.16. miRNAs diferencialmente expressos em três ou mais comparações.....	58
Figura 5.17. mRNAs diferencialmente expressos em todas as análises.....	61
Figura 5.18. Vias metabólicas relacionadas à mielócitos e linfócitos. A) Primeira comparação realizada: Linfócitos vs. Mielócitos. B) As três vias metabólicas com menor p-valor relacionadas à comparação.....	64
Figura 5.19. Vias metabólicas relacionadas à GMP e MEP. A) Segunda comparação realizada: GMP vs. MEP. B) As três vias metabólicas com menor p-valor relacionadas à comparação. ....	65
Figura 5.20. Vias metabólicas relacionadas à neutrófilos e monócitos. A) Terceira comparação realizada: Neutrófilos vs. Monócitos. B) As três vias metabólicas com menor p-valor relacionadas à comparação.....	66

**Figura 5.21. Vias metabólicas relacionadas à monócitos e macrófagos. A) Quarta comparação realizada: Monócitos vs. Macrófagos. B) As três vias metabólicas com menor p-valor relacionadas à comparação.....67**

## LISTA DE TABELAS

Tabela 4.1. Tipos celulares e quantidade de amostras utilizadas para avaliação do perfil de expressão de miRNAs pelo projeto BLUEPRINT. ....	29
Tabela 4.2. Quantidade de amostras por tipo celular usada na comparação de mRNAs. ....	32
Tabela 5.1. Totais de “reads” obtidos de pacientes com adenocarcinoma. ....	34
Tabela 5.2. Os cinco snoRNAs ou piRNAs diferencialmente expressos com maiores e menores logFC entre amostras normal de não fumantes e fumantes. ....	37
Tabela 5.3. Os cinco piRNAs ou snoRNAs diferencialmente expressos com maiores e menores logFC entre amostras tumorais de não fumantes e fumantes. ....	41
Tabela 5.4. snoRNAs e piRNAs diferencialmente expressos em ambas comparações. ....	44
Tabela 5.5. Os 10 snoRNAs constitutivamente expressos com menor desvio padrão em amostras de não fumantes. ....	46
Tabela 5.6. Os 10 snoRNAs constitutivamente expressos com menor desvio padrão em fumantes. ....	47
Tabela 5.7. snoRNAs constitutivos em comum entre não fumantes e fumantes. ....	47
Tabela 5.8. <i>Reads</i> alinhados do projeto Blueprint .....	48
Tabela 5.9. Quantidade de miRNAs encontrados diferencialmente expressos. ....	50
Tabela 5.10. Os 10 miRNAs diferencialmente expressos entre linfócitos e os outros tipos celulares com maiores probabilidades posteriores. ....	50
Tabela 5.11. Os 10 miRNAs diferencialmente expressos entre GMP e MEP com maiores probabilidades posteriores. ....	52
Tabela 5.12. Os 10 miRNAs diferencialmente expressos entre neutrófilos e monócitos e macrófagos M1 e M2 com maior probabilidade posterior. ....	53
Tabela 5.13. Os 10 miRNAs diferencialmente expressos entre monócitos e macrófagos M1 e M2 com maiores probabilidades posteriores. ....	55
Tabela 5.14. Os 10 miRNAs diferencialmente entre macrófagos M1 e macrófagos M2 com maior probabilidade posterior. ....	56
Tabela 5.15. mRNAs diferencialmente expressos por comparação realizada. ....	59



Tabela 5.16. Os 10 mRNAs diferencialmente expressos com maior probabilidade posterior entre linfócitos e os outros tipos celulares. ....	59
Tabela 5.17. Os 10 mRNAs diferencialmente expressos com maior probabilidade posterior entre GMP vs. MEP.....	59
Tabela 5.18. Os 10 mRNAs diferencialmente expressos com maior probabilidade posterior entre neutrófilo e monócito e macrófago M1 e M2 .....	60
Tabela 5.19. Os 10 mRNAs diferencialmente expressos com maior probabilidade posterior entre monócito e macrófago M1 e M2 . .....	60
Tabela 5.20. Os 10 mRNAs diferencialmente expressos com maior probabilidade posterior entre macrófago M1 e M2 .. .....	60
Tabela 5.21. Quantidade de alvos menos expressos do miRNAs mais expressos encontradas. ....	62
Tabela 5.22. Quantidade de vias metabólicas alteradas por comparação. ....	62
Tabela 5.23. As 10 vias metabólicas inibidas com menor p-valor em linfócitos T CD4+.....	63
Tabela 5.24. As 10 vias metabólicas inibidas com menor p-valor em Megacariócitos e Eritroblatos. ....	65
Tabela 5.25. As 10 vias metabólicas inibidas com menor p-valor em neutrófilos. ....	66
Tabela 5.26. As 10 vias metabólicas com menor p-valor em monócitos.....	67

## LISTA DE SIGLAS E ABREVIATURAS

CLP	Precursor Linfóide Comum, do inglês “Common Lymphoid Progenitor”
CMP	Precursor Mielóide Comum, do inglês “Common Myeloid Progenitor”
DNA	Ácido desoxirribonucleico, do inglês “Deoxyribonucleic Acid”
FDR	Taxa de falsa descoberta, do inglês “False Discovery Rate”
GMP	Precursor Granulócito Monócito, do inglês “Granulocyte Monocyte Progenitor”
GTF	“General Feature Format”
HSC	Célula Tronco Hematopoiética, do inglês “Haematopoietic Stem Cell”
LogFC	Logaritmo de “Fold Change”
lncRNA	Longo RNA não codificador, do inglês “Long non coding RNA”
MEP	Precursor Megacariócito Eritroblato, do inglês “Megakaryocyte Erythrocyte Progenitor”
miRNA	MicroRNA
mRNA	RNA mensageiro, do inglês “Messenger RNA”
ncRNA	RNA não codificador, do inglês “non coding RNA”
NGS	Sequenciamento de nova geração, do inglês “Next generation sequencing”
NNonS	Amostra de tecido Normal de paciente Não Fumante, do inglês “Normal Non Smoker”
NS	Amostra de tecido Normal de paciente Fumante, do inglês “Normal Smoker”
piRNA	RNA que interage com PIWI, do inglês “PIWI interacting RNA”
pré-piRNA	PiRNAs precursoros
pri-miRNA	miRNA primário
RISC	Complexo de Indução do Silenciamento do RNA, do inglês “RNA induced silencing complex”
RNA	Ácido ribonucleico, do inglês “Ribonucleic Acid”
rRNA	RNA ribossomal
siRNA	RNAs de interferência pequenos, do inglês “small interfering RNA”
snRNA	RNA nuclear pequeno, do inglês “small nuclear RNA”
sncRNA	RNA não codificador pequeno, do inglês “small non coding RNA”
SNORD	SnoRNA C/D box

snoRNA	RNA nucleolar pequeno, do inglês “small nucleolar RNA”
snoRNP	Pequena ribonucleoproteína nucleolar, do inglês “small ribonucleoprotein”
TCGA	“The Cancer Genome Atlas”
TMM	“Trimmed Means of M-values”
TNonS	Amostra de tecido de Tumor de paciente Não Fumante, do inglês “Tumor Non Smoker”
TS	Amostra de tecido de Tumor de paciente Fumante, do inglês “Tumor Smoker”

# 1. INTRODUÇÃO

## 1.1. O Projeto Genoma Humano

O Projeto Genoma Humano foi um consórcio internacional que envolveu diversas instituições de pesquisa nos Estados Unidos, Grã-Bretanha, França, China, entre outros países. Iniciado oficialmente em 1990 e patrocinado pelo “National Institute of Health” (NIH) norte-americano e o “Wellcome Trust” britânico, os objetivos deste consórcio eram determinar a sequência dos mais de três bilhões de nucleotídeos presentes no genoma humano, mapear os genes presentes neste genoma e armazenar essas informações em bancos de dados públicos; tudo no prazo de até 15 anos (Watson 1990).

Oito anos depois, a empresa privada de biotecnologia “Celera Genomics” anunciou que também sequenciaria o genoma humano (Roberts 2001). Duas técnicas distintas de sequenciamento foram empregadas pelas iniciativas. Enquanto o consórcio público empregou o “hierarchical shotgun sequencing” (Lander et al. 2001), a iniciativa privada utilizou “whole genome shotgun sequencing” (Venter et al. 2001).

Em 2000, o presidente norte-americano Bill Clinton e o primeiro ministro inglês Tony Blair, anunciaram, em conjunto, que a Universidade da Califórnia havia concluído uma primeira versão do genoma humano. No ano seguinte, ambas as iniciativas, pública e privada, publicaram os primeiros artigos sobre o assunto nas revistas “Nature” (Lander et al. 2001) e “Science” (Venter et al. 2001), respectivamente. O genoma completo foi publicado em 14 de abril de 2003 e, no ano seguinte, a avaliação da qualidade do sequenciamento indicou que mais de 92% do genoma possuía acurácia de 99.99% (Schmutz et al. 2004).

Entre as descobertas mais importantes decorrentes do sequenciamento do genoma humano, destacam-se: a existência de aproximadamente 22.000 genes codificadores de proteínas, a existência de regiões repetitivas e que somente 7% das famílias proteicas são específicas de vertebrados (Lander et al. 2001).

## 1.2. O Projeto Encode

Após os primeiros resultados do Projeto Genoma Humano, ficou claro que ainda havia muito a ser compreendido sobre a biologia dos genomas. Assim, em 2003, foi criado o projeto ENCODE (em inglês, “Encyclopedia of DNA Elements”), cujo objetivo era mapear detalhadamente uma porção do genoma humano, incluindo regiões promotoras, genes, “enhancers”, repressores, éxons, origens de replicação, sítios de terminação, sítios de ligação de fatores de transcrição, sítios de metilação, modificações de cromatina, regiões conservadas e transcritas (Weinstock 2007).

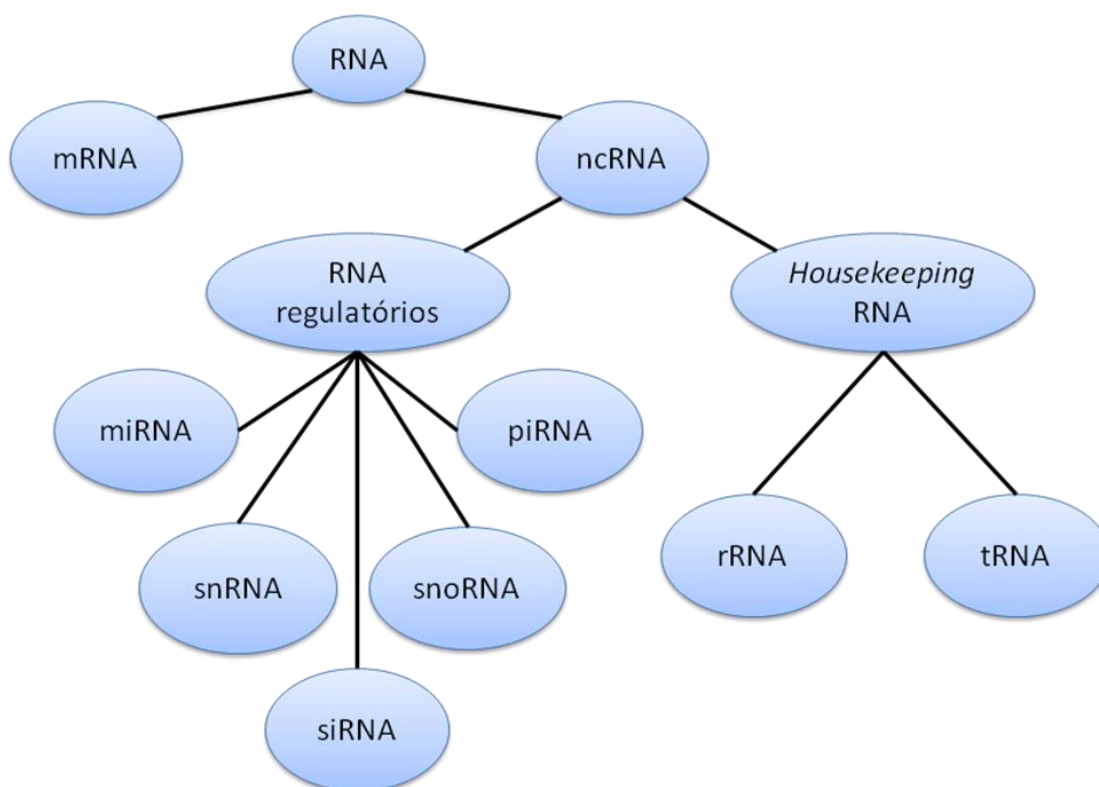
Na primeira fase deste projeto, 35 grupos de pesquisa usaram técnicas de alta vazão para examinar detalhadamente um pedaço de 1% do genoma humano que possuía tanto regiões codificadoras de proteínas quanto regiões de função desconhecida (ENCODE Project Consortium et al. 2007). Quatro anos depois, os resultados deste projeto revelaram que a maior parte do genoma humano é transcrita, identificaram novos transcritos não codificadores e caracterizaram novas regiões promotoras, entre outros elementos genéticos (ENCODE Project Consortium et al. 2007).

### 1.2.1. Os RNAs não codificadores

RNAs não codificadores (ncRNAs) são moléculas funcionais de RNA que não codificam proteínas (Griffiths-Jones 2004). Os primeiros RNAs não codificadores encontrados foram os RNAs transportadores, descritos nos anos 60 (Holley et al. 1965). Desde então, diversas novas famílias foram caracterizadas e, acredita-se que muitas ainda serão descobertas.

Assim como genes codificadores, os ncRNAs também são classificados em famílias de acordo com sua estrutura (Griffiths-Jones 2004). ncRNAs também são classificados em “housekeeping”, que inclui os RNAs transportadores (tRNAs) e RNAs ribossomais (rRNAs) e regulatórios, cuja expressão varia de acordo com as condições celulares (Inamura 2017). Os ncRNAs regulatórios que possuem até 200 nucleotídeos de tamanho são considerados ncRNAs pequenos (sncRNA) e os ncRNAs maiores que 200 nucleotídeos são longos ncRNAs (lncRNA) (Han Li e Chen 2015).

Atualmente, existem diversas famílias de sncRNAs em mamíferos, sendo algumas delas: microRNAs (miRNAs), RNAs de interferência pequenos (siRNAs), RNAs que interagem com proteínas PIWI (piRNAs), RNAs nucleolares pequenos (snoRNAs), RNA ribossomal (rRNA) e RNA transportador (tRNA) (Figura 1.1) (Le Thomas et al. 2014). As diferentes famílias de sncRNAs possuem diferentes funções que compreendem de componentes estruturais (por exemplo rRNAs), à modificação de outros RNAs (por exemplo, snoRNAs) e regulação da transcrição (por exemplo, miRNAs, siRNAs e tRNA) (Griffiths-Jones 2004). Recentemente, diferentes grupos de pesquisa têm demonstrado a importância dos sncRNAs na regulação da expressão gênica em vias metabólicas fundamentais tais como tradução, replicação, diferenciação e apoptose. Estas moléculas também são fundamentais no desenvolvimento e manutenção de diferentes doenças, tais como câncer, Alzheimer e doenças cardiovasculares (Taft et al. 2010).



**Figura 1.1. Famílias de ncRNAs.**

Recentemente, sncRNAs têm sido considerados como candidatos para biomarcadores para diagnóstico e acompanhamento do tratamento de câncer (Gyoba et al. 2016; Joyce et al. 2016; Letelier et al. 2016; Saus et al. 2016). Por exemplo, Xue e colaboradores (2014) inibiram o crescimento tumoral de

adenocarcinoma de pulmão e aumentaram a taxa de apoptose, *in vitro* e *in vivo*, por meio da inserção do miRNA miR-34a e um siRNA capaz de reduzir a expressão de *KRAS*. Gyoba e colaboradores (2016) revisaram o uso de miRNAs como biomarcadores para diagnóstico de câncer de pulmão e Esteller (2011) revisou o papel de sncRNAs em diferentes doenças.

Nesta tese, trabalhamos com três famílias de ncRNAs, sendo elas: miRNAs, piRNAs e snoRNAs.

### 1.2.1.1. *MicroRNAs (miRNAs)*

Uma das famílias de sncRNAs mais estudadas são os miRNAs. Esta família é composta por sncRNAs de tamanho médio de 22 nucleotídeos (Lund 2004). As moléculas desta família são capazes de inibir a tradução de mRNAs com sequência complementar através da união do miRNA com o complexo proteico RISC e o pareamento do complexo ribonucleico com a região 3' UTR do mRNA alvo (Lund 2004).

A região do miRNA responsável pelo reconhecimento dos alvos é chamada de “seed” e composta por seis nucleotídeos situados entre a 2ª e 8ª posição a partir do terminal 5' do miRNA (Figura 1.2). Esta região pode ser aumentada quando há uma complementaridade com o 8º nucleotídeo ou há a presença de uma adenosina no primeiro nucleotídeo. Alguns autores sugerem que estas características aumentam a eficiência da repressão da expressão do mRNA (Grimson et al. 2007). Outra característica desta região é que ela apresenta conservação entre diferentes espécies, sugerindo a atuação de pressões seletivas negativas (Lewis et al. 2003).

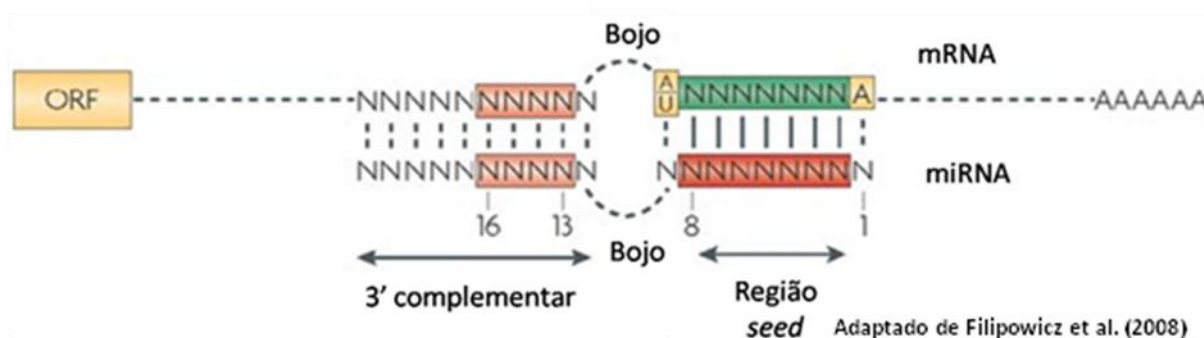
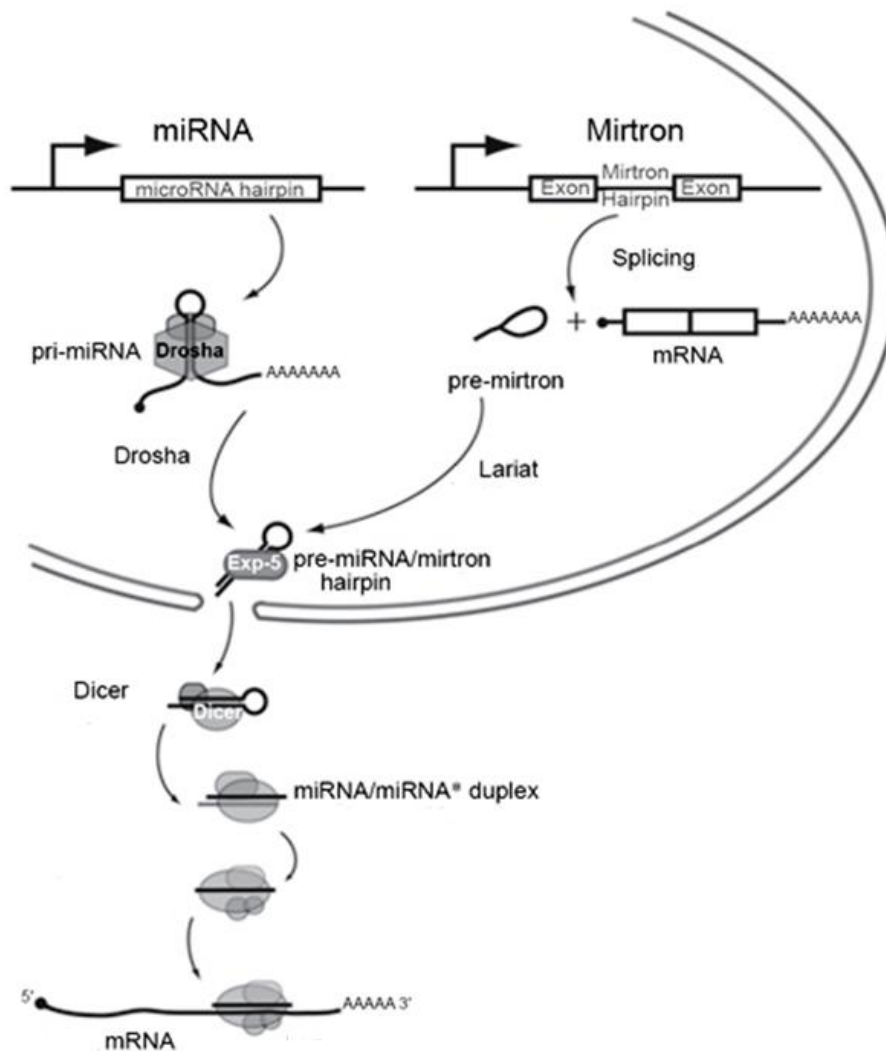


Figura 1.2. Região “seed” do miRNA.

Os miRNAs aparecem codificados no genoma em íntrons de genes codificadores ou em genes próprios (Lee et al. 2004). Estas regiões são transcritas em miRNAs primários que são clivados no núcleo pela enzima Drosha em um transcrito no formato de “hairpin” chamado miRNA precursor (pré-miRNA). O pré-miRNA é exportado para o citoplasma pela enzima exportina 5, onde é clivado em uma fita dupla de miRNA pela enzima Dicer. As fitas de miRNA são separadas por helicases e se unem a proteína Argonata 2 (AGO2) com o complexo de silenciamento induzido por RNA (RISC) para reconhecerem o mRNA alvo através de pareamento incompleto entre o miRNA e o mRNA alvo (Lund 2004) (Figura 1.3).

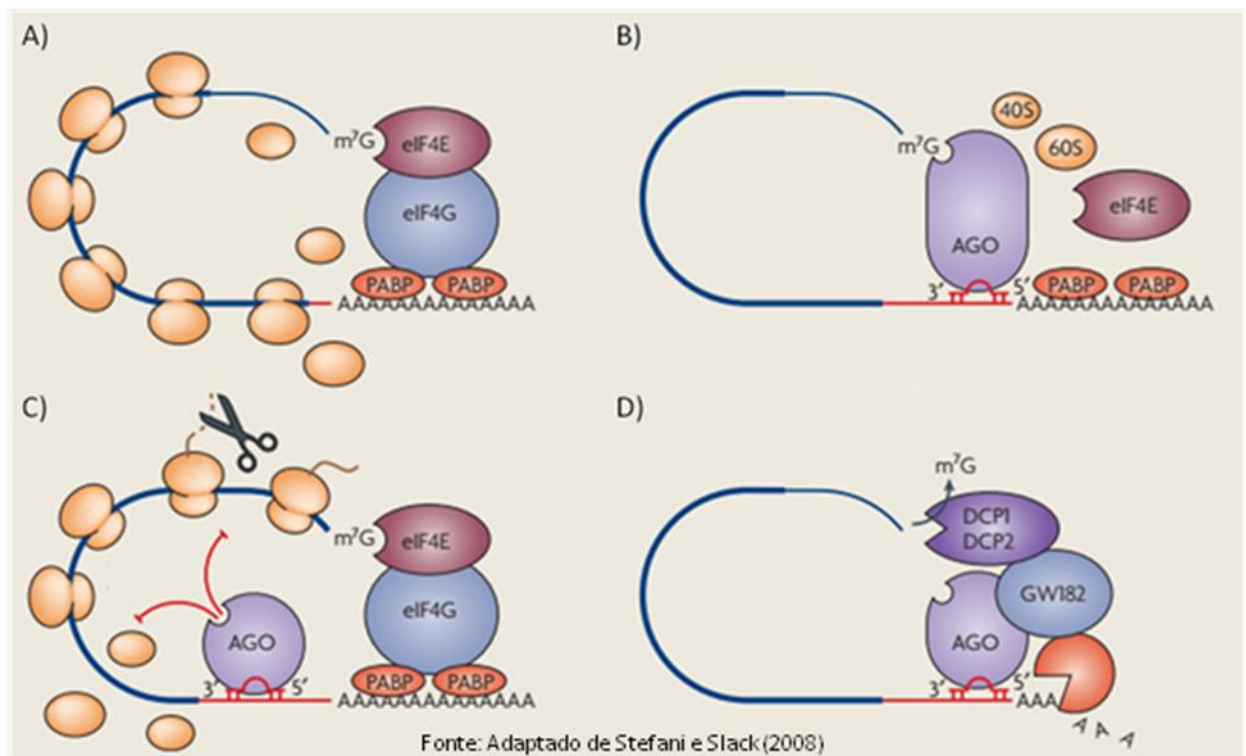


Fonte: Liu et al. 2008

Figura 1.3. Biossíntese de miRNA.



Existem três hipóteses sobre o mecanismo pelo qual os miRNAs interferem na tradução: inibição do início da tradução, promoção do término prematuro da tradução ou desestabilização do mRNA alvo (Li e Rana 2014) (Figura 1.4). Na primeira hipótese, a ligação do complexo miRNA-RISC com o mRNA alvo impede o reconhecimento do fator de iniciação 4F (eIF4F), recrutamento da subunidade 40S ou acoplação da subunidade 60S. Na segunda hipótese, os ribossomos se desacoplam prematuramente e os peptídeos iniciais são degradados. A terceira hipótese envolve o recrutamento de enzimas que removem a proteção 5' ou a cauda de poli-adeninas (poli(A)). Alguns alvos são transportados para os corpos-P<sup>1</sup>, onde são degradados ou armazenados e, dependendo das condições celulares, podem voltar a ser traduzidos (Li e Rana 2014).



**Figura 1.4. Mecanismos de atuação dos miRNAs.** A) Tradução normal. B) Inibição do início da tradução. C) Término prematuro da tradução. D) Recrutamento de enzimas que removem o 5' CAP ou a cauda poli A.

<sup>1</sup> Corpos-P: corpos-P são estruturas presentes no citosol que contém proteínas e enzimas envolvidas na degradação de mRNAs (Li e Rana 2014).

Os mRNAs controlados por miRNAs participam de processos biológicos fundamentais, tais como controle da tradução, transcrição, replicação e apoptose, incluindo genes supressores tumorais e oncogenes (Lund 2004). Assim, diversos miRNAs apresentam alterações no seu perfil de expressão em diversas doenças, por exemplo, os miRNAs miR-15a e miR16-1 estão freqüentemente deletados em leucemia crônica (Calin et al. 2002) e diversos miRNAs da família let-7 atuam inibindo o gene RAS, atuando como um gene supressor tumoral (Johnson et al. 2005).

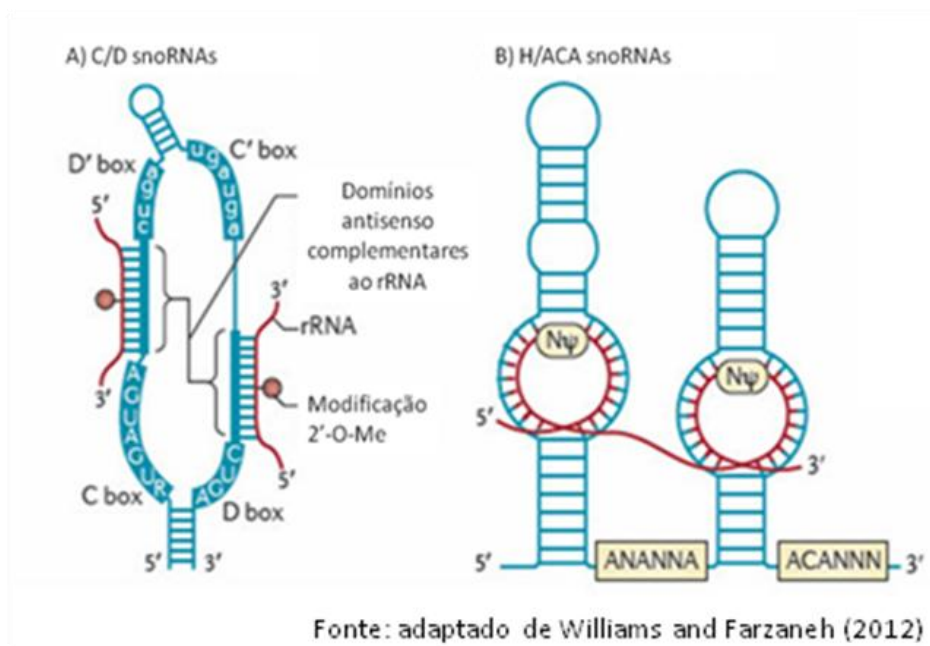
Os aspectos de pré-processamento e análise de dados de seqüenciamento de miRNA usando plataformas de seqüenciamento de alta vazão foram discutidos pelo nosso grupo em duas oportunidades (Jorge et al., 2012; Passetti et al., 2014) e serão apresentadas e discutidas ao longo deste texto.

### **1.2.1.2. RNAs nucleolares pequenos (snoRNAs)**

Uma das famílias de sncRNAs mais populosas é a dos snoRNAs e inclui moléculas desde 60 até 300 nucleotídeos capazes de se associarem à enzimas específicas para promover a metilação e pseudo-uridilação de rRNAs (Thorenor e Slaby 2015). snoRNAs podem ser divididos em dois subtipos, de acordo com o motivo presente na molécula: o motivo H/ACA (SNORA) guia a pseudo uridilação e o motivo C/D box (SNORD) guia a metilação (Thorenor e Slaby 2015) (Figura 1.5). Em 2002, um snoRNA com ambos os domínios C/D e H/ACA, que se acumula nos corpos de Cajal<sup>2</sup>, foi descrito (Darzacq 2002). Estes snoRNAs são conhecidos como SCARNA (Thorenor e Slaby 2015).

---

<sup>2</sup> Corpos de Cajal são suborganelas presentes no núcleo de células transcricionalmente ativas de plantas, fungos e animais (Gall 2003).



**Figura 1.5. Motivos de snoRNA.**

Em vertebrados, a maioria dos snoRNAs está codificada em regiões intrônicas de genes codificadores ou de RNAs não codificadores longos (lncRNAs) (Mannoor et al. 2012) (Figura 1.6), as exceções são os snoRNAs SNORD3, SNORD8, SNORD13 e SNORD118, que possuem promotores próprios (Falaleeva et al. 2017) e são transcritos pela RNA polimerase II (Falaleeva e Stamm 2013). As extremidades 5' e 3' do íntron não relacionadas ao snoRNA são removidas pelas exonucleases XRN1/2 para formação do snoRNA maduro (Falaleeva et al. 2017).

Após ser transcrito e processado, os snoRNAs podem permanecer no núcleo ou serem direcionados a outros compartimentos celulares onde dão origem outros sncRNAs com diferentes funções (Falaleeva e Stamm 2013). Assim, enquanto os snoRNAs presentes no núcleo participam do processo de “splicing” (Falaleeva et al. 2017), no citoplasma eles podem ser clivados em moléculas de RNA menores (Ender et al. 2008; Stepanov et al. 2015) tais como miRNAs, piRNAs e snRNAs (Falaleeva et al. 2017) e exercer a função destas classes de snRNA ou atuar na metilação de rRNAs. Contudo, alguns snoRNAs não possuem alvos identificados, sugerindo que possam existir outras funções para estas moléculas (Stepanov et al. 2015). Um exemplo foi recentemente relatado por Huang e colaboradores (2017), que mostraram que o SNORD50A inibe o processamento da região 3' de mRNAs ao se ligar em sítios de poliadenilação, impedindo a atuação da proteína Fip1, um

componente do complexo de clivagem e poli-adenilação. Vale ressaltar que a inibição dos níveis de SNORD50A alteraram o perfil de poliadenilação e os níveis de diferentes transcritos.

Alterações no padrão de expressão de snoRNAs foram relatadas em diversas doenças. Os snoRNAs HBII-52 e SNORD116 estão frequentemente deletados na síndrome de Pradder-Willis (Kishore e Stamm 2006; Purtell et al. 2017). O snoRNA HBII-52 é necessário para o correto “splicing” do transcrito do receptor de serotonina 2C e a sua ausência leva à formação de um receptor truncado após o terceiro domínio transmembrana (Kishore e Stamm 2006). Diversos snoRNAs também apresentaram padrão de expressão alterado em hipóxia (Liu et al. 2011) e estresse oxidativo (Michel et al. 2011). Em 2008, Dong e colaboradores descreveram uma deleção de dois nucleotídeos no snoRNA U50 em amostras de câncer de próstata, a correção desta deleção reduziu a formação de colônias (Dong et al. 2008). No ano seguinte, o mesmo grupo descreveu a mesma alteração em câncer de mama (Dong et al. 2009). Além disso, foi encontrado um maior número de cópias e expressão elevada do gene de SNORA42 em amostras de tumores de câncer de pulmão de não pequenas células (NSCLC). Os níveis encontrados deste snoRNA estão inversamente relacionados com a sobrevivência do paciente ao câncer (Mei et al. 2012). No sentido oposto, níveis baixos de RNU44 foram associados ao prognóstico ruim em câncer de mama e de cabeça e pescoço (Gee et al. 2011).

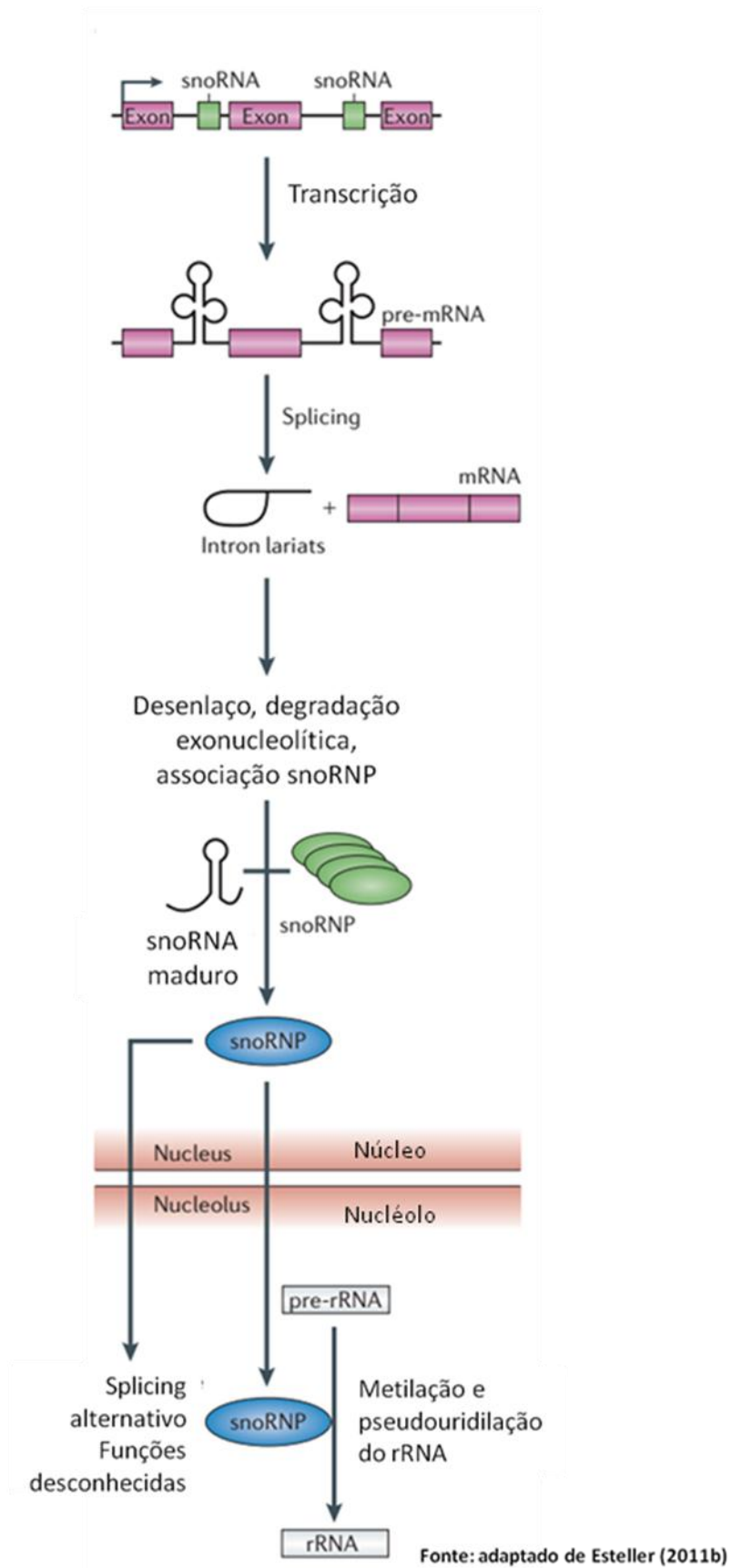


Figura 1.6. Biosíntese snoRNAs.

### **1.2.1.3. RNA que interage com PIWI**

A família de RNAs que interagem com as proteínas PIWI (piRNAs) é uma das menos estudadas. Os membros desta família possuem de 26 a 31 nucleotídeos e muito de sua biogênese e mecanismo de atuação ainda é desconhecido.

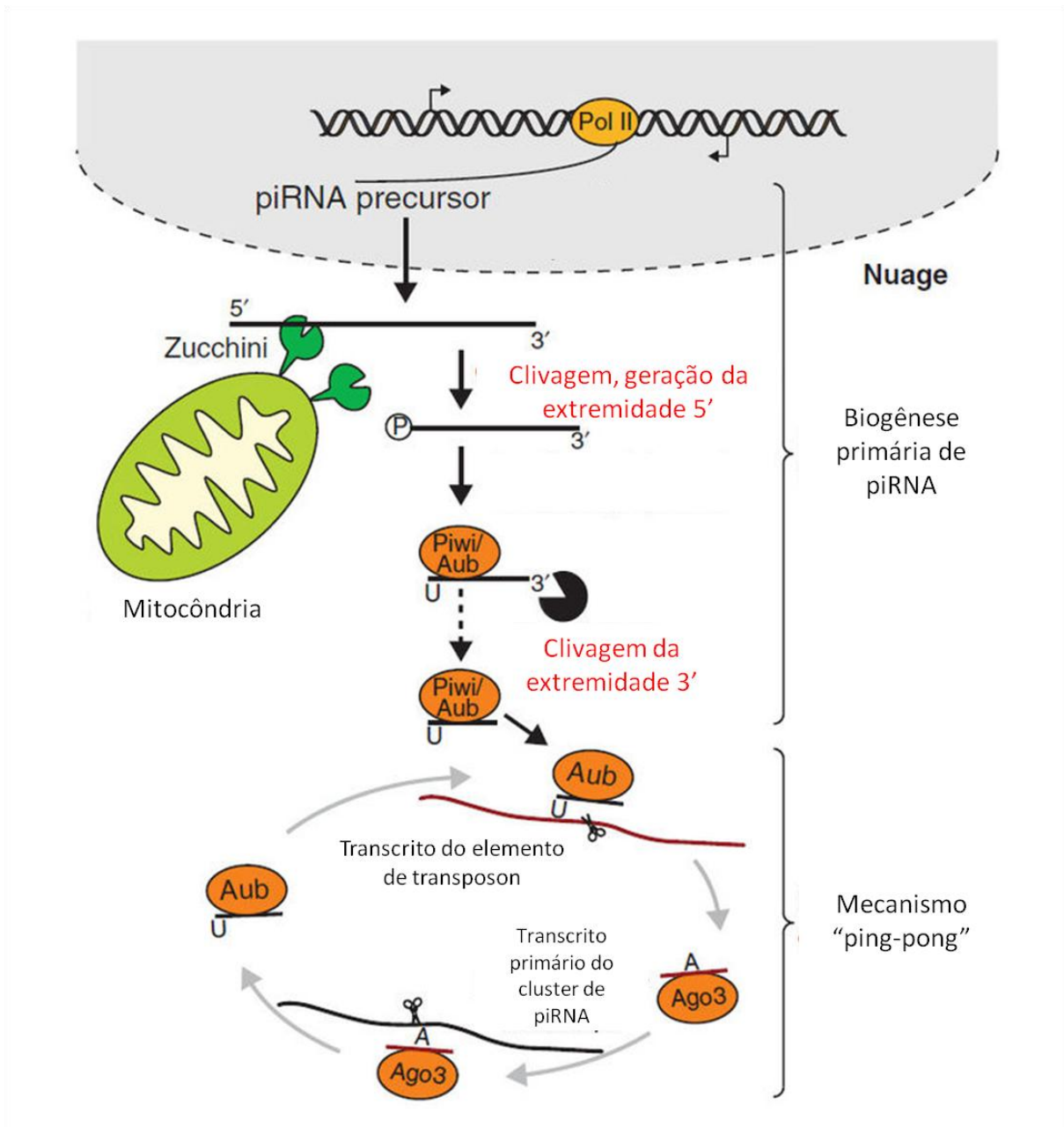
A principal função conhecida dos piRNAs é de silenciar elementos transponíveis em células germinativas (Siomi et al. 2011; Le Thomas et al. 2014), atuando então na divisão celular, apoptose, controle epigenético dos transposons e telômeros e controle da tradução (Busch et al. 2015). Contudo, muitos piRNAs têm sido encontrados em células somáticas (Yan et al. 2011; Yang et al. 2013; Fu et al. 2015), onde atuam no controle da expressão gênica através da indução de modificações em histonas e metilação de DNA (Rajasethupathy et al. 2012; Ross et al. 2014). O complexo piRNA-proteína PIWI reconhece o mRNA alvo a partir do pareamento entre o piRNA e a região alvo. Em camundongos, a proteína MIWI requer um pareamento quase perfeito para que possa clivar o RNA alvo (Reuter et al. 2011). Por outro lado, experimentos com *C. elegans* identificaram alvos com pareamento imperfeito (Lee et al. 2012).

A forma na qual ocorre a interrupção da expressão gênica também difere. Em camundongos, enquanto a proteína MIWI é capaz de clivar o RNA alvo (Reuter et al. 2011), as proteínas MILI e MIWI2 induzem a metilação do DNA na região alvo do piRNA (Aravin et al. 2008). As proteínas PIWI observadas em *C. elegans* também promovem a metilação do DNA, porém recrutam RNAs de interferência pequenos endógenos para silenciar o alvo (Lee et al. 2012). Outro aspecto intrigante dos piRNAs é a ausência de padrão na sequência primária ou de um motivo estrutural conservado, levantando hipóteses sobre como estas moléculas são reconhecidas para transcrição e processamento (Le Thomas et al. 2014).

A maioria dos piRNAs está codificada em grupos no genoma humano (Siomi et al. 2011; Le Thomas et al. 2014), porém alguns foram mapeados em regiões 3' UTR de genes codificadores (Le Thomas et al. 2014). Apesar do mecanismo de biogênese não ser completamente conhecido, acredita-se que os transcritos precursores longos sejam processados pela nuclease Zucchini a fim de gerar o resíduo 5' fosfato, presente na maioria dos piRNAs maduros. Esta molécula

intermediária se uniria às proteínas PIWI e sua extremidade 3' seria clivada por uma enzima ainda desconhecida, mas relacionada à proteína PIWI que o piRNA está inserido (Le Thomas et al. 2014). Outra hipótese acrescenta que piRNAs maduros são capazes de originar outros piRNAs maduros através de um mecanismo denominado ping-pong (Le Thomas et al. 2014). Neste mecanismo, o complexo piRNA-proteína PIWI reconheceria a região complementar ao piRNA no transcrito precursor longo e se ligaria a esta região. Endonucleases ainda desconhecidas clivariam as extremidades 5' e 3' não pareadas do transcrito longo, gerando uma nova molécula de piRNA. Por sua vez, esta nova molécula de piRNA seria capaz de se unir a outra proteína PIWI, reiniciando o processo. A Figura 1.7 mostra um esquema sobre a biossíntese de piRNAs.

Diversos estudos têm relatado padrões alterados de expressão de piRNAs em diferentes tipos de câncer (Cheng et al. 2011; Busch et al. 2015; Fu et al. 2015; Müller et al. 2015). Por exemplo, Cheng e colaboradores (Cheng et al. 2011) detectaram níveis alterados de piR-665 em câncer de estômago, colón, pulmão e mama. Esses autores também relatam a diminuição no crescimento tumoral após a inibição do piRNA, sugerindo uma possível atuação deste piRNA no desenvolvimento tumoral. Busch e colaboradores (Busch et al. 2015) associaram níveis elevados de piR-57125 com a recorrência e sobrevivência de carcinoma renal de células claras metastáticas. Recentemente, Peng e colaboradores (2016) relataram uma associação entre baixos níveis de piR-55490 com sobrevida de paciente com carcinoma de pulmão. No mesmo trabalho, os autores mostram que a restauração aos níveis normais resultou em redução da taxa de proliferação tumoral. Outro exemplo que podemos citar é o estudo de Fu e colaboradores (2015), no qual os autores investigaram a influência de um polimorfismo no piR-021285 em linhagens MCF7 de câncer de mama. Os autores encontraram alterações significativas no padrão de metilação de diversos genes, incluindo piRNAs, relacionados a câncer de mama.



Fonte: adaptado de Le Thomas et al. 2014

Figura 1.7. Biogênese de piRNAs.

### 1.3. Sequenciamento de alta vazão (HTS)

A abordagem que mais foi empregada para o estudo em larga escala do perfil de expressão gênica foi a técnica de "microarray". Após análises por bioinformática, alguns candidatos são selecionados para validação usando técnicas mais



específicas, tais como PCR em tempo real (RT-qPCR) (Tainsky 2009). Porém, esta abordagem requer o uso de moléculas de sondas ou “primers”, indicando que o pesquisador deve possuir algum conhecimento *a priori* do que está sendo analisado e limitando a identificação de novos genes.

O sequenciamento do genoma humano estimulou o desenvolvimento de outras técnicas de sequenciamento que evoluíram de uma única molécula sequenciada por vez, nos sequenciamentos de Maxam-Gilbert e Sanger nos anos 70, até o sequenciamento simultâneo de várias moléculas diferentes pelo sequenciamento de alta vazão (HTS) (Aziz et al. 2015). Várias técnicas para o sequenciamento em larga escala foram desenvolvidas (Roy et al. 2016). Cada tecnologia possui diferentes especificações, tais como vazão, tamanho da sequência gerada, frequência de erros, custo por base e número de sequências geradas (Zhou et al. 2011).

Recentemente, as tecnologias de HTS têm sido usadas para estudar a expressão de diferentes genes devido a sua alta precisão (Zhou et al. 2011) e sensibilidade, permitindo estimar níveis de expressão dos genes presentes nas amostras, detectar rearranjos cromossômicos, mutações, novos transcritos, novas isoformas, genes pouco expressos e identificar novos genes (Meyerson et al. 2010; Jorge et al. 2012). Por exemplo, Müller e colaboradores (2015) utilizaram HTS para avaliar a expressão de genes codificadores e não codificadores em seis amostras de câncer pancreático. Além de diversos genes alterados, os autores encontraram os snoRNAs HBII-296B e U104 e o piRNA piR-017061 diferencialmente expressos em tumores quando comparados a células normais. Outro exemplo é o trabalho de Witten e colaboradores (2010) nos quais foram avaliados o perfil de expressão de miRNAs em 29 amostras de tecido normal e câncer cervical das mesmas pacientes. Os autores relatam diversos genes diferencialmente expressos entre as amostras normais e tumorais e identificaram 64 novos miRNAs.

#### **1.4. Consórcios de Sequenciamento**

O grande sucesso do consórcio internacional do Projeto Genoma Humano aliado ao maior acesso à tecnologia de HTS permitiu que diversos centros de pesquisa se unissem para a criação de diferentes consórcios. Estes grupos

possuem objetivos diversos, que vão desde sequenciar e comparar mais e 2000 genomas de indivíduos de diferentes etnias, tais como o Projeto 1000 Genomas (Auton et al. 2015) e avaliar a regulação da expressão gênica de diferentes tecidos como o GTEX (Lonsdale et al. 2013).

Existem inúmeras vantagens na criação de um consórcio. A maior quantidade de institutos envolvidos em um projeto claramente se traduz em um maior número de amostras e menores gastos, dado que o custo e mão de obra relacionados ao processamento são divididos entre os centros envolvidos (Rung e Brazma 2012). Além disso, todos os institutos envolvidos aplicam os mesmos mecanismos de controle de qualidade e algoritmos de processamento dos dados, garantindo confiabilidade e uniformidade (Rung e Brazma 2012).

A integração dos dados genômicos, epigenômicos e fenotípicos disponibilizados pelos diversos consórcios, auxilia na melhor compreensão dos mecanismos de diferentes patologias e na personalização de tratamentos (Apellaniz-Ruiz et al. 2016). Neste trabalho, utilizamos dados públicos de dois consórcios: “The Cancer Genome Atlas” (TCGA) (Tomczak et al. 2015) e “BLUEPRINT” (Abbott 2011).

#### **1.4.1. O Projeto TCGA**

O projeto *The Cancer Genome Atlas* é uma colaboração entre o *National Cancer Institute (NCI)* e o *National Human Genome Research Institute (NHGRI)* cujo objetivo é produzir e armazenar mapas de perfil de expressão e alterações genômicas e transcriptômicas de mais de 33 tipos de câncer (Tomczak et al. 2015). Este consórcio envolve diferentes centros de pesquisa responsáveis pela coleta das amostras, processamento, sequenciamento, análise de bioinformática padronizada, armazenamento e disponibilidade dos dados brutos e processados.

Este banco de dados é público, porém de acesso controlado, e contém dados de HTS do genoma, transcriptoma e miRNAs de tumores, tecidos normais adjacentes ao tumor e sangue. Os dados processados incluem informações sobre a expressão de genes, transcritos e miRNAs, variações de nucleotídeo único encontradas, variações do número de cópias, perdas de heterozigidade, mutações, metilação de DNA e expressão de proteínas (Tomczak et al. 2015)

Neste trabalho, utilizamos amostras de HTS de RNAs pequenos de tecido tumoral e normal adjacente de pacientes de adenocarcinoma de pulmão (Collisson et al. 2014).

#### **1.4.1.1. Adenocarcinoma de pulmão**

O câncer é a maior causa de mortes no mundo (Larrea et al. 2016) e o câncer de pulmão é uma das maiores causas de óbito relacionadas a câncer, em homens e mulheres, no mundo (Saalberg e Wolff 2016). No Brasil, o câncer de pulmão é o segundo mais frequente em homens e o quarto mais frequente em mulheres (INCA, 2016).

O câncer de pulmão é dividido em dois grupos: o câncer de células pequenas e o de células não pequenas (NSCLC) (Inamura 2017), sendo este o tipo mais comum totalizando quase 85% dos casos (Mendes et al. 2016). O NSCLC também pode ser dividido em adenocarcinoma, carcinoma escamoso e carcinoma de células claras (Inamura 2017).

Devido à ausência de sintomas em estágios iniciais e de métodos de diagnóstico específicos, o câncer de pulmão é frequentemente detectado quando já se encontra em estado clínico avançado (Kunz et al. 2016), adquirindo, portanto, um prognóstico desfavorável e alta taxa de mortalidade (Herbst et al. 2008; Gyoba et al. 2016; Li et al. 2016). Estima-se que a taxa de sobrevida após 5 anos é de 5% (Kunz et al. 2016).

O fumo é um dos principais fatores de risco para câncer de pulmão (Gumus et al. 2008). Já foram documentados que os diversos carcinógenos presentes no fumo são capazes de se ligarem ao DNA e gerar mutações somáticas, tais como as observadas nos genes *KRAS* em câncer de pulmão e no gene *TP53* em diversos tipos de câncer (Bialous e Sarna 2017). Estudos demonstraram as alterações no perfil de expressão de diversos genes provocados por fumaça de cigarro nas linhagens celulares MSK-Leuk1 (Gumus et al. 2008), células epiteliais (Spira et al. 2004) e mucosa oral (Cao et al. 2015). Além disso, foram relatadas também alterações nas vias metabólicas de transdução de sinal, asma e proliferação celular (Gumus et al. 2008; Wang e Li 2010; Minicã et al. 2017). Contudo, a maioria dos fumantes não desenvolve câncer de pulmão e 20% das mulheres com esta doença

são não fumantes (Vavalà et al. 2016). Assim, outros fatores além do fumo podem contribuir para o desenvolvimento da doença. (Huang et al. 2015).

Diversos grupos de pesquisa apontaram alterações no padrão de expressão de diversos sncRNAs em câncer de pulmão. Os miRNAs miR-21, let-7, miR-145 e miR-155 foram apontados como biomarcadores para o diagnóstico precoce de câncer de pulmão (Kunz et al. 2016). Assim como os snoRNAs SNORA78, SNORA21, SNORD28 e SNORD66 foram associados com a sobrevida dos pacientes (Gao et al. 2015) e o piRNA piR-57125 e os miRNAs miR-30a-3p e miR-210-3p com metástases à distância (Daugaard et al. 2017). Além disso, a expressão aumentada de SNORD78 foi apontada como necessária para proliferação e invasão a outros tecidos (Zheng et al. 2015).

O fumo também foi apontado como um fator que altera os níveis de expressão de sncRNAs em câncer de pulmão. Os miRNAs miR-182, miR-183, miR-210 e miR-126 foram indicados para distinguir indivíduos fumantes e pacientes com NSCLC (Zhu et al. 2016). Recentemente, Wang e colaboradores (2017) relataram 34 miRNAs cujas funções estão relacionadas com o desenvolvimento do pulmão, diferenciação das vias aéreas inflamação e câncer, foram relatados alterados entre fumantes e não fumantes. Doze destes 34 miRNAs permaneceram alterados mesmo após a cessação de fumo por três meses. Krishnan e colaboradores (2017) relataram 13 piRNAs alterados entre pacientes fumantes e não fumantes de câncer de cabeça e pescoço.

#### **1.4.2. O Projeto BLUEPRINT**

O projeto BLUEPRINT é outro exemplo de projeto que seguramente foi possível ser desenvolvido a partir dos desenvolvimentos tecnológicos oriundos do sequenciamento do genoma humano. Este projeto é uma iniciativa patrocinada pela União Europeia cujo objetivo é gerar, no mínimo, 100 mapas epigenéticos de referência de células sanguíneas de indivíduos saudáveis e pacientes diagnosticados com leucemia (Abbott 2011).

Este projeto foi iniciado em 2011 e, somente dois anos após o seu início, publicou 12 epigenomas completos de neutrófilos e monócitos de amostras de sangue periférico de adultos e de cordão umbilical. As informações obtidas incluem modificações de histonas, bases nucleotídicas metiladas, expressão de RNA

mensageiro e acessibilidade do genoma (Martens e Stunnenberg 2013). A Figura 1.8 mostra o processo de diferenciação celular da árvore hematopoiética.

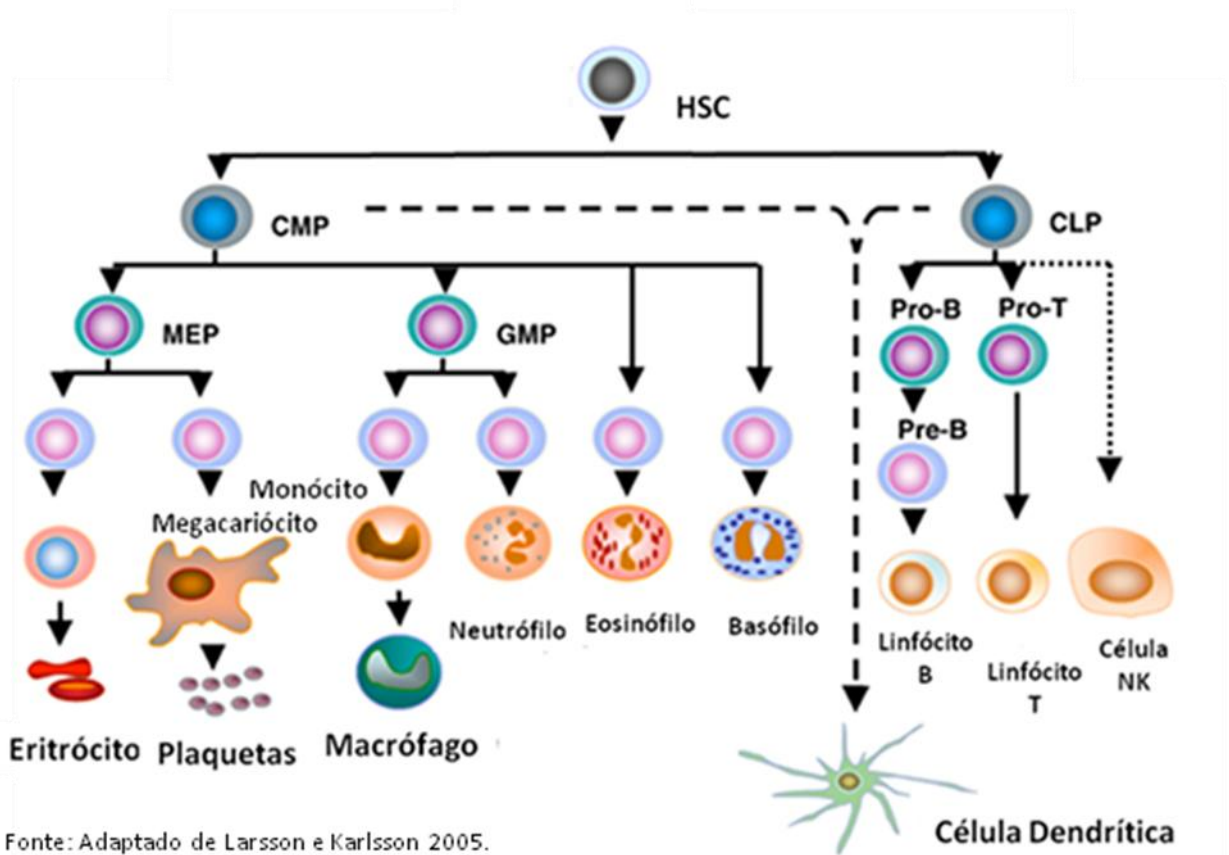


Figura 1.8. Árvore Hematopoiética.

### 1.4.2.1. Linfócitos T CD4+

Linfócitos T CD4+, ou linfócito T auxiliar, são leucócitos que participam da imunidade mediada por células. Estas células se diferenciam a partir de linfoblastos presentes na medula óssea sofrem maturação no timo, onde começam a expressar os receptores CD4 e TCR que caracterizam este tipo celular.

A principal função dos linfócitos auxiliares é a produção de citocinas e interleucinas que coordenam a resposta à vírus, bactérias ou fungos. Estas células são ativadas ao reconhecerem receptores peptídicos ligados à moléculas de MHC de células apresentadoras de antígenos, tais como macrófagos, células dendríticas e linfócitos B (Delamarre 2005). Uma vez estimulados, os linfócitos secretam a citocina IL-2, proliferam e se diferenciam em subtipos que são determinados pelas

interleucinas produzidas. Os linfócitos T CD4+ do tipo 1 produzem interferon gama e IL-12 que estimularão linfócitos T CD8+ e macrófagos a destruírem células infectadas, os do tipo 2 produzem IL-4, IL-5, IL-6 e IL-13 para estimular a produção de imunoglobulina E por linfócitos B contra helmintos, os do tipo três suprimem a resposta imune e os do tipo 17 secretam IL-6 e IL-17 a fim de gerar uma resposta inflamatória contra bactérias extracelulares (Abbas et al. 2015).

Estudos constataram que a expressão miRNAs são fundamentais para o processo de diferenciação de linfócitos. Por exemplo, o miR-155 e miR-125b estão expressos em células tronco hematopoiéticas (HSC) que irão se diferenciar em células precursoras linfóides comum (CLP) (Utikal et al. 2015). O grupo de miRNA miR-17-92 também foi encontrado mais expresso em células precursoras que originarão linfócitos T (Baltimore et al. 2008; Sonkoly et al. 2008; Fiedler e Brunner 2012), enquanto que o miR-181 foi relatado menos expresso no mesmo precursor (Baltimore et al. 2008; Sonkoly et al. 2008; Fiedler e Brunner 2012). O linfócito T também expressa diversos miRNAs, tais como o miR-146, miR-155, miR-150, e miR142-3p (Baltimore et al. 2008; Sonkoly et al. 2008; O'Connell et al. 2011; Sayed e Abdellatif 2011).

#### **1.4.2.2. Megacariócitos**

Megacariócitos são células gigantes da medula óssea responsáveis pela produção de plaquetas. Elas são derivadas de células tronco hematopoiéticas que foram estimuladas pela trombopoietina a se diferenciarem em megacariócitos. Durante o processo de maturação, estas células aumentam o seu volume através de replicações endomitóticas, nas quais o núcleo se multiplica, mas não há divisão celular (Rabellino 1981).

As células maduras iniciam então a formação de processos proto-plaquetários, que são membranas internas que armazenam as plaquetas até o momento de liberação por explosão ou por pseudópodes que realizam a liberação contínua de plaquetas (Italiano et al. 1999).

miRNAs também foram relacionados com a diferenciação e atuação de megacariócitos. Os miRNAs miR-29a e miR-9 foram encontrados em HSC que se diferenciarão em células precursoras mieloides comum (CMP) e miR-146b e miR-

150 em células que se diferenciarão em precursores de megacariócitos ou eritroblatos (MEP) (Utikal et al. 2015). Estudos também relatam que para a diferenciação para megacariócitos é necessária a queda na expressão de miR-130 e miR-155 (Baltimore et al. 2008; O'Connell et al. 2011; Fiedler e Brunner 2012) e o aumento de miR-150 (Sayed e Abdellatif 2011; Fiedler e Brunner 2012; He et al. 2014).

### **1.4.2.3. Eritroblatos**

Eritroblatos são células sanguíneas vermelhas que ainda retém o núcleo, sendo precursoras imediatas do eritrócitos, cuja função é transportar oxigênio dos pulmões para os tecidos através da circulação sanguínea. Estas células se diferenciam de células tronco pluripotentes presentes na medula óssea. Neste processo, elas perdem o seu núcleo e a maioria das organelas, adquirindo seu formato bicôncavo característico. O citoplasma dos eritrócitos é rico em hemoglobina, responsável pela ligação ao oxigênio (Pierigè et al. 2008).

É necessária a queda do miRNA miR-150 para que as células precursoras de megacariócito eritrócitos (MEP) se diferenciem em eritroblatos (Fiedler e Brunner 2012; He et al. 2014), além disso, também há o aumento da expressão do miR-451 (Baltimore et al. 2008; O'Connell et al. 2011; Sayed e Abdellatif 2011; Fiedler e Brunner 2012)

### **1.4.2.4. Neutrófilos**

Neutrófilos são as células brancas fagocíticas encontradas com maior frequência na circulação sanguínea, seu núcleo possui de três a cinco lóbulos e seu citoplasma possui grânulos contendo enzimas como lisozima, colagenase, elastase, defensinas e catelicidinas (Abbas et al. 2015). Estas células são as primeiras recrutadas em casos de infecção e são marcadores para infecções agudas (Jacobs et al. 2010)

Os neutrófilos são produzidos na medula óssea a partir do mesmo precursor que produz fagócitos monocleares estimulado pelo fator estimulador de colônias (Abbas et al. 2015).

MiRNAs também estão envolvidos na diferenciação de neutrófilos. A expressão dos miRNAs miR-155 e miR-223 aumenta para auxiliar na diferenciação de células CMP para os precursores granulócitos e monócitos (Utikal et al. 2015). Os miRNAs miR-223 e miR-34 também são mais expressos para a diferenciação final em neutrófilos, enquanto o miR-21 e o miR-196b são reprimidos (O'Connell et al. 2011; Sayed e Abdellatif 2011; Fiedler e Brunner 2012)

#### **1.4.2.5. Monócitos e Macrófagos**

Os monócitos são células fagocíticas mononucleares circulantes diferenciadas a partir de células pluripotentes na medula óssea. As células pluripotentes são estimuladas pelo fator estimulador de colônia de monócito, diferenciando-se nos monócitos maduros que migram para a circulação sanguínea. Os monócitos possuem somente um núcleo e seu citoplasma contém lisossomos, vacúolos fagocíticos e filamentos de actina necessários para fagocitose e produção e armazenamento de espécies reativas de oxigênio e nitrogênio e enzimas de digestão proteolíticas que são tóxicas para os organismos invasores (Abbas et al. 2015).

Durante as reações inflamatórias, os monócitos migram para os tecidos onde se diferenciam em macrófagos. Muitos tecidos são povoados com macrófagos residentes, tais como as células de Kupffer no fígado, macrófagos alveolares nos pulmões e células da micróglia no cérebro (Abbas et al. 2015). De acordo com o estímulo recebido, os macrófagos também se diferenciam em tipo M1 ou M2. Os macrófagos do tipo M1 se diferenciam a partir de citocinas produzidas por linfócitos auxiliares do Tipo 1, ao contrário dos macrófagos M2 que se diferenciam por citocinas produzidas por linfócitos auxiliares do Tipo 2.

Os miRNAs miR-155 e miR-221 tem sua expressão reduzida em monócitos (Baltimore et al. 2008; Lu et al. 2011), enquanto o miRNA miR-146a é mais expresso (Baltimore et al. 2008; O'Connell et al. 2011; Utikal et al. 2015). Os níveis do miR-



146 se mantém altos em macrófagos também, assim como os níveis do miR-155 que é mais expresso nestas células (Sayed e Abdellatif 2011).

## **1.5. Reuso de dados públicos de HTS**

A tecnologia de HTS tem evoluído para a obtenção de cada vez mais dados de sequência com custos cada vez menores. Os dados produzidos por esta tecnologia podem ser armazenados em bancos de dados públicos e disponibilizados gratuitamente para a comunidade científica. Neste sentido, torna-se importante padronizar os formatos nos quais os dados serão disponibilizados, os procedimentos para obtê-los e informações sobre as amostras utilizadas (Chervitz et al. 2011).

Esta padronização facilita tanto a divulgação quanto a interpretação dos dados por outros cientistas (Chervitz et al. 2011). Desta forma, com informações sobre as amostras e formatos acessíveis, diferentes grupos de pesquisa podem recriar o experimento, testar novos algoritmos de análise ou combinar com outros dados para responder outras perguntas biológicas (Perez-Riverol et al. 2015). Ao combinar os dados de diferentes experimentos, após controle de qualidade normalização e avaliação de cada experimento, consegue-se aumentar a probabilidade de detecção de sinais fracos e gerar novo conhecimento, sem o alto custo atrelado, caso somente um grupo fosse realizar tudo (Rung e Brazma 2012).

Rung e Brazma (2012) realizaram uma pesquisa de quantas vezes o banco de dados público ArrayExpress foi mencionado em artigos publicados no ano de 2011, os autores descobriram que quase 25% dos artigos que mencionavam o banco de dados, utilizaram a informação contida nele para responder perguntas biológicas diferentes da questão do estudo original. Por exemplo, Kröger e colaboradores (2016) usaram diversos experimentos públicos de microarray para identificar alterações genéticas em células sanguíneas mononucleares de pacientes com lúpus eritematoso sistêmico, enquanto Gonzalez-Porta e colaboradores (2012) utilizaram dados públicos do HapMap para avaliar diferentes padrões de “splicing” em populações de origem caucasiana e yorubas. Além disso, Cao e colaboradores (2015) usaram dados públicos de microarray para identificar mais de 300 genes diferencialmente expressos entre fumantes e não fumantes.

Assim, é importante também desenvolver metodologias de análise que não estejam relacionadas somente a um método ou fabricante, mas que possam lidar

com dados de diferentes origens e combiná-los (Chervitz et al. 2011). Um grande esforço neste sentido é exemplificado pela iniciativa Bioconductor (<https://www.bioconductor.org/>).

O projeto Bioconductor é uma iniciativa para a distribuição e criação colaborativa de “open source” software para análises de dados em bioinformática e biologia computacional (Gentleman et al. 2004; Huber et al. 2015). O objetivo é permitir a análise integrada de diferentes tipos de dados de alta vazão, tais como HTS, “microarray”, proteômica, entre outros. Atualmente, esta iniciativa é baseada na linguagem de programação R e possui mais de 934 software capazes de interagirem uns com os outros, mas criados por diferentes cientistas (Huber et al. 2015). Assim, torna-se possível a criação de “pipelines” capazes de combinarem diferentes tipos de dados e ferramentas que auxiliam em análises exploratórias e estatísticas e na visualização das informações obtidas (Huber et al. 2015).

O projeto genoma humano causou um grande impacto na ciência atual, não somente pelos seus achados primários, mas também pelo estímulo a criação de novos consórcios e pelas tecnologias desenvolvidas a partir dele. A grande quantidade de dados obtida com as metodologias atuais de alta vazão é disponibilizada à comunidade científica que desenvolve metodologias capazes de integrar estas informações e auxiliar na compreensão de diferentes fenômenos biológicos. Este trabalho utiliza dados e metodologias públicos e livres para avaliar o perfil de expressão de snoRNAs, piRNAs e miRNAs em diferentes amostras humanas.

## 2. JUSTIFICATIVA

De acordo com o Instituto de câncer norte americano, biomarcadores são “moléculas biológicas encontradas em sangue, fluidos ou tecidos que são sinais de processos normais, anormais condições específicas ou doenças” (National Cancer Institute 2016). Os biomarcadores podem ser divididos em três categorias: diagnóstico, prognóstico ou predição. Desta forma, um biomarcador ideal deve ter alta sensibilidade, especificidade e poder preditivo (Larrea et al. 2016).

O sequenciamento de alta vazão é uma tecnologia que vem sendo cada vez mais aplicada no estudo do transcriptoma de diferentes tipos celulares devido às vantagens de se poder identificar e quantificar transcritos novos e já anotados presentes em uma amostra. Recentemente, esta tecnologia tem evoluído para uma detecção mais sensível, rápida e barata das informações presentes nas amostras, de forma que, muitos grupos têm pensado na sua aplicação na rotina clínica.

O acúmulo de dados de HTS excede nossa capacidade tecnológica de controlar e interpretar a informação (Roy et al. 2016). Dessa forma, o reuso de dados em bancos públicos é estimulado e diferentes grupos de pesquisa podem utilizar os dados gerados por outros centros para responder novas perguntas biológicas com um conjunto robusto de amostras.

O perfil de expressão de diferentes sncRNAs é específico do tecido onde são expressos e apresenta alterações em diferentes condições patológicas, tais como diferentes tipos de câncer e em diferentes estágios de desenvolvimento do tumoral (Larrea et al. 2016).

Desta forma, uma melhor compreensão dos perfis de expressão de sncRNAs através do estudo de dados de HTS produzidos por diferentes institutos de pesquisa pode contribuir para o melhor entendimento do metabolismo de diferentes tipos celulares e estados patológicos.

Neste trabalho, utilizamos dados de sequenciamento de alta vazão para avaliar e comparar o perfil de expressão de RNAs não codificadores pequenos em amostras de pacientes fumantes e não fumantes de adenocarcinoma de pulmão e em diferentes células sanguíneas.

### **3. OBJETIVOS**

#### **3.1. Objetivo Geral**

Analisar comparativamente o perfil de expressão de RNAs não codificadores pequenos de amostras de pacientes fumantes e não fumantes com câncer de pulmão, bem como entre diferentes tipos celulares do sangue humano.

#### **3.2. Objetivos Especificos**

##### **3.2.1. *Perfil de expressão de RNAs não codificadores pequenos em pacientes fumantes e não fumantes de adenocarcinoma de pulmão.***

- Obter o perfil de expressão de miRNAs, snoRNAs e piRNAs de amostras normais e tumorais de pacientes não fumantes com adenocarcinoma de pulmão.
- Obter o perfil de expressão de miRNAs, snoRNAs e piRNAs de amostras normais e tumorais de pacientes fumantes com adenocarcinoma de pulmão.
- Comparar o perfil de expressão dos snoRNAs e piRNAs encontrados nas amostras de pacientes não fumantes e fumantes

##### **3.2.2. *Perfil de expressão de miRNAs em células sanguíneas***

- Obter o perfil de expressão de miRNAs em amostras de linfócitos T CD4, Megacariócitos, Eritroblatos, Neutrófilos, Monócitos e Macrófagos M1 e Macrófagos M2 do projeto Blueprint.
- Comparar o perfil de expressão dos miRNAs encontrados nas amostras de sangue de voluntários saudáveis do projeto BLUEPRINT.

## 4. MATERIAL E MÉTODOS

Todos os programas e dados foram processados e armazenados no Laboratório de Bioinformática e Biologia Computacional no Instituto Nacional de Câncer (INCA), no Laboratório de Genômica Funcional e Bioinformática do Instituto Oswaldo Cruz da Fundação Oswaldo Cruz (IOC/FIOCRUZ) e na Plataforma de Bioinformática da Fiocruz RPT04A/RJ da Fiocruz.

### 4.1. Perfil de expressão de RNAs não codificadores pequenos em pacientes fumantes e não fumantes de adenocarcinoma de pulmão.

#### 4.1.1. Dados de sequenciamento

Os dados de sequenciamento de RNAs pequenos de mulheres com adenocarcinoma de pulmão foram obtidos através do trabalho de Kim e colaboradores (Kim et al. 2013) e do Banco de dados do TCGA (autorização do comitê de acesso do dbGAP pelo “Project ID 43224-3”).

Kim e colaboradores extraíram o RNA total de amostras pareadas, tecido tumoral e tecido normal adjacente ao tumor, de 6 mulheres coreanas que apresentaram câncer de pulmão e que nunca fumaram, em seguida selecionaram os RNAs pequenos (entre 22 a 30 nucleotídeos) e utilizaram a tecnologia do Genome Analyzer Iix da Illumina para o sequenciamento específico das amostras enriquecidas com pequenos RNAs de 22 a 30 nucleotídeos de tamanho [NCBI GEO:GSE37764]. Os *reads* de 36 nt de tamanho foram utilizados para obtenção do perfil de miRNAs presentes nas amostras. Os autores avaliaram e compararam o perfil de miRNAs entre as amostras normais e tumorais, mas relataram no artigo que outros sncRNAs também poderiam estar presentes.

Diferentes centros de pesquisa utilizaram ou o Illumina Genome Analyzer ou o Illumina HiSeq para sequenciar amostras enriquecidas com pequenos RNAs de 17 a 27 nucleotídeos de tamanhos a partir de RNA total e sequenciar os estes RNAs pequenos de seis mulheres caucasianas fumantes associados ao TCGA (Download: Janeiro/2016).

#### 4.1.2. Pré-processamento e análise

Os dados de Kim e colaboradores (2013) foram obtidos no formato fastq sem qualquer processamento, enquanto que os dados do TCGA foram obtidos no formato BAM após o processamento. Este processamento do TCGA incluiu remoção de sequências adaptadoras, remoção de “reads” menores que 10 nucleotídeos ou maiores que 30 nucleotídeos e alinhamento contra o genoma humano versão hg19 permitindo um total de 1 “mismatch” (Chu et al. 2016).

As etapas de pré-processamento dos dados de Kim e colaboradores (2013) estão descritas na Figura 4.1. Em suma, os adaptadores inseridos durante o sequenciamento foram removidos das sequências dos *reads* utilizando o programa Cutadapt (Martin 2011) (cutadapt -m 15 -a adaptador arquivo fastq). Em seguida, os fragmentos foram filtrados de acordo com sua qualidade utilizando o programa Fastx-tools (Gordon & Hannon, não publicado), mantendo somente aquelas que possuísem mais de 90% de suas bases com valor de qualidade “phred” acima de  $20^3$  (fastq\_quality\_filter -Q33 -q 20 -p 90 -i fastq trimado -o fastq filtrado). As sequências restantes foram alinhadas ao genoma humano usando o programa Novoalign (versão 3.02.13, <http://www.novocraft.com/>), mantendo somente os alinhamentos únicos (novoalign -r None -d hg19\_no\_random\_no\_hap.index -o SAM -f fastq filtrado).

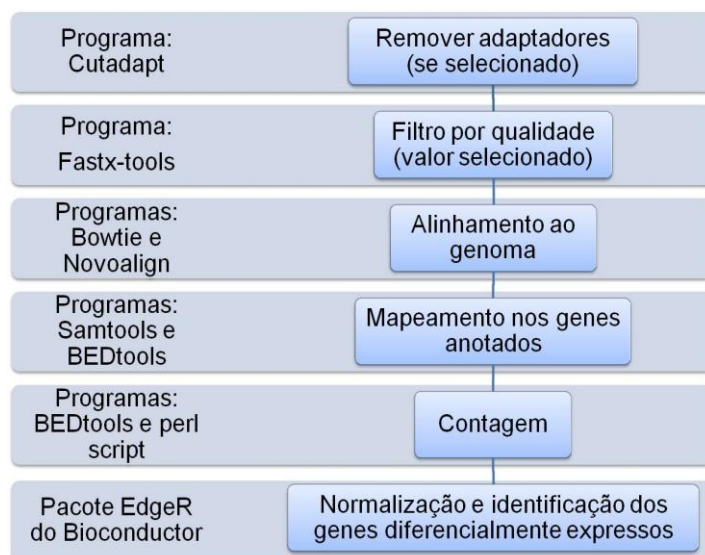


Figura 4.1. Pipeline de processamento das amostras de Adenocarcinoma de pulmão.

<sup>3</sup> O valor de “phred” é calculado a partir da probabilidade de erro da identificação do nucleotídeo. O valor de 20 indica uma probabilidade de erro de 1 base a cada 100 bases, representando precisão de 99%.

A qualidade das sequências no formato BAM oriundas do TCGA foi avaliada utilizando o programa FastQC (Andrews 2010). Os dados do TCGA não possuíam sequências adaptadoras e já haviam passado pelo filtro de qualidade, assim, elas foram convertidas para o formato fastq utilizando o programa BAMtoBED do pacote BEDtools (Quinlan e Hall 2010) e alinhadas ao genoma humano utilizando o programa Novoalign (Hercus 2008) e mantendo somente os alinhamentos únicos.

O arquivo de anotação de piRNAs foi obtido do banco de dados piRNABank (Sai Lakshmi e Agrawal 2008) no formato GTF<sup>4</sup>. Os piRNAs deste arquivo de anotação que se encontravam nas fitas senso ou antisenso em regiões sobrepostas foram unidos em grupos utilizando programas na linguagem Perl. O arquivo de anotação dos snoRNAs foi obtido pelo banco de dados UCSC (Rosenbloom et al. 2015). As anotações de snoRNAs e de piRNAs agrupados foram unidas utilizando o programa merge do pacote BEDtools (Quinlan e Hall 2010). Este arquivo de anotação contendo as informações de piRNA e snoRNA foi comparado com a anotação de miRNAs maduros do miRBase versão 21 (Kozomara e Griffiths-Jones 2011). As regiões que se sobrepuseram às regiões de miRNAs foram consideradas como miRNAs e removidas das análises.

A identificação e quantificação dos snoRNAs ou piRNAs expressos nas amostras normais e tumorais de pacientes não fumantes e fumantes com câncer de pulmão foi realizada utilizando o programa CoverageBED do pacote BEDtools e o arquivo de anotação contendo piRNAs e snoRNAs. As contagens brutas foram normalizadas e os genes diferencialmente expressos foram identificados utilizando o pacote EdgeR versão 3.18.1 (Robinson et al. 2010) do projeto Bioconductor. Somente os sncRNAs com pelo menos 50 “reads” mapeados em pelo menos metade das amostras e contagem de 0 em no máximo duas amostras foram mantidos para cálculo de normalização e avaliação da expressão diferencial. O pacote do bioconductor EdgeR utiliza o método de TMM para normalização e o teste binomial negativo para identificação dos genes diferencialmente expressos. O método de TMM considera que a maioria dos genes não é diferencialmente expresso e utiliza os valores de contagens brutos de cada gene expresso e o total de “reads” produzidos em cada amostra para calcular o fator de correção que será

---

<sup>4</sup> Arquivos no formato GTF contém diversas informações sobre transcritos ou genes de interesse anotados, tais como suas coordenadas cromossômicas e a fita de DNA em que se localiza a anotação de interesse.

aplicado em cada amostra. Foram considerados como diferencialmente expressos, os genes que apresentaram  $FDR < 0,01$  e magnitude da alteração no perfil de expressão ( $\log FC$ ) maior que 2.

Os genes constitutivamente expressos foram obtidos segundo a metodologia descrita por Eisenberg and Levanon (2013), na qual são considerados genes constitutivos todos os genes que apresentaram pelo menos 1 contagem por milhão (cpm) em todas as condições. Os genes cuja expressão normalizada não é duas vezes maior que a média das amostras e o desvio padrão é menor que 1, são considerados possíveis genes *housekeeping*.

## 4.2. Comparação do perfil de miRNAs em células sanguíneas

Foram sequenciadas 29 amostras de diferentes tipos celulares (Tabela 4.1). Os arquivos destes dados estão sob tutela do Dr. Ernest Turro, supervisor do doutorado sanduíche desta estudante. Os arquivos fastq disponibilizados para análise já possuíam boa qualidade e por isso, nenhum filtro de qualidade foi aplicado e os adaptadores foram removidos usando o programa Trim Galore! (Krueger 2012) (`trim_galore -f fastq -e 0.05 -q 15 -O 3 -a adaptador`). Os *reads* foram alinhados às sequências de microRNAs maduros presentes no banco de dados miRBase versão 21 (Kozomara e Griffiths-Jones 2011) utilizando o programa RapMap (Srivastava et al. 2016) (`rapmap quasimap -c -s -z 0.9 -i indice -r fastq -t 8`).

**Tabela 4.1. Tipos celulares e quantidade de amostras utilizadas para avaliação do perfil de expressão de miRNAs pelo projeto BLUEPRINT.**

Tipo celular	Quantidade de amostras
Linfócito T CD4	4
Megacariócito	2
Eritroblato	1
Neutrófilo	8
Monócito	5
Macrófago M1	4
Macrófago M2	5



A quantificação dos miRNAs presentes foi realizada utilizando o programa MMSEQ (Turro et al. 2011), perfis de expressão de miRNAs de cada tipo celular e os miRNAs diferencialmente expressos foram obtidos utilizando o programa MMDIFF (Turro et al. 2014). O programa MMSEQ infere a expressão dos genes presentes nas amostras a partir do conjunto de *reads* alinhados e de arquivos de anotação. Este programa considera que os *reads* de um determinado experimento podem alinhar em mais de um transcrito que podem ou não ter se originado do mesmo gene. Assim, este programa obtém a contagem de *reads* mapeados em cada transcrito e aplica o modelo de Poisson para inferir a expressão de cada transcrito. As distribuições posteriores dos parâmetros de expressão para os grupos de transcritos pertencentes ao mesmo gene são agregadas para fornecer estimativas de expressão do gene. Esta metodologia foi desenvolvida para incluir *reads* mapeados em mais de um *locus*, permitindo também obter estimativas de expressão distintas para diferentes membros de famílias gênicas e diferentes isoformas do mesmo transcrito. Este programa é executado em duas etapas, a primeira identifica os *reads* alinhados em regiões de transcritos (bam2hits arquivo fasta de referência arquivo bam de alinhamento > arquivo.hits) e a segunda infere as estimativas de expressão (mmseq arquivo.hits diretório de saída) O programa MMDIFF também faz parte do pacote do programa MMSEQ e realiza a comparação de modelos de regressão para identificar genes diferencialmente expressos (mmdiff -m matriz de comparação arquivos mmseq > arquivo.mmdiff).

A fim de determinar o papel de miRNAs na diferenciação celular hematopoiética, foram realizadas cinco análises de expressão diferencial em pontos-chaves da diferenciação das células (Figura 4.2), a saber:

- Mielócitos versus Linfócitos: amostras TCD4 versus todas as outras;
- Precursor Megacariócito Eritroblato (MEP) versus Precursor Granulócitos Monócito (GMP): Megacariócitos e Eritroblatos versus Neutrófilos, Monócitos e Macrófagos M1 e M2;
- Neutrófilos versus Monócitos: Neutrófilos versus Monócitos e Macrófagos M1 e M2;
- Monócitos versus Macrófagos M1 e M2
- Macrófagos M1 versus Macrófagos M2

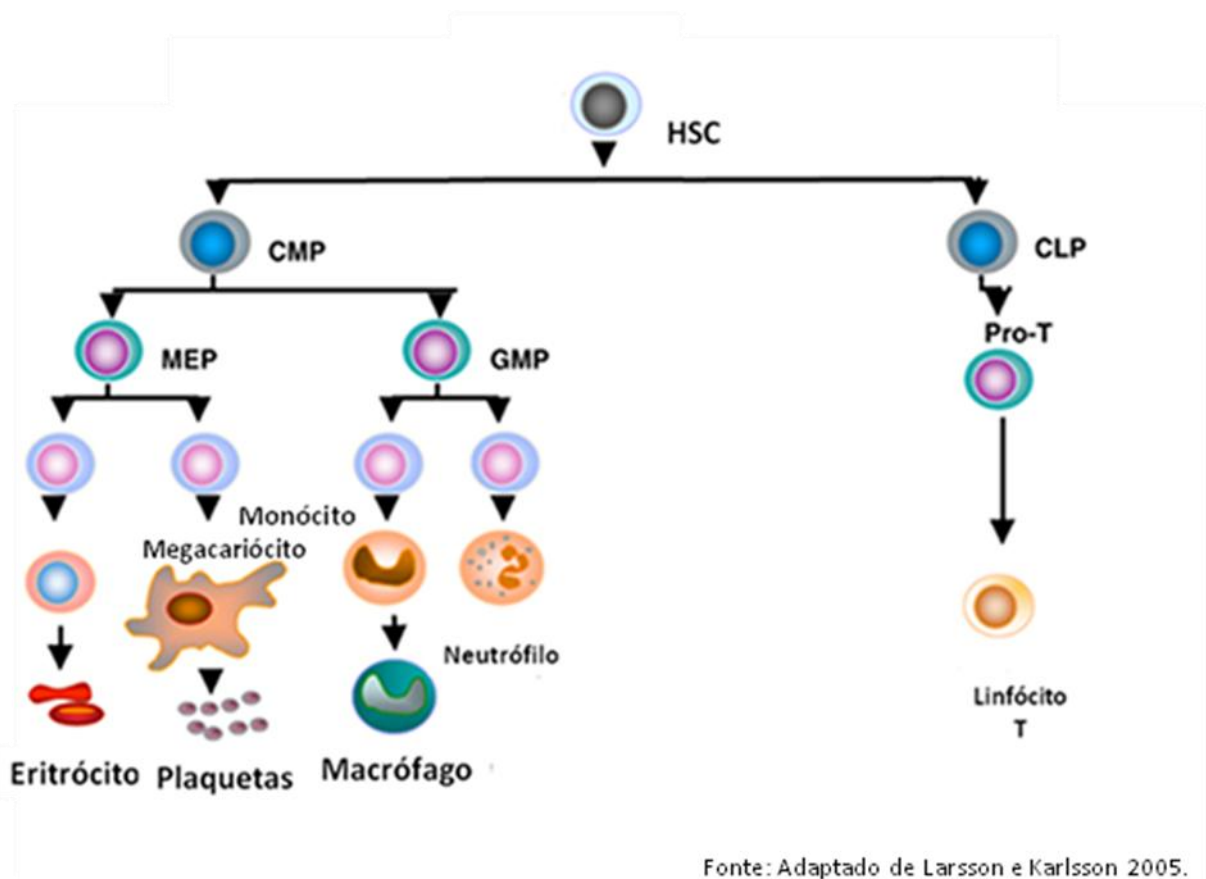


Figura 4.2. Árvore Hematopoiética. HSC: célula tronco hematopoética; CMP: precursor comum mieloide; CLP: precursor comum linfóide.

Os alvos validados dos miRNAs diferencialmente expressos foram obtidos utilizando o banco de dados miRTarBase versão 6.0 (versão de 15 de Setembro de 2015) (Download: 16 de Dezembro de 2016). (Hsu et al. 2014).

As mesmas comparações foram realizadas para dados de sequenciamento de mRNA dos mesmos tipos celulares (Tabela 4.2). O perfil de expressão dos alvos de miRNA diferencialmente expressos foi relacionado ao perfil dos mRNAs diferencialmente expressos. Os alvos menos expressos dos miRNAs mais expressos foram utilizados para identificar as vias metabólicas relacionadas à diferenciação e específicas de cada tipo celular utilizando o pacote do R GSEasy (Greene e Turro, não publicado).

**Tabela 4.2. Quantidade de amostras por tipo celular usada na comparação de mRNAs.**

Tipo celular	Número de amostras
Linfócito T CD4+	14
Megacariócito	7
Eritroblato	2
Neutrófilo	11
Monócito	8
Macrófago M1	6
Macrófago M2	7

## 5. RESULTADOS

### 5.1. Perfil de expressão de RNAs não codificadores pequenos em pacientes fumantes e não fumantes de adenocarcinoma de pulmão.

Os resultados aprensetados aqui fazem parte do artigo intitulado “snoRNA and piRNA expression levels modified by tobacco use in women with lung adenocarcinoma” (Jorge et al., 2017) (Anexo 1).

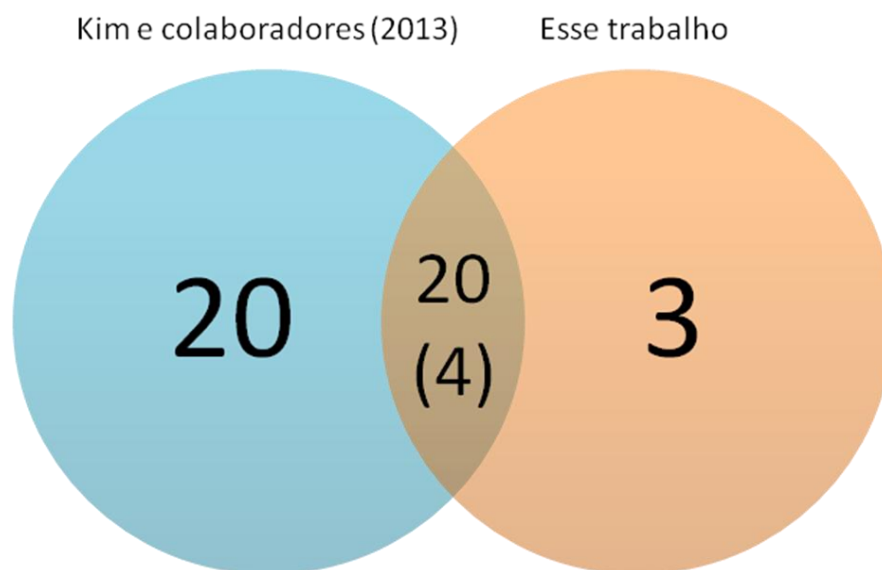
Em ambos conjuntos de dados, fumantes e não fumantes, aproximadamente 50% dos *reads* obtidos após filtragem de qualidade e remoção de adaptador foram alinhados unicamente. A Tabela 5.1 mostram a quantidade de *reads* totais e alinhados unicamente para fumantes e não fumantes, respectivamente.

#### 5.1.1. Validação da metodologia

Kim e colaboradores (Kim et al. 2013) encontraram 40 miRNAs diferencialmente expressos. Utilizando a mesma abordagem descrita na seção de 4.1.2 e a anotação da versão 21 do banco de dados do miRBase, encontramos 23 miRNAs diferencialmente expressos entre os tecidos normais e tumorais de não fumantes. Vinte destes miRNAs também foram encontrados por Kim e colaboradores (Figura 5.1) e outros quatro miRNAs encontrados por aqueles autores também foram encontrados nesse trabalho, mas removidos pelos filtros de FDR e logFC.

**Tabela 5.1. Totais de “reads” obtidos de pacientes com adenocarcinoma.**

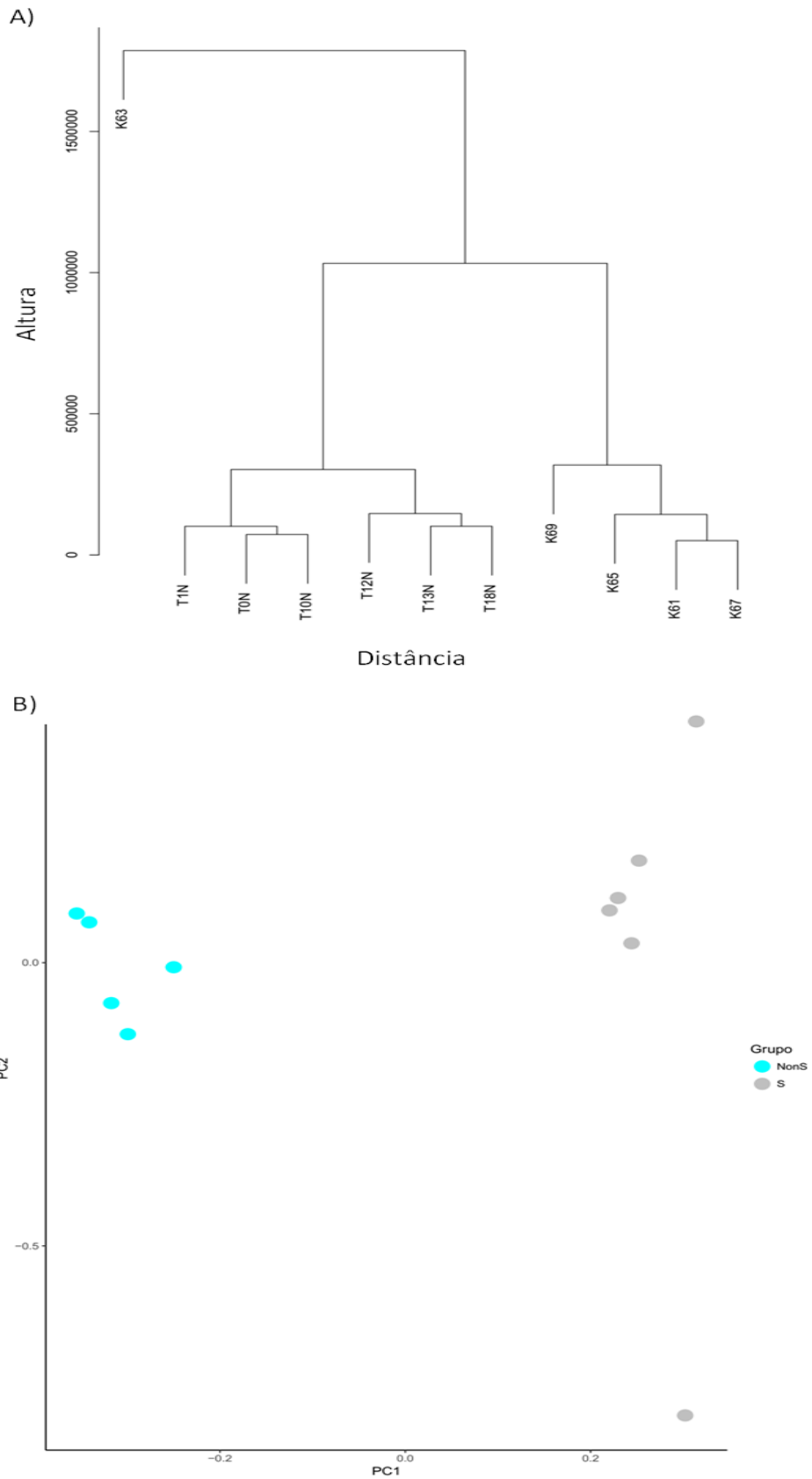
Nome do Arquivo	Tipo	Fumo	ID	Paciente	Total de Reads	Reads alinhados unicamente	Porcentagem
TCGA-91-6835-11A-01H-2169-13	Normal	Fumante	T0N	0	13.372.963	6.985.515	52,24
TCGA-91-6835-01A-11H-1857-13	Tumor	Fumante	T0T	0	12.622.721	5.401.630	42,79
TCGA-44-6777-11A-01H-2169-13	Normal	Fumante	T10N	10	5.660.175	2.911.521	51,44
TCGA-44-6777-01A-11H-1857-13	Tumor	Fumante	T10T	10	12.567.241	5.327.711	42,39
TCGA-49-6743-11A-01H-2169-13	Normal	Fumante	T12N	12	14.964.335	7.016.841	46,89
TCGA-49-6743-01A-11H-1857-13	Tumor	Fumante	T12T	12	10.566.238	3.690.428	34,93
TCGA-91-6836-11A-01H-2169-13	Normal	Fumante	T13N	13	13.624.532	6.666.210	48,93
TCGA-91-6836-01A-21H-1857-13	Tumor	Fumante	T13T	13	10.933.762	3.799.116	34,75
TCGA-44-6776-11A-01H-2169-13	Normal	Fumante	T18N	18	18.797.778	6.864.285	36,52
TCGA-44-6776-01A-11H-1857-13	Tumor	Fumante	T18T	18	10.672.335	5.101.922	47,81
TCGA-49-6744-11A-01H-2169-13	Normal	Fumante	T1N	1	10.776.834	4.958.274	46,01
TCGA-49-6744-01A-11H-1857-13	Tumor	Fumante	T1T	1	16.729.337	7.282.854	43,53
SRR493961	Normal	Não Fumante	K61	2	28.645.348	12.152.274	51,10
SRR493962	Tumoral	Não Fumante	K62	2	34.742.127	19.693.426	62,24
SRR493963	Normal	Não Fumante	K63	3	19.693.426	11.664.193	40,14
SRR493964	Tumoral	Não Fumante	K64	3	30.531.469	14.329.858	49,61
SRR493965	Normal	Não Fumante	K65	4	29.140.944	9.923.732	37,21
SRR493966	Tumoral	Não Fumante	K66	4	9.923.732	1.3141.624	47,38
SRR493967	Normal	Não Fumante	K67	5	30.970.140	13.430.860	46,56
SRR493968	Tumoral	Não Fumante	K68	5	30.105.887	16.592.787	60,35
SRR493969	Normal	Não Fumante	K69	6	31.808.386	12.091.840	41,23
SRR493970	Tumoral	Não Fumante	K70	6	12.091.840	11.978.451	46,62



**Figura 5.1. Diagrama de Venn entre Kim e colaboradores e o trabalho atual.** O círculo azul representa a quantidade de miRNAs diferencialmente expressos encontrados por Kim e colaboradores e o círculo vermelho a quantidade encontrada no trabalho atual, entre parêntesis, a quantidade de miRNAs que não passou pelo nosso filtro.

**5.1.2. Comparação do perfil de expressão de piRNAs e snoRNAs em amostras normais**

A fim de avaliar se as amostras Normal de Não fumante (NNonS) e Normal de Fumante (NS) (Figura 5.2) apresentavam perfis distintos de expressão de piRNAs e snoRNAs, foi realizada o agrupamento hierárquico das amostras (Figura 5.2a) e a análise de componente principal (Figura 5.2b). Ambas análises apresentaram clara distinção entre os dois grupos.



**Figura 5.2. Agrupamento das amostras normais.** A) Agrupamento hierárquico. As amostras que começam com T pertencem a pacientes fumantes e as que começam com K, a pacientes não fumantes. B) Análise de componente principal. NonS representam amostras de pacientes não fumantes e S, de pacientes fumantes.

A análise de expressão diferencial revelou 49 snoRNAs diferencialmente expressos entre os dois grupos (Figura 5.3). Ao todo, 29 genes tiveram a expressão reduzida em fumantes e 20 genes aumentada (Figura 5.4). Nenhum piRNA foi encontrado diferencialmente expresso nesta análise. O gene U60 (SNORD60) foi o snoRNA com a maior alteração encontrada, apresentando maior expressão nas amostras normais de não fumantes (logFC = 6.22). O gene menos expresso nas amostras normais de não fumantes foi o HBII-420 (SNORD99). As alterações encontradas entre os grupos de genes mais expressos e genes menos expressos apresentam ordem de magnitude semelhante. A Tabela 5.2 mostra os 10 genes com menor FDR e maior logFC entre os diferencialmente expressos e a lista completa de todos os genes diferencialmente expressos encontra-se no anexo 2.

**Tabela 5.2. Os cinco snoRNAs ou piRNAs diferencialmente expressos com maiores e menores logFC entre amostras normal de não fumantes e fumantes.**

Gene	logFC	logCPM	FDR
U60	6,22	18,78	2,31E-57
U76	5,84	11,88	6,63E-30
U15A	5,50	11,34	5,55E-24
U44	4,23	12,19	6,63E-30
HBII-296A	3,82	11,00	9,94E-21
HBII-142	-3,44	15,29	4,53E-22
snR38B	-3,85	15,37	1,52E-21
ACA7B	-4,21	9,70	1,21E-20
U30	-5,23	16,87	7,16E-36
HBII-420	-5,25	13,92	2,01E-30

logFC: logaritmo de "Fold Change"; logCPM: logaritmo de CPM; FDR: "false discovery rate"; logFC positivos indicam genes mais expressos em amostras normais de não fumantes.



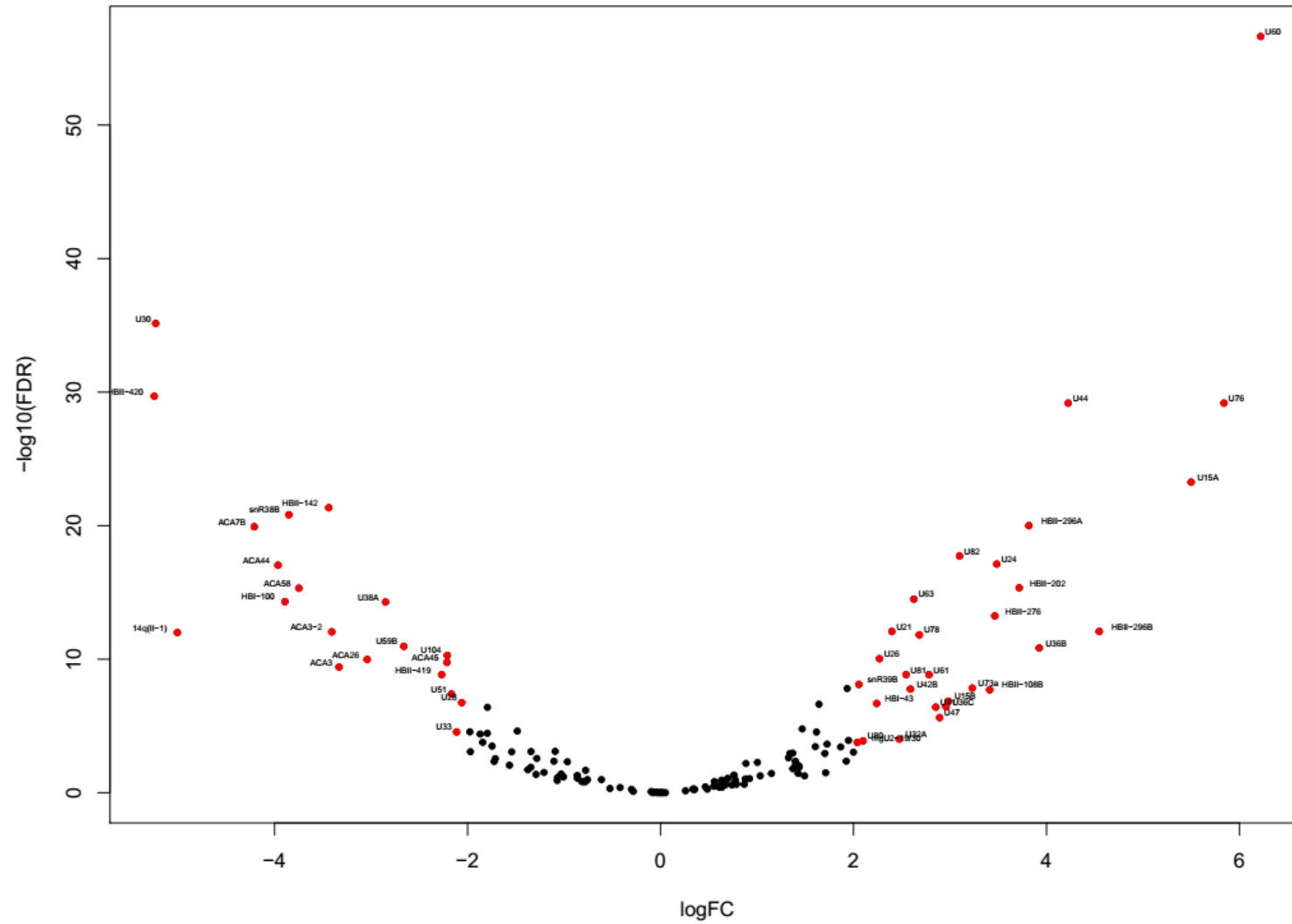
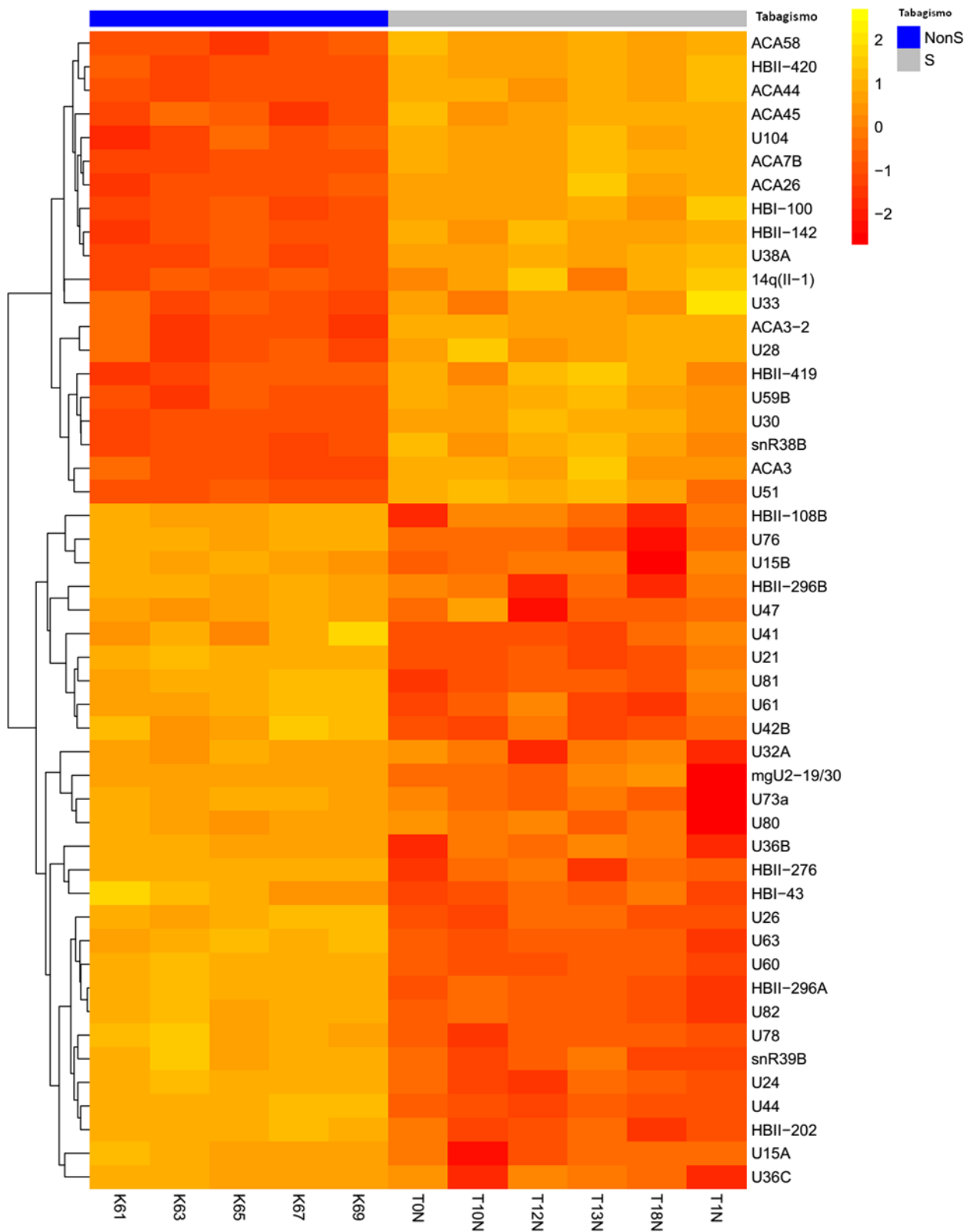


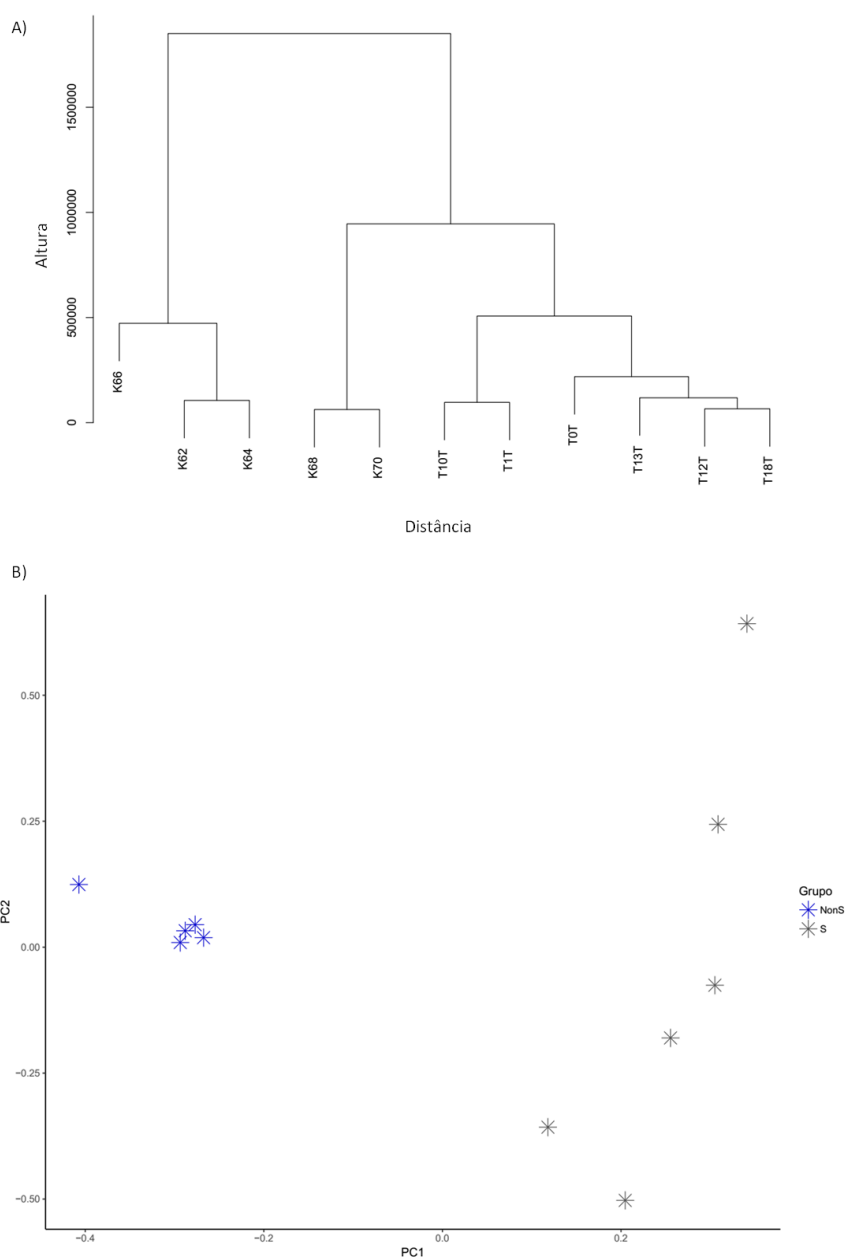
Figura 5.3. “Vulcano” plot dos genes diferencialmente expressos em amostras normais. Os genes diferencialmente expressos estão em vermelho.



**Figura 5.4. “Heatmap” dos genes diferencialmente expressos em amostras normais.** A barra azul indica as amostras de pacientes não fumantes e a barra cinza indica as amostras de pacientes fumantes.

### 5.1.3. Comparação do perfil de expressão de piRNAs e snoRNAs em amostras tumorais

As amostras tumorais também apresentaram distinção entre as amostras de pacientes não fumantes (TNonS) e de pacientes fumantes (TS) nas análises de agrupamento hierárquico (Figura 5.5a) e análise de componente principal (Figura 5.5b).



**Figura 5.5. Agrupamento das amostras tumorais.** A) Agrupamento hierárquico. As amostras que começam com T pertencem a pacientes fumantes e as que começam com K, a pacientes não fumantes. B) Análise de componente principal. NonS representam amostras de pacientes não fumantes e S, de pacientes fumantes.

A análise de expressão diferencial revelou 55 snoRNAs ou piRNAs alterados entre as amostras tumorais de não fumantes (TNonS) e fumantes (TS) (Figura 5.6).

Destes, 34 apresentavam-se mais expressos e 21 menos expressos em fumantes (Figura 5.7). O U60 (SNORD99) foi o gene mais expresso nas amostras de não fumantes ( $\logFC = 5,63$ ), e o gene menos expresso nestas amostras foi U30 (SNORD30). A Tabela 5.3 indica os 10 genes com menor FDR e maior  $\logFC$  encontrados nesta comparação. Apesar de não haver diferenças significativas na ordem de magnitude das alterações encontradas entre os genes mais expressos e menos expressos desta comparação, as alterações encontradas nestas análises são menores que as encontradas na comparação entre as amostras normais. As informações sobre todos os genes alterados nesta análise encontram-se no anexo 3.

Dois piRNAs apresentaram alterações significativas no seu padrão de expressão, sendo eles: hsa-piR-010894-3 ( $\logFC = 3.43$ ) e hsa-piR-001168-4 ( $\logFC = 2.95$ ). Estes dois piRNAs são mais expressos em amostras tumorais de não fumantes.

**Tabela 5.3. Os cinco piRNAs ou snoRNAs diferencialmente expressos com maiores e menores  $\logFC$  entre amostras tumorais de não fumantes e fumantes.**

Gene	$\logFC$	$\logCPM$	FDR
U60	5,63	19,02	1,69E-28
U76	4,00	11,35	6,78E-22
U15A	3,89	10,75	2,34E-12
U63	3,01	13,33	2,67E-12
HBII-142	-3,00	13,91	2,34E-12
U58C	-3,60	9,94	5,95E-13
U37	-3,69	9,45	5,46E-13
HBII-99	-4,15	8,07	3,21E-15
HBII-420	-4,80	13,20	7,43E-30
U30	-6,72	17,81	5,38E-35

$\logFC$ : logaritmo de "Fold Change";  $\logCPM$ : logaritmo de CPM; FDR: "false discovery rate";  $\logFC$  positivos indicam genes mais expressos em amostras normais de não fumantes.

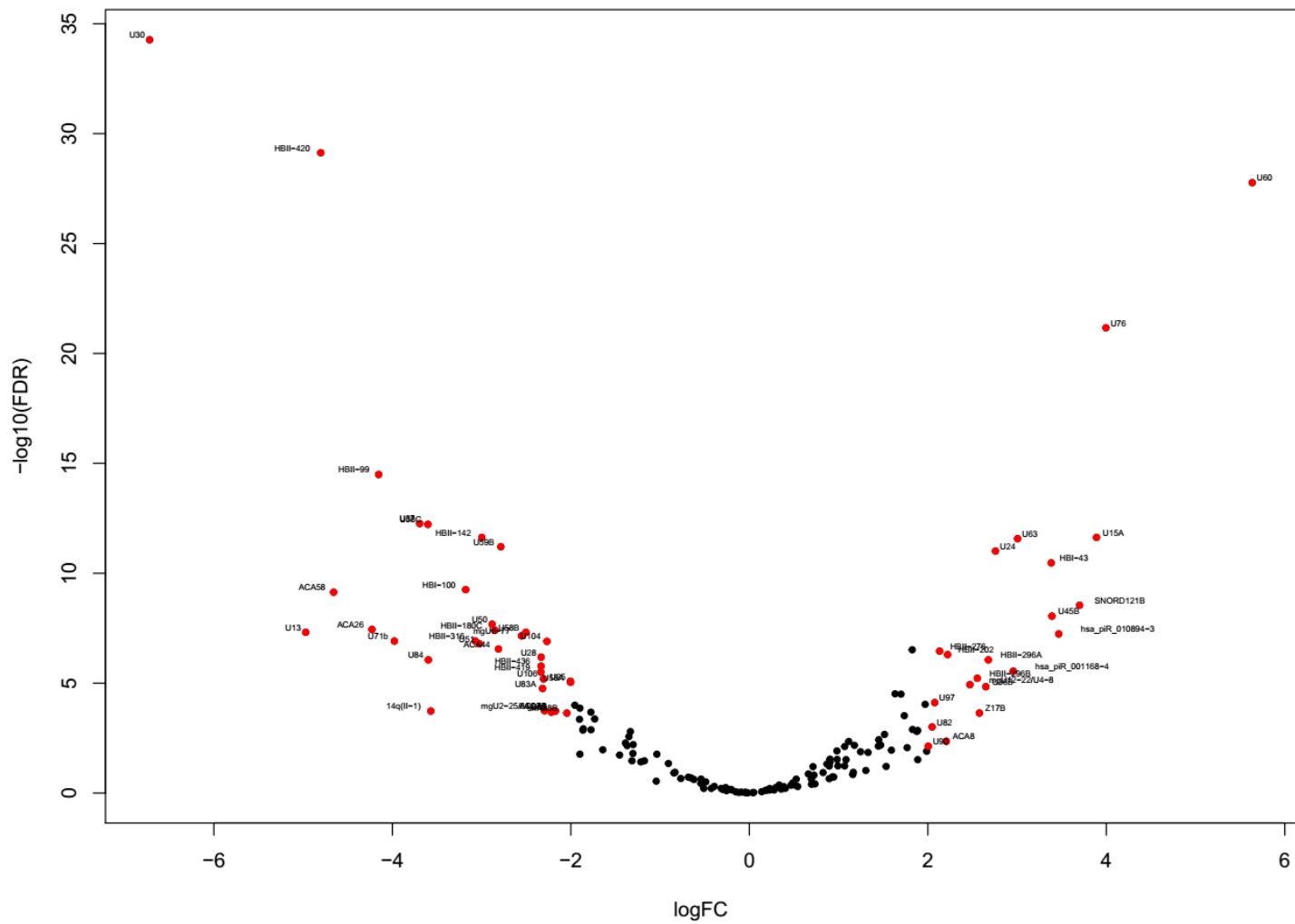
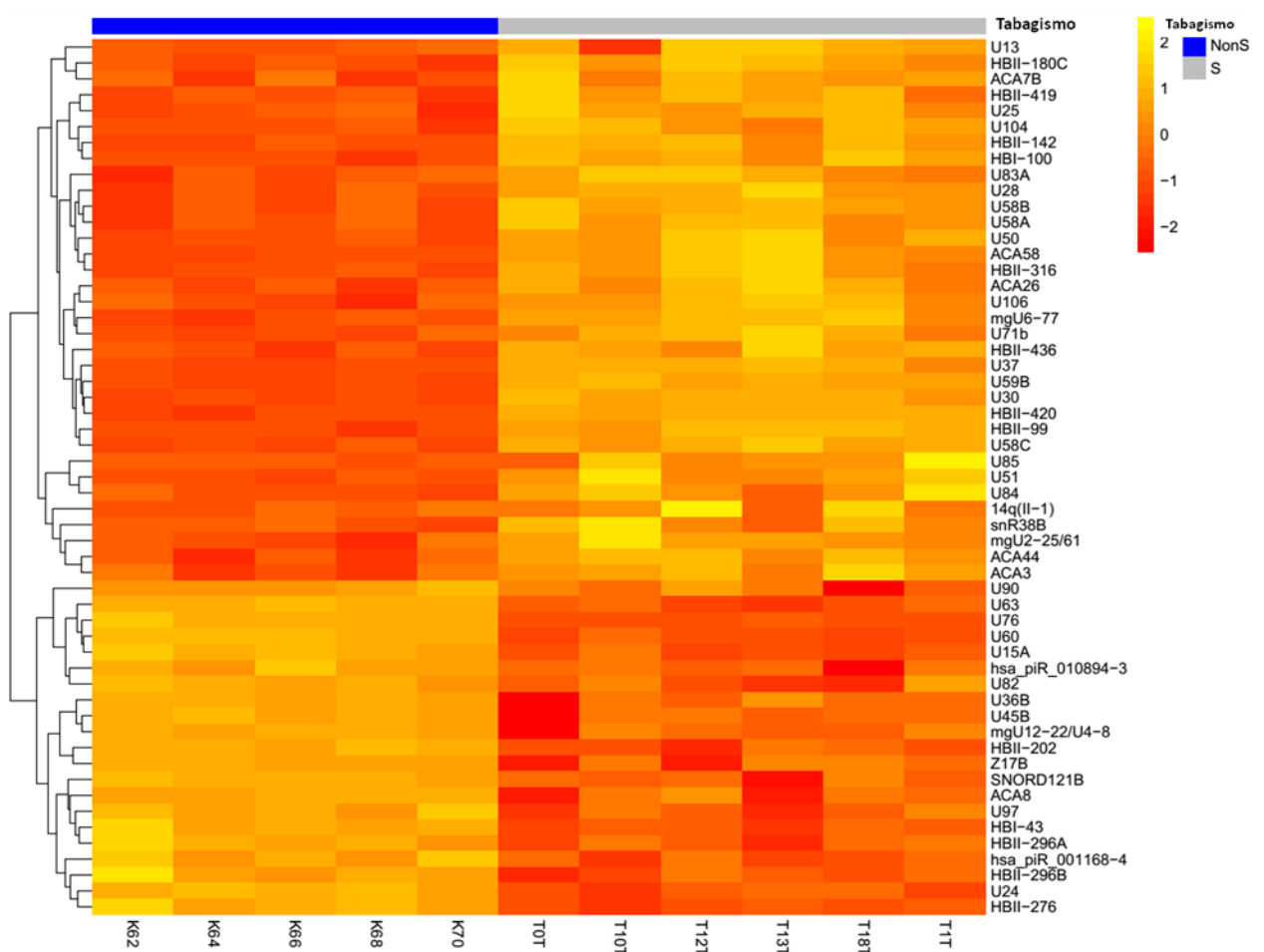


Figura 5.6. “Volcano” plot dos genes diferencialmente expressos em amostras tumorais. Os genes diferencialmente expressos estão em vermelho.



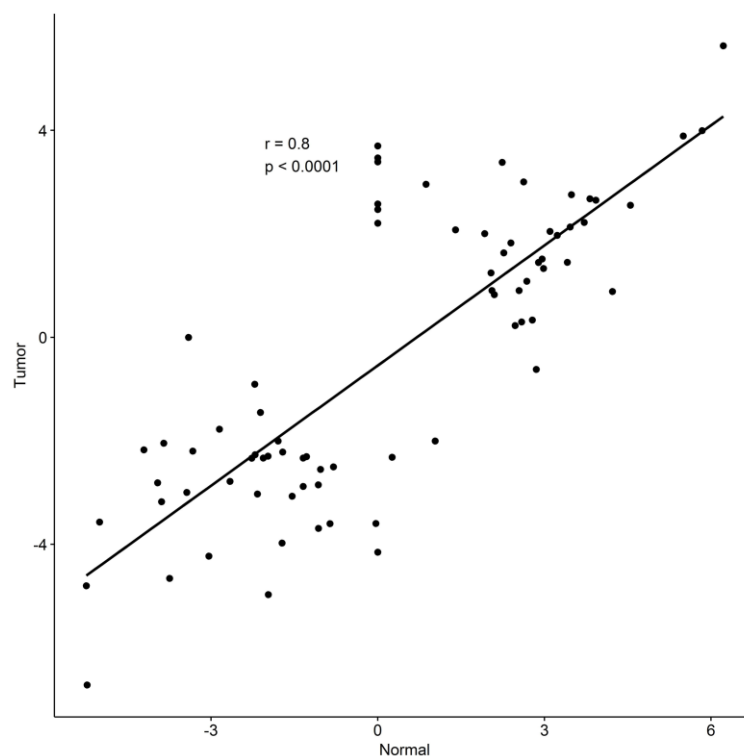
**Figura 5.7. Heatmap dos genes diferencialmente expressos em amostras tumorais.** A barra azul indica as amostras de pacientes não fumantes e a barra cinza indica as amostras de pacientes fumantes.

Ao todo, 28 genes apresentaram alterações na sua expressão em ambas análises (Tabela 5.4Tabela 5.3Figura 5.5). Estes genes apresentavam a mesma tendência de alteração, independente se a amostra era tumoral ou normal (coeficiente de relação de Pearson: 0,80) (Figura 5.8). Como será apresentado em detalhes na seção 5.1.4, 11 genes apresentaram expressão que não se altera independente do estado patológico e que mostraram alterações significativas no seu padrão de expressão quando realizadas as comparações envolvendo o uso de tabaco (em negrito na Tabela 5.4).

**Tabela 5.4. snoRNAs e piRNAs diferencialmente expressos em ambas comparações.**

Gene	Normal		Tumor	
	logFC	logCPM	logFC	logCPM
<b>U60</b>	<b>6,22</b>	<b>18,78</b>	<b>5,63</b>	<b>19,02</b>
U76	5,84	11,88	4,00	11,35
U15A	5,50	11,34	3,89	10,75
HBII-296B	4,55	9,25	2,55	8,61
U36B	3,93	8,90	2,65	7,92
HBII-296 <sup>a</sup>	3,82	11,00	2,68	10,41
HBII-202	3,72	10,82	2,22	9,74
U24	3,49	12,99	2,76	12,15
HBII-276	3,47	10,62	2,13	10,11
U82	3,10	12,44	2,05	11,71
<b>U63</b>	<b>2,63</b>	<b>13,61</b>	<b>3,01</b>	<b>13,33</b>
HBI-43	2,24	10,87	3,38	10,32
<b>U28</b>	<b>-2,06</b>	<b>10,96</b>	<b>-2,33</b>	<b>10,89</b>
<b>U51</b>	<b>-2,17</b>	<b>10,76</b>	<b>-3,03</b>	<b>10,96</b>
<b>U104</b>	<b>-2,21</b>	<b>16,35</b>	<b>-2,27</b>	<b>15,16</b>
<b>HBII-419</b>	<b>-2,27</b>	<b>13,59</b>	<b>-2,34</b>	<b>13,21</b>
<b>U59B</b>	<b>-2,66</b>	<b>10,53</b>	<b>-2,78</b>	<b>9,88</b>
ACA26	-3,04	7,84	-4,23	9,84
ACA3	-3,33	7,33	-2,20	7,47
<b>HBII-142</b>	<b>-3,44</b>	<b>15,29</b>	<b>-3,00</b>	<b>13,91</b>
ACA58	-3,75	9,25	-4,66	9,44
snR38B	-3,85	15,37	-2,05	13,25
<b>HBI-100</b>	<b>-3,89</b>	<b>10,90</b>	<b>-3,18</b>	<b>9,98</b>
ACA44	-3,96	8,77	-2,81	8,15
ACA7B	-4,21	9,7	-2,17	8,92
14q(II-1)	-5,01	11,04	-3,57	9,52
<b>U30</b>	<b>-5,23</b>	<b>16,87</b>	<b>-6,72</b>	<b>17,81</b>
<b>HBII-420</b>	<b>-5,25</b>	<b>13,92</b>	<b>-4,80</b>	<b>13,20</b>

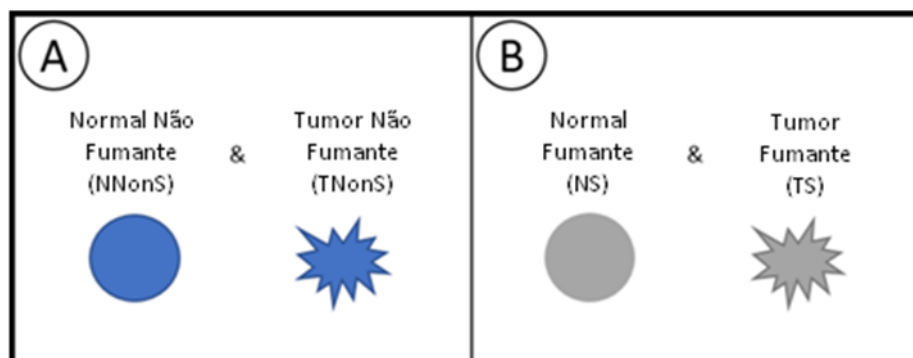
logFC: logaritmo de “Fold Change”; logCPM: logaritmo de CPM; logFC maiores que 0 indicam genes mais expressos em amostras não fumantes. Genes em negrito apresentaram baixa dispersão entre as amostras normais e tumorais.



**Figura 5.8. Gráfico de correlação entre os logFC dos genes diferencialmente expressos.** Os genes mais expressos e menos expressos formam dos grupos distintos na parte superior direita e inferior esquerda, respectivamente.

#### 5.1.4. *sncRNAs constitutivos em não fumantes*

A fim de determinar os genes que não se alteram entre as amostras normais e tumorais de não fumantes, foi realizada a análise de dispersão (Figura 5.9).



**Figura 5.9. Comparação realizada para identificação de genes constitutivamente expressos.** A) Comparação entre as amostras de não fumantes. B) Comparação entre as amostras de fumantes.

Foram encontrados 179 genes cuja expressão normalizada apresenta desvio padrão menor que 1. A Tabela 5.5 mostra os 10 genes com menor desvio padrão e a tabela com todos os genes encontrados está no anexo 4. Amostras com final ímpar correspondem ao tecido normal adjacente ao tumor, e as amostras de final



par, ao tecido tumoral. As amostras de final ímpar são pareadas com a de final par seguinte. Por exemplo, a amostra normal K61 é pareada com a amostra tumoral K62, indicando que ambas as amostras pertencem a uma mesma paciente.

**Tabela 5.5. Os 10 snoRNAs constitutivamente expressos com menor desvio padrão em amostras de não fumantes.**

Gene	K61	K62	K63	K64	K65	K66	K67	K68	K69	K70	Var	DP
U51	8,45	8,48	8,54	8,58	8,79	8,37	8,57	8,76	8,53	8,35	0,02	0,15
U57	12,20	12,38	12,34	12,30	12,07	11,92	12,37	12,36	12,25	11,67	0,05	0,23
HBII-429	9,42	9,37	9,57	9,40	9,81	9,60	9,34	8,92	9,72	9,42	0,06	0,25
HBII-95	5,89	5,62	5,37	5,94	5,87	5,94	5,77	6,12	5,38	5,63	0,06	0,25
U21	13,34	13,27	14,15	13,56	13,60	13,52	13,54	13,68	13,77	13,30	0,07	0,26
U35B	4,98	4,85	4,84	4,14	4,83	4,45	4,93	4,82	4,96	4,74	0,07	0,26
U37	6,65	6,37	5,98	5,95	6,25	6,13	6,21	6,73	6,32	6,67	0,08	0,28
U75	9,08	9,38	8,77	9,23	8,66	8,88	9,01	8,94	8,90	8,42	0,08	0,28
U95	12,08	12,11	12,30	11,66	12,48	11,85	12,26	11,87	12,36	12,45	0,08	0,28
ACA8	5,96	5,27	4,95	5,11	5,50	5,64	5,45	5,52	5,36	5,68	0,08	0,29

DP: desvio padrão; Var: variância. Amostras que terminam em números ímpares pertencem a tecidos normais e as amostras que terminam em números pares pertencem a tecidos tumorais.

### 5.1.5. *sncRNAs constitutivos em Fumantes*

A análise de dispersão das amostras normais e tumorais de fumantes (Figura 5.9) revelou 33 genes constitutivamente expressos. A Tabela 5.6 mostra os 10 genes com menor desvio padrão e no anexo 5 encontram-se os dados de todos os genes constitutivamente expressos.

Todos os genes considerados constitutivos nesta análise também apresentaram baixa dispersão entre as amostras normais e tumorais de não fumantes (Tabela 5.7). Em geral, nota-se que a expressão destes genes é mais uniforme nas amostras de pacientes não fumantes devido às menores variâncias e desvios padrão encontrados.

**Tabela 5.6. Os 10 snoRNAs constitutivamente expressos com menor desvio padrão em fumantes.**

Gene	T0N	T0T	T10N	T10T	T12N	T12T	T13N	T13T	T18N	T18T	T1N	T1T	Var	SD
U59B	11,50	11,05	11,56	11,58	11,53	11,03	11,90	11,49	11,74	10,87	11,08	11,23	0,10	0,32
U25	13,24	13,95	12,68	12,93	12,97	12,98	13,14	13,57	13,13	13,52	12,62	12,57	0,17	0,41
U57	12,71	12,29	12,37	12,40	12,28	11,78	12,25	12,91	12,67	11,72	11,69	11,88	0,17	0,41
U48	12,81	13,26	13,23	13,56	13,49	13,46	13,34	12,75	13,38	14,38	13,39	12,89	0,19	0,43
U27	14,04	14,30	13,27	14,65	14,07	13,93	14,05	14,04	13,93	13,40	12,96	13,92	0,21	0,46
U21	11,90	11,95	12,52	11,94	12,28	12,87	11,28	12,62	12,36	12,85	13,21	12,03	0,29	0,53
HBII-420	14,95	14,53	15,03	14,01	14,24	14,78	14,67	14,85	14,49	14,90	16,25	14,67	0,30	0,55
U20	13,15	13,07	13,00	11,70	12,94	12,96	12,70	13,61	12,90	11,75	12,98	12,22	0,33	0,57
U43	13,73	12,91	12,84	12,45	13,74	13,25	13,43	12,72	14,26	13,67	14,02	12,59	0,36	0,60
snR39B	12,67	12,55	12,00	11,33	12,42	13,40	12,90	12,97	12,19	12,85	11,79	13,31	0,38	0,62

DP: desvio padrão; Var: variância. Amostras que terminam em “N” pertencem a tecidos normais e as amostras que terminam “T” pertencem a tecidos tumorais.

**Tabela 5.7. snoRNAs constitutivos em comum entre não fumantes e fumantes.**

Gene	Não fumante			Fumante		
	Média	Var	DP	Média	Var	DP
U59B	7,78	0,09	0,30	11,38	0,10	0,32
U25	10,34	0,27	0,52	13,11	0,17	0,41
U57	12,19	0,05	0,23	12,25	0,17	0,41
U48	11,47	0,18	0,43	13,33	0,19	0,43
U27	14,34	0,14	0,37	13,88	0,21	0,46
U21	13,57	0,07	0,26	12,32	0,29	0,53
HBII-420	8,87	0,20	0,45	14,78	0,30	0,55
U20	10,69	0,11	0,33	12,75	0,33	0,57
U43	12,68	0,29	0,53	13,30	0,36	0,60
snR39B	13,19	0,23	0,48	12,53	0,38	0,62
U63	13,91	0,14	0,38	11,90	0,38	0,62
U104	13,66	0,37	0,61	16,78	0,42	0,65
HBII-336	11,03	0,44	0,66	11,49	0,43	0,66
HBII-419	11,27	0,23	0,48	14,34	0,44	0,66
SNORD119	12,97	0,29	0,53	12,55	0,45	0,67
U28	8,88	0,24	0,49	11,96	0,46	0,68
U42A	9,70	0,41	0,64	12,09	0,46	0,68
U38A	8,92	0,40	0,64	12,10	0,48	0,69
U59A	9,03	0,15	0,38	10,89	0,49	0,70
HBII-55	9,94	0,21	0,46	11,12	0,52	0,72
U31	13,95	0,17	0,42	15,06	0,53	0,73
HBII-142	11,65	0,25	0,50	15,74	0,55	0,74
ACA45	10,14	0,35	0,59	12,69	0,57	0,75
HBII-210	11,67	0,16	0,40	11,67	0,58	0,76
U52	11,03	0,34	0,58	11,64	0,60	0,78
U95	12,14	0,08	0,28	13,54	0,64	0,80
HBII-251	11,78	0,22	0,47	12,24	0,65	0,81
HBI-100	7,22	0,32	0,57	11,54	0,67	0,82
U60	19,47	0,53	0,73	14,43	0,71	0,84
U51	8,54	0,02	0,15	11,78	0,82	0,91
U50	8,63	0,17	0,41	11,53	0,88	0,94
HBII-295	11,43	0,29	0,54	13,22	0,90	0,95
U30	11,75	0,14	0,37	18,49	0,90	0,95

DP: desvio padrão; Var: variância.

## 5.2. Análise de miRNAs diferencialmente expressos entre diferentes tipos sanguíneos celulares do projeto BLUEPRINT

As amostras do tecido sanguíneos do projeto Blueprint também apresentaram aproximadamente 50% dos *reads* alinhados em cada amostra (Tabela 5.8).

**Tabela 5.8. Reads alinhados do projeto Blueprint**

Tipo celular	Arquivo	Total reads	Reads Descartados	Reads Alinhados	Porcentagem
Linfócito T CD4+	BLUE_1_9319_ACAGTG.fastq	11.291.954	564.554	5.434.810	48,13
Linfócito T CD4+	BLUE_1_9319_GTAGAG.fastq	15.703.273	3.447.057	3.629.565	23,11
Linfócito T CD4+	BLUE_2_9320_ACAGTG.fastq	12.085.251	2.214.458	3.808.303	31,51
Linfócito T CD4+	BLUE_3_9321_CGTACG.fastq	17.138.713	3.574.411	1.380.318	8,05
Eritroblato	BLUE_3_9321_AGTTCG.fastq	5.829.568	2.039.966	800.253	13,73
Neutrófilo	BLUE_3_9321_ATGTCA.fastq	19.796.984	576.714	9.889.719	49,96
Neutrófilo	BLUE_2_9320_TGACCA.fastq	27.015.307	2.688.567	11.804.403	43,70
Macrófago M1	BLUE_1_9319_TGACCA.fastq	14.052.794	4.997.495	1.998.049	14,22
Macrófago M1	BLUE_2_9320_AGTTCG.fastq	7.731.708	504.007	4.892.486	63,28
Macrófago M1	BLUE_3_9321_CCGTCC.fastq	11.283.560	783.132	7.350.923	65,15
Macrófago M2	BLUE_1_9319_ATCACG.fastq	28.758.627	8.096.252	2.739.714	9,53
Macrófago M2	BLUE_1_9319_CCGTCC.fastq	12.197.590	1.055.959	5.661.747	46,42
Macrófago M2	BLUE_1_9319_GTTTCG.fastq	13.026.246	4.026.211	1.322.519	10,15
Macrófago M2	BLUE_2_9320_ACTTGA.fastq	9.809.343	205.565	1.421.885	14,50
Macrófago M2	BLUE_2_9320_GTGGCC.fastq	14.108.527	4.837.715	856.370	6,07
Macrófago M2	BLUE_3_9321_GGCTAC.fastq	9.778.004	826.779	4.512.282	46,15
Megacariócito	BLUE_2_9320_ATGTCA.fastq	14.763.798	5.018.689	2.448.120	16,58
Megacariócito	BLUE_3_9321_TTAGGC.fastq	11.601.289	2.984.418	3.221.139	27,77
Monócito	BLUE_1_9319_TAGCTT.fastq	28.146.728	13.962.419	2.148.021	7,63
Monócito	BLUE_2_9320_ATCACG.fastq	20.358.780	5.153.367	5.125.614	25,18
Monócito	BLUE_3_9321_ACAGTG.fastq	11.914.955	1.486.691	6.208.439	52,11
Monócito	BLUE_3_9321_GCCAAT.fastq	12.908.182	1.937.113	5.600.748	43,39
Neutrófilo	BLUE_1_9319_CAGATC.fastq	33.375.068	1.376.011	14.042.103	42,07
Neutrófilo	BLUE_1_9319_GAGTGG.fastq	11.836.475	2.388.748	474.993	4,01
Neutrófilo	BLUE_2_9320_GAGTGG.fastq	20.567.746	3.193.454	8.256.599	40,14
Neutrófilo	BLUE_2_9320_GCCAAT.fastq	7.777.189	332.376	4.875.688	62,69
Neutrófilo	BLUE_2_9320_GTCCGC.fastq	14.869.223	3.189.215	4.521.925	30,41
Neutrófilo	BLUE_3_9321_CAGATC.fastq	8.457.178	344.496	5.014.813	59,30
Neutrófilo	BLUE_3_9321_CGATGT.fastq	9.666.338	804.975	3.192.224	33,02

Inicialmente, utilizamos o programa MMDIFF para obter estimativas dos níveis de expressão de cada miRNA avaliado de acordo com os diferentes tipos celulares

estudados. Esta análise revelou perfis distintos entre células T CD4+, megacariócitos e eritroblatos, neutrófilos, monócitos, macrófagos M1 e macrófagos M2 (Figura 5.10).

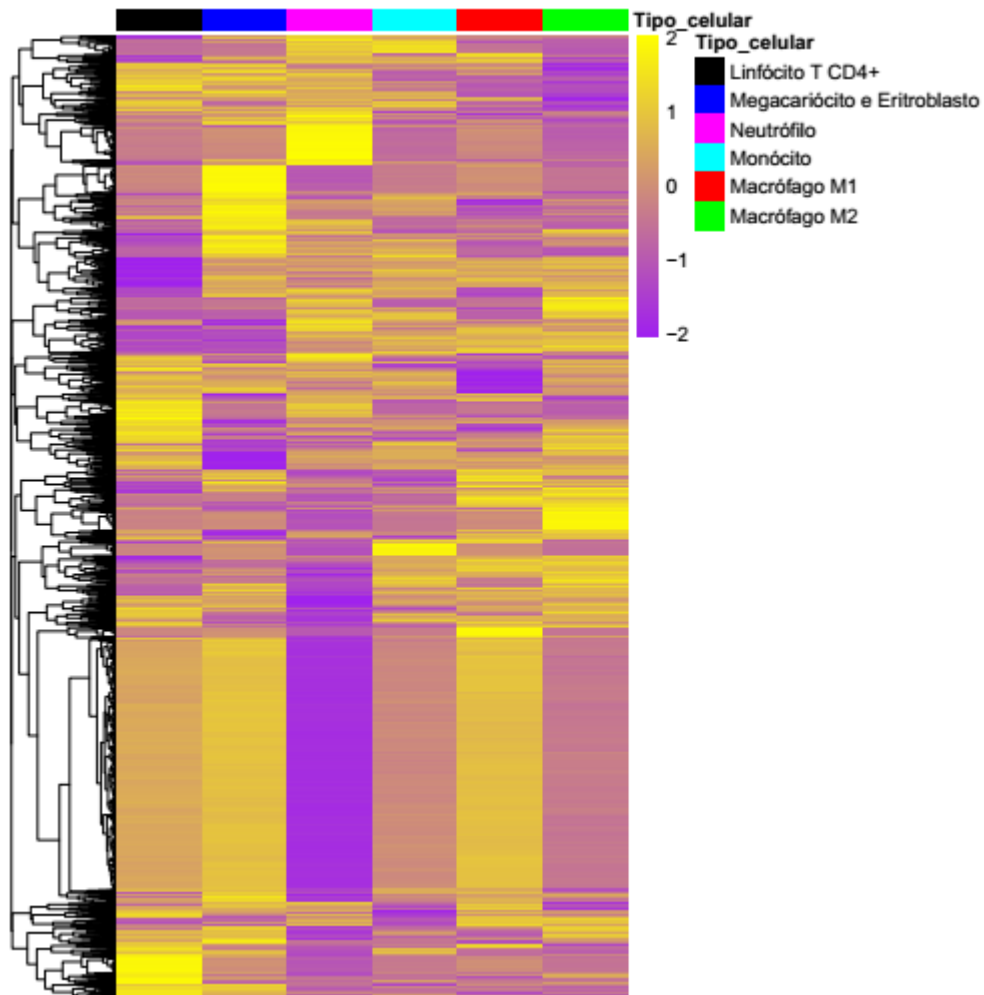


Figura 5.10. Perfil de expressão de miRNA por tipo celular do BLUEPRINT.

Em seguida, o mesmo programa foi utilizado para identificar os miRNAs diferencialmente expressos em diferentes níveis da árvore hematopoiética descrita na Figura 4.2. A Tabela 5.9 resume a quantidade de miRNAs diferencialmente expressos encontrados nessas comparações. Por exemplo, no nível hierárquico mais alto da árvore hematopoiética, é possível comparar as linhagens linfocíticas e mielocíticas quando são comparadas as expressões das amostras de linfócitos T CD4+ contra todos os outros tipos celulares.

**Tabela 5.9. Quantidade de miRNAs encontrados diferencialmente expressos.**

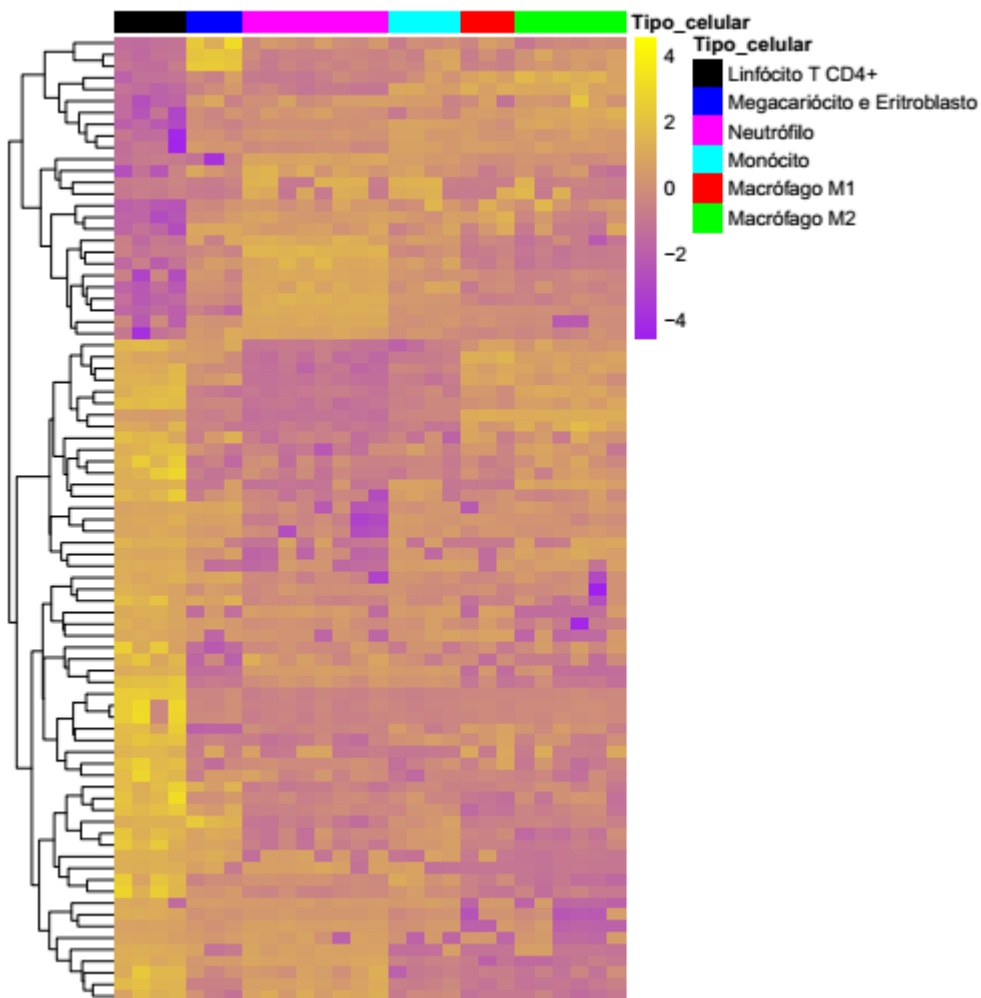
Comparação	Mais expressos	Menos expressos	Total
Mielócitos vs. Linfócitos	57	26	83
GMP vs. MEP	51	19	70
Monoócitos vs. Neutrófilos	122	111	233
Macrófagos vs. Monocitos	20	42	62
Macrófago M2 vs. Macrófago M1	5	4	9

Mais ou menos expresso refere-se sempre ao segundo elemento da comparação.

Assim, a primeira comparação realizada entre as amostras de linfócitos T CD4+ contra todos os outros tipos celulares, identificou 83 miRNAs diferencialmente expressos (Figura 5.11), sendo 57 miRNAs mais expressos e 26 miRNAs menos expressos em linfócitos T CD4+. A Tabela 5.10 mostra os 10 miRNAs com maiores probabilidades posteriores e a lista contendo todos os miRNA diferencialmente expressos encontra-se no anexo 6.

**Tabela 5.10. Os 10 miRNAs diferencialmente expressos entre linfócitos e os outros tipos celulares com maiores probabilidades posteriores.**

miRNA	Probabilidade Posterior	LogFC
hsa-miR-6761-3p	1,000	-0,727
hsa-miR-6876-5p	1,000	-0,965
hsa-miR-874-3p	1,000	3,809
hsa-miR-338-5p	1,000	-3,356
hsa-miR-155-5p	1,000	1,799
hsa-miR-3934-5p	1,000	4,227
hsa-miR-345-5p	1,000	-3,335
hsa-miR-1271-5p	0,999	2,475
hsa-miR-95-3p	0,999	5,226
hsa-miR-142-5p	0,999	2,874



**Figura 5.11. miRNAs diferencialmente expressos entre mielócitos e linfócitos.**

Esta abordagem também identificou 70 miRNAs diferencialmente expressos em megacariócitos e eritroblatos (Figura 5.12), sendo 51 mais expressos e 19 menos expressos entre GMP e MEP. A lista com todos os miRNAs diferencialmente expressos está no anexo 7 e a lista com os 10 miRNAs com maiores probabilidades posteriores encontra-se na Tabela 5.11.

Tabela 5.11. Os 10 miRNAs diferencialmente expressos entre GMP e MEP com maiores probabilidades posteriores.

miRNA	Probabilidade Posterior	LogFC
hsa-miR-193a-5p	0,999	-4,315
hsa-miR-338-5p	0,999	-3,310
hsa-miR-296-3p	0,998	3,820
hsa-miR-582-3p	0,996	-4,225
hsa-miR-125b-2-3p	0,996	3,389
hsa-miR-1249-3p	0,996	-3,212
hsa-miR-335-3p	0,996	5,317
hsa-miR-195-3p	0,994	2,538
hsa-miR-582-5p	0,992	-4,198
hsa-miR-151a-3p	0,991	3,133

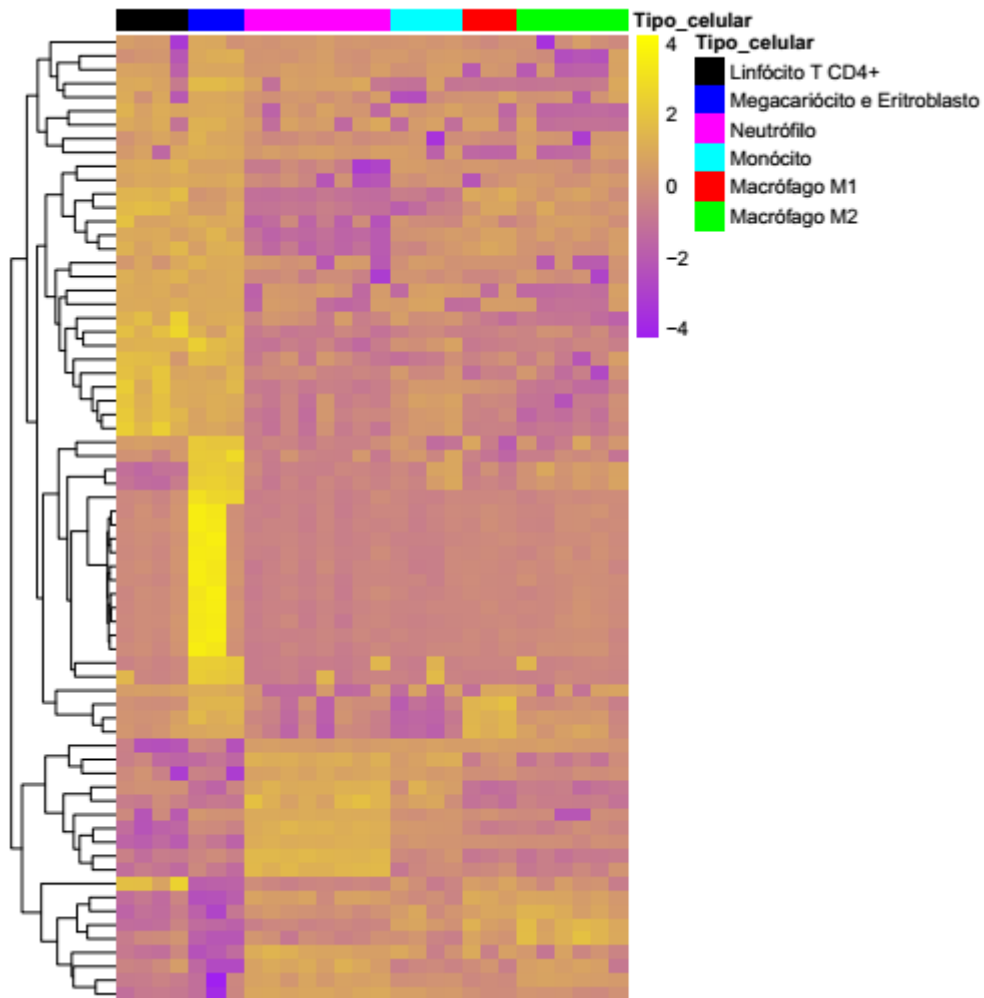


Figura 5.12. miRNAs diferencialmente expressos entre megacariócitos e eritroblatos e os outros tipos celulares avaliados.

A análise para identificação dos miRNAs diferencialmente expressos em neutrófilos foi a que apresentou mais miRNAs diferencialmente expressos: 144 ao todo (Figura 5.13). A maioria dos miRNAs apresentou-se mais expresso, 123 ao todo, enquanto 121 miRNAs foram encontrados menos expressos em neutrófilos. A lista com todos os miRNAs diferencialmente expresso está no anexo 8 e os 10 miRNAs com maiores probabilidades posteriores encontram-se na Tabela 5.12.

**Tabela 5.12. Os 10 miRNAs diferencialmente expressos entre neutrófilos e monócitos e macrófagos M1 e M2 com maior probabilidade posterior.**

miRNA	Probabilidade Posterior	LogFC
hsa-miR-143-3p	1,000	5,451
hsa-miR-143-5p	1,000	5,344
hsa-miR-145-3p	1,000	5,241
hsa-miR-155-5p	1,000	-2,939
hsa-miR-194-5p	1,000	3,402
hsa-miR-2115-3p	1,000	3,791
hsa-miR-2115-5p	1,000	3,525
hsa-miR-338-5p	1,000	2,423
hsa-miR-378a-3p	1,000	-2,700
hsa-miR-378c	1,000	-1,796



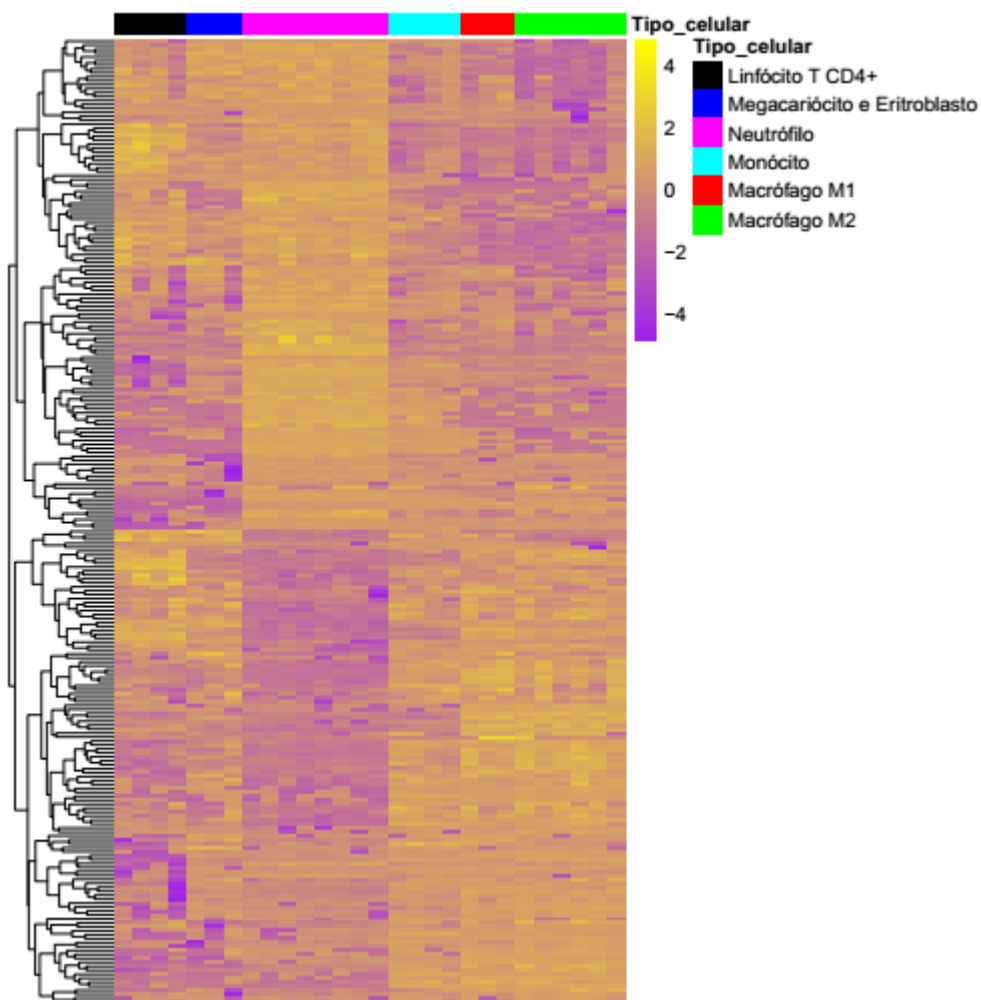


Figura 5.13. miRNAs diferencialmente expressos em neutrófilos.

A comparação para identificação dos miRNAs diferencialmente expressos em monócitos revelou 66 miRNAs com expressão alterada (Figura 5.14), sendo 22 mais expressos e 44 menos expressos entre monócitos e macrófagos. Todos os miRNAs alterados estão no anexo 9 e a lista com os 10 miRNAs com maiores probabilidades posteriores encontra-se na Tabela 5.13.

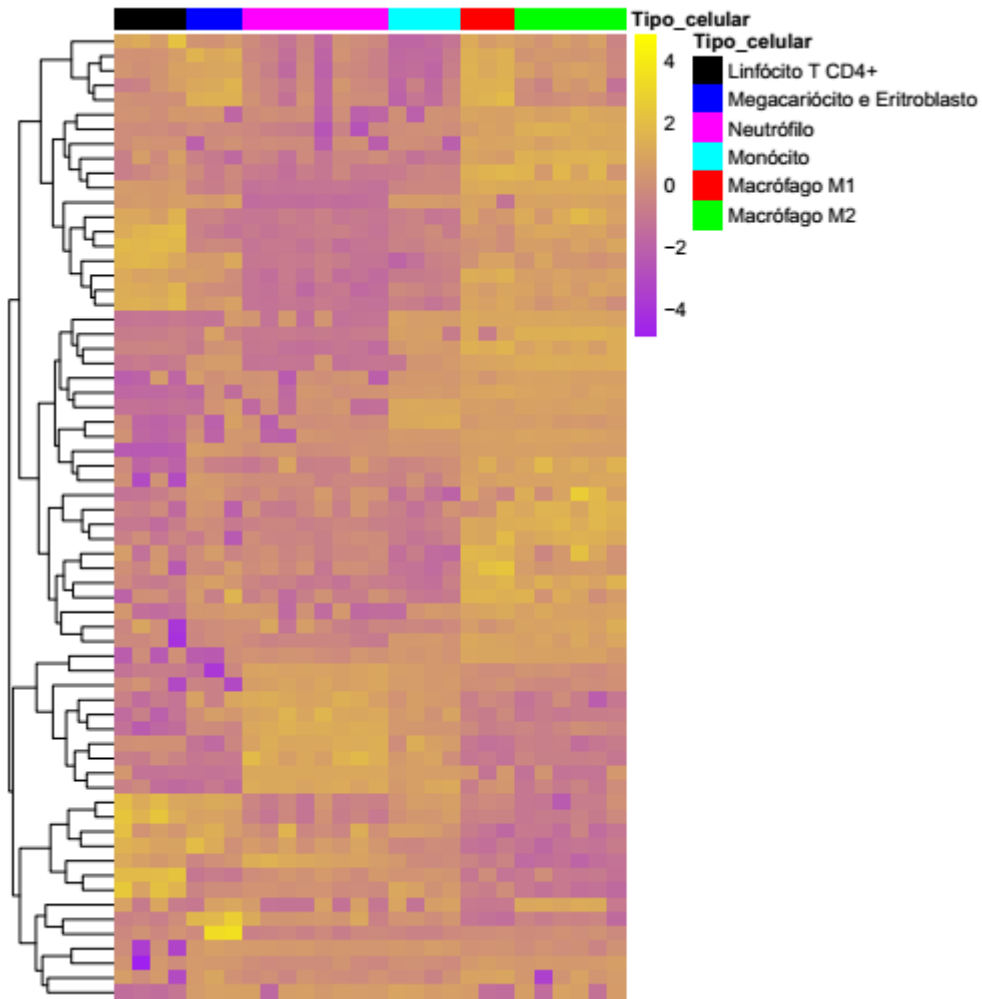


Figura 5.14. miRNAs diferencialmente expressos em monócitos.

Tabela 5.13. Os 10 miRNAs diferencialmente expressos entre monócitos e macrófagos M1 e M2 com maiores probabilidades posteriores.

miRNA	Probabilidade Posterior	LogFC
hsa-miR-155-5p	1,000	-2,853
hsa-miR-4734	1,000	0,673
hsa-miR-497-3p	1,000	0,053
hsa-miR-221-3p	1,000	-2,914
hsa-miR-511-5p	1,000	-5,262
hsa-miR-34a-5p	1,000	-4,964
hsa-miR-511-3p	1,000	-5,200
hsa-miR-3200-3p	1,000	-5,133
hsa-miR-146a-5p	1,000	-3,761
hsa-miR-212-5p	0,999	-2,498

A última comparação, para identificação dos miRNAs alterados em macrófagos M1 revelou 13 miRNAs diferencialmente expressos (Figura 5.15), sendo cinco mais expressos e sete menos expressos entre macrófagos M1 e macrófagos M2. A Tabela 5.14 mostra a lista com todos os miRNAs diferencialmente expressos.

**Tabela 5.14. Os 10 miRNAs diferencialmente entre macrófagos M1 e macrófagos M2 com maior probabilidade posterior.**

miRNA	Probabilidade Posterior	LogFC
hsa-miR-6727-3p	1,000	-0,245
hsa-miR-6761-3p	1,000	-0,219
hsa-miR-6876-3p	1,000	-1,012
hsa-miR-7155-3p	1,000	0,138
hsa-miR-6501-5p	0,996	-2,709
hsa-miR-511-3p	0,993	-2,317
hsa-miR-3614-5p	0,984	-2,325
hsa-miR-181d-5p	0,963	1,680
hsa-miR-511-5p	0,949	-2,800
hsa-miR-99b-3p	0,938	2,371
hsa-miR-181b-5p	0,929	1,178
hsa-miR-181c-3p	0,917	1,838

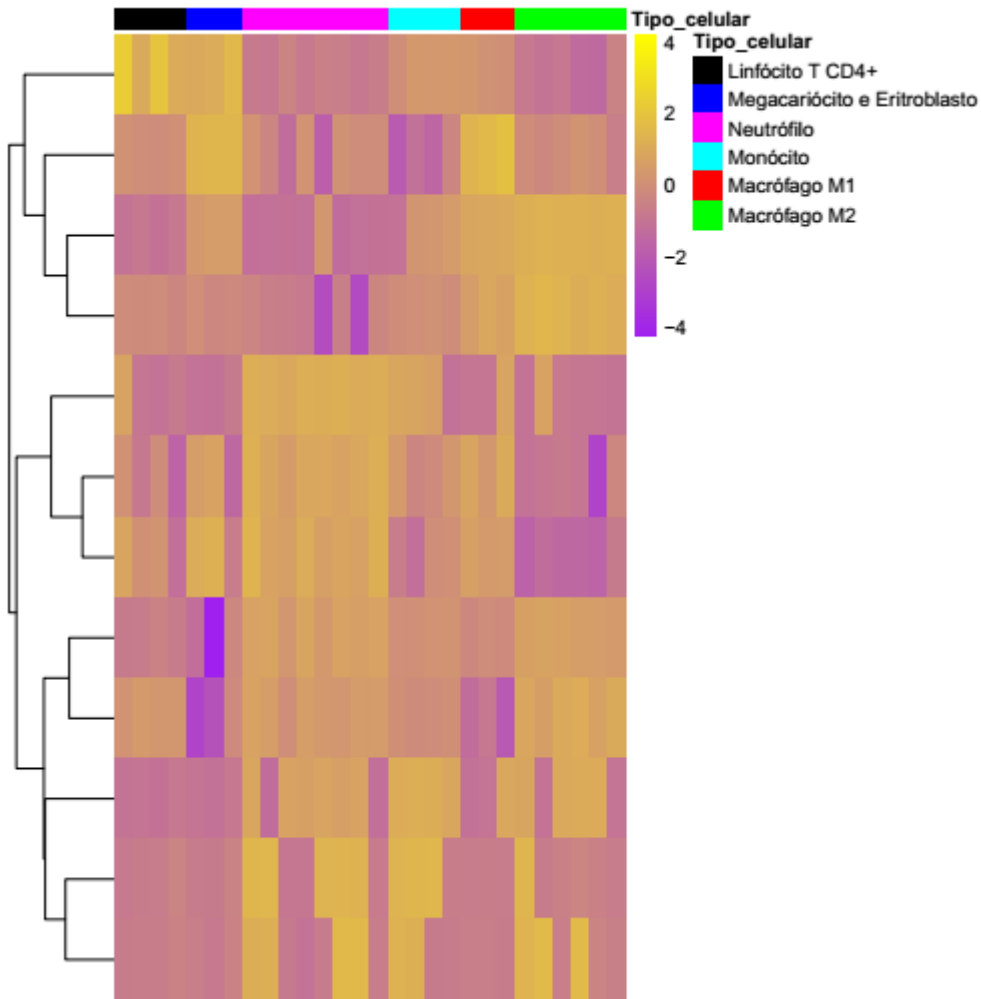


Figura 5.15. miRNAs alterados em macrófagos M1.

Comparamos também os miRNAs diferencialmente expressos encontrados em todas as análises para identificar aqueles cuja expressão é mais alterada na árvore hematopoiética.

Nenhum miRNA foi encontrado alterado em todas as análises. No entanto, a expressão de três miRNAs (miR-618, miR-146b e miR-223-3p) não foi alterada apenas na última comparação (macrófago M1 x macrófago M2). Outros 23 miRNAs estão alterados em três comparações e 56 miRNAs em duas comparações. A avaliação dos miRNAs encontrados em comum em três análises ou mais revelou que estes genes alteram sua expressão na linhagem de granulócitos (Figura 5.16). O anexo 10 lista todos os miRNAs alterados em mais de uma comparação e seus respectivos logFC.

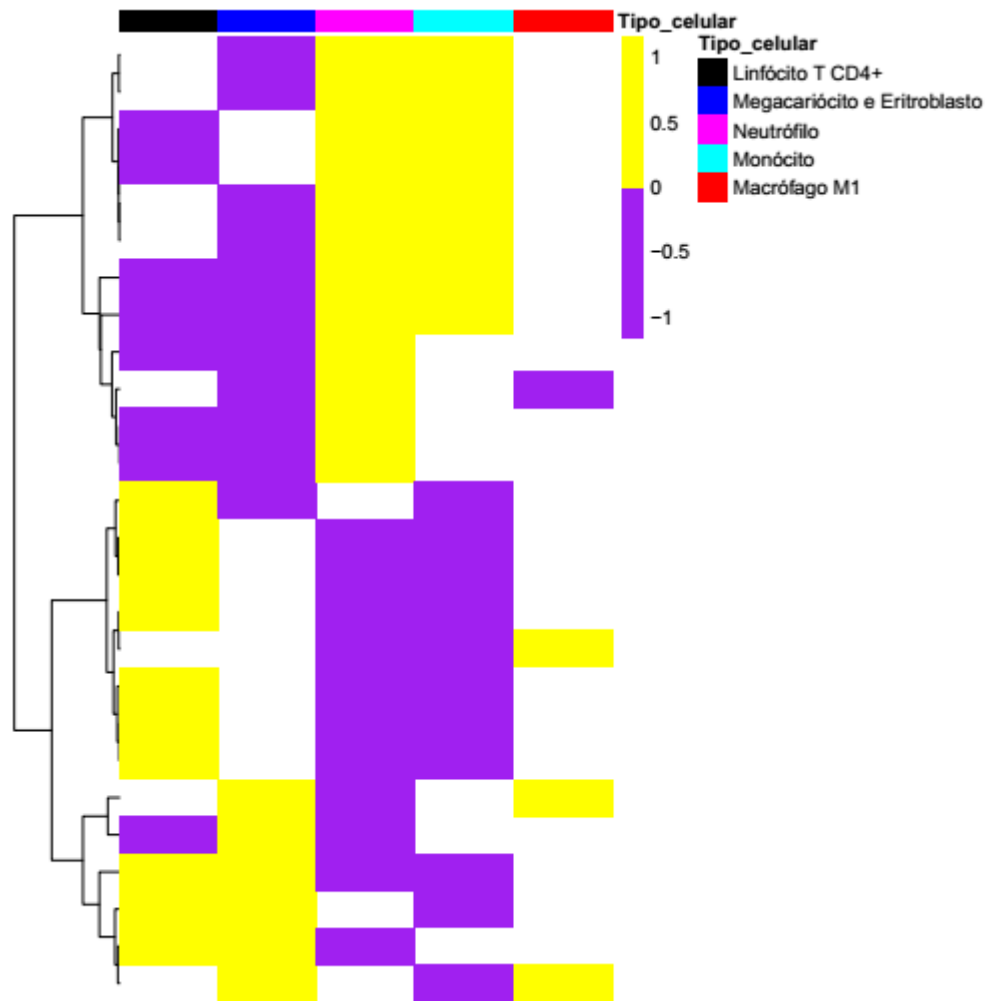


Figura 5.16. miRNAs diferencialmente expressos em três ou mais comparações.

Apesar do objetivo principal do projeto BLUEPRINT ser a obtenção e comparação de perfis epigenéticos, os centros de pesquisa envolvidos também geram dados de HTS de miRNA e de mRNAs. Portanto, a fim de identificar as vias metabólicas alteradas relacionadas aos miRNAs diferencialmente expressos em cada tipo celular, as mesmas comparações foram realizadas com dados de HTS de mRNAs do projeto Blueprint. A Tabela 5.15 indica a quantidade mRNAs encontrados mais e menos expresso em cada comparação e as tabelas Tabela 5.16 a Tabela 5.20 mostram os 10 mRNAs com maior probabilidade posterior e maiores e menores logFC.

**Tabela 5.15. mRNAs diferencialmente expressos por comparação realizada.**

Comparação	Mais expresso	Menos Expresso	Total
Mielócitos vs. Linfócitos	5.549	4.463	10.012
GMP vs. MEPEritroblato	4.497	3.727	8.226
Monoócitos vs. Neutrófilos	7.209	5.145	12.354
Macrófagos vs. Monocitos	4.438	3.960	8.398
Macrófago M2 vs. Macrófago M1	1.281	1.438	2.719

Mais ou menos expresso refere-se sempre ao segundo elemento da comparação.

**Tabela 5.16. Os 10 mRNAs diferencialmente expressos com maior probabilidade posterior entre linfócitos e os outros tipos celulares.**

Gene	Probabilidade Posterior	logFC
TRAV6	1	9,38
BCL11B	1	8,05
THEMIS	1	7,71
CD3D	1	7,56
LINC00861	1	7,55
SIGLEC5	1	-8,10
IL8	1	-7,97
TREML3P	1	-7,45
FCGR2A	1	-7,33
AQP9	1	-6,94

**Tabela 5.17. Os 10 mRNAs diferencialmente expressos com maior probabilidade posterior entre GMP vs. MEP.**

Gene	Probabilidade Posterior	logFC
RP11-8L18.2	1	6,24
DMTN	1	5,98
AC007253.1	1	5,17
FN3K	1	4,75
MAD2L1	1	4,26
GLT1D1	1	-7,68
BCL2A1	1	-7,59
CXCL16	1	-6,07
RP11-408H20.2	1	-5,05
LINC00877	1	-4,67

**Tabela 5.18. Os 10 mRNAs diferencialmente expressos com maior probabilidade posterior entre neutrófilo e monócito e macrófago M1 e M2 ..**

Gene	Probabilidade Posterior	logFC
RP11-290L7.3	1	16,98
RP11-109D20.1	1	15,42
AC008074.5	1	14,43
RP4-753D10.5	1	14,14
RP11-329N22.1	1	13,46
CTSL	1	-6,65
DDO	1	-6,59
PLA2G7	1	-6,54
DAGLA	1	-6,38
FAM129B	1	-6,18

**Tabela 5.19. Os 10 mRNAs diferencialmente expressos com maior probabilidade posterior entre monócito e macrófago M1 e M2 .**

Gene	Probabilidade Posterior	logFC
VCAN-AS1	1,00	9,67
SELL	1,00	8,04
SERPINB10	1,00	7,60
TPPP3	1,00	7,05
ASGR2	1,00	6,97
RP3-473B4.3	0,97	-13,67
SLC15A5	0,99	-14,77
RP11-20G13.3	1,00	-15,49
RP11-20G13.2	1,00	-16,32
RP11-96C23.11	1,00	-17,26

**Tabela 5.20. Os 10 mRNAs diferencialmente expressos com maior probabilidade posterior entre macrófago M1 e M2 ..**

Gene	Probabilidade Posterior	logFC
NCF1C	1	5,95
SEMA4A	1	4,23
CHST2	1	4,15
SOD2	1	3,76
RP11-714G18.1	1	0,18
RP11-268F1.3	1	-0,53
AC010969.1	1	-0,63
RP11-408H20.2	1	-0,74
RP11-15D14.2	1	-2,18
UCP3	1	-4,62

Também comparamos as listas de mRNAs diferencialmente expressos a fim de identificar os genes alterados em mais de uma comparação. Ao todo, 319

mRNAs foram encontrados diferencialmente expressos em todas as análises (Figura 5.17), 1.960 mRNAs em quatro análises, 4.039 em três análises e 4.444 em duas análises.

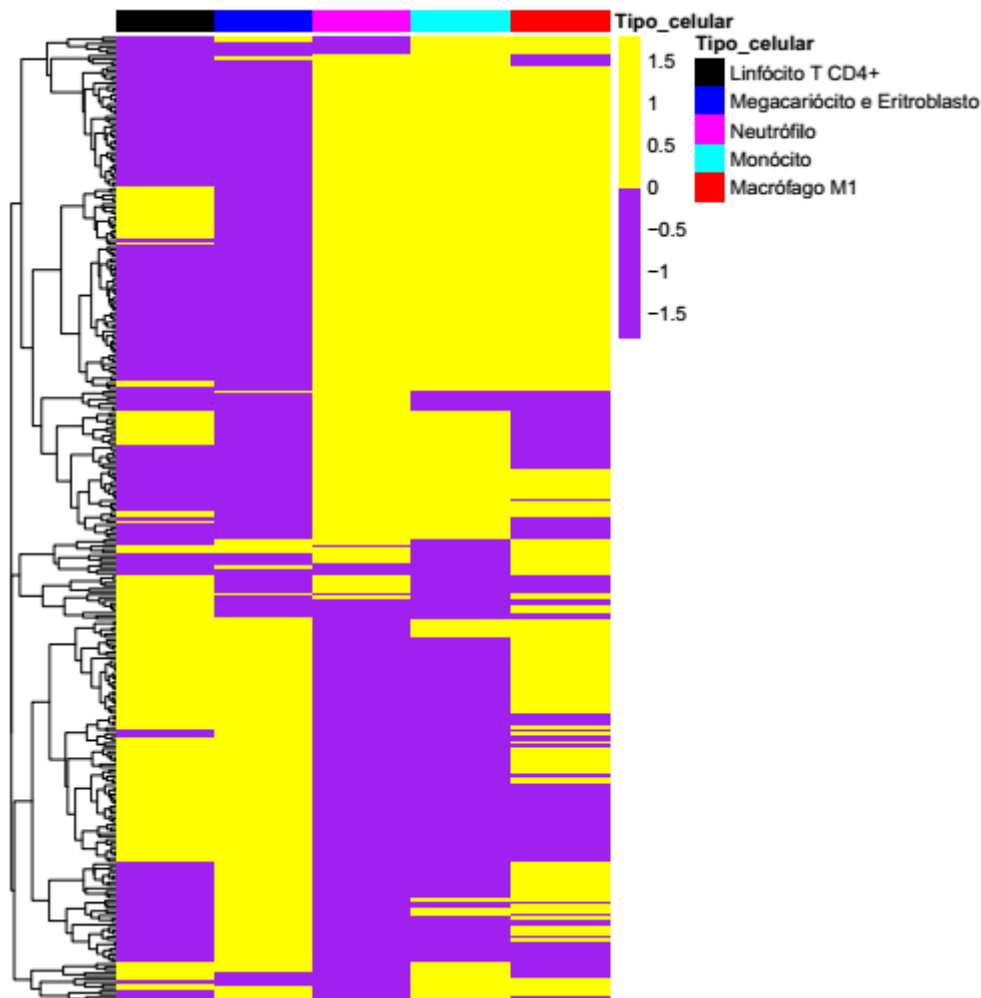


Figura 5.17. mRNAs diferencialmente expressos em todas as análises.

A fim de obter as vias metabólicas que efetivamente são afetadas pela expressão do miRNA, selecionamos somente os alvos menos expressos dos miRNAs mais expressos para a análise de enriquecimento de vias metabólicas. A Tabela 5.21 mostra a quantidade de alvos menos expressos de miRNAs mais expressos encontrados em cada comparação.



**Tabela 5.21. Quantidade de alvos menos expressos do miRNAs mais expressos encontradas.**

Comparação	Quantidade de alvos
Mielócitos VS. Linfócitos	1.747
GMP VS. MEP	817
Monócitos VS. Neutrófilos	2.974
Macrófagos VS. Monócitos	889
Macrófago M2 vs. Macrófago M12	25

Menos expresso refere-se sempre ao segundo elemento da comparação.

A análise de enriquecimento de vias metabólicas foi realizada usando o ambiente R, o banco de dados GO e o pacote do R GsEasy, foram consideradas alteradas as vias metabólicas com p-valor menor que 0,01. A Tabela 5.22 mostra a quantidade de vias alteradas encontradas nas comparações. Somente a comparação entre macrófagos M1 e macrófagos M2 não apresentou nenhuma via metabólica dentro do limite de p-valor estabelecido. A lista completa das vias metabólicas afetadas se encontram nos anexos 11 a 14. As Tabela 5.23 a Tabela 5.26 mostram as 10 vias metabólicas inibidas com menor p-valor e as figuras Figura 5.18 a Figura 5.21 mostram os 10 miRNAs mais expressos e seus alvos menos expressos para as três vias metabólicas com menor p-valor relacionadas aos tipos celulares da nossa comparação.

**Tabela 5.22. Quantidade de vias metabólicas alteradas por comparação.**

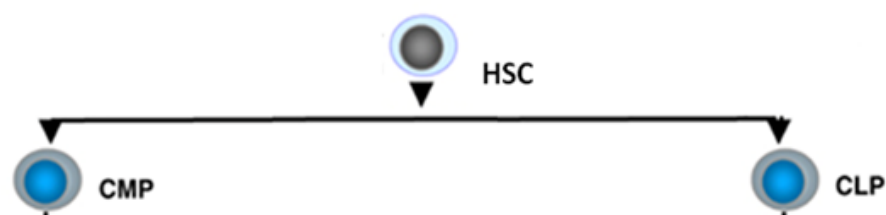
Comparação	Quantidade de vias metabólicas alteradas
Mielócitos VS. Linfócitos	152
GMP VS. MEP	173
Monócitos VS. Neutrófilos	222
Macrófagos VS. Monócitos	113
Macrófago M2 VS. Macrófago M1	0

As vias estão alteradas sempre ao segundo elemento da comparação.

**Tabela 5.23. As 10 vias metabólicas inibidas com menor p-valor em linfócitos T CD4+.**

Via metabólica	P-valor
Regulação positiva de angiogênese	2,00E-05
Regulação da resposta a ferimento	4,00E-05
Membrana do granulo terciário	6,00E-05
Membrana granulo alfa de plaqueta	7,00E-05
Regulação da recuperação de ferimento	7,00E-05
Regulação positiva da fosforilação de peptidil-tirosina	8,00E-05
Atividade oxidoreductase, oxidação de íons metálicos	9,00E-05
Estabelecimento da localização do fuso mitótico	1,80E-04
Ligação a carboidrato	2,00E-04
Regulação positiva do desenvolvimento de vascularização.	2,10E-04

A)



B)

Membrana do granulo alfa de plaquetas (p-valor: 7e-05)					
miRNA	Probabilidade posterior do miRNA	logFc do miRNA	Gene alvo	Probabilidade posterior do gene alvo	LogFC do gene alvo
hsa-miR-874-3p	1,00	3,81	LHFPL2	1,00	-4,14
hsa-miR-155-5p	1,00	1,80	CD36	1,00	-5,91
hsa-miR-3941	0,99	3,56	SYTL4	1,00	-3,81
hsa-miR-125b-2-3p	0,98	2,35	LHFPL2	1,00	-4,14
hsa-miR-125b-2-3p	0,98	2,35	SYTL4	1,00	-3,81
hsa-miR-30b-3p	0,97	1,12	ITGB3	1,00	-4,03
hsa-miR-455-5p	0,97	2,79	CD36	1,00	-5,91
hsa-miR-363-3p	0,95	2,54	LHFPL2	1,00	-4,14
hsa-let-7c-5p	0,94	1,48	ITGB3	1,00	-4,03

Regulação positiva da resposta a ferimento (p-valor: 3,9e-4)					
miRNA	Probabilidade posterior do miRNA	logFc do miRNA	Gene alvo	Probabilidade posterior do gene alvo	LogFC do gene alvo
hsa-miR-155-5p	1,00	1,80	CD36	1,00	-5,91
hsa-miR-155-5p	1,00	1,80	RREB1	1,00	-0,54
hsa-miR-155-5p	1,00	1,80	THBS1	0,99	-3,76
hsa-miR-155-5p	1,00	1,80	MYLK	0,98	-2,86
hsa-miR-142-5p	1,00	2,87	NFE2L2	1,00	-0,82
hsa-miR-99a-5p	0,99	2,49	SERPINE1	0,97	-2,45
hsa-let-7g-5p	0,99	2,27	THBS1	0,99	-3,76
hsa-miR-194-5p	0,99	2,01	HBEGF	1,00	-2,60
hsa-miR-766-5p	0,99	2,20	THBS1	0,99	-3,76
hsa-miR-4478	0,99	2,62	SCARF1	1,00	-4,36

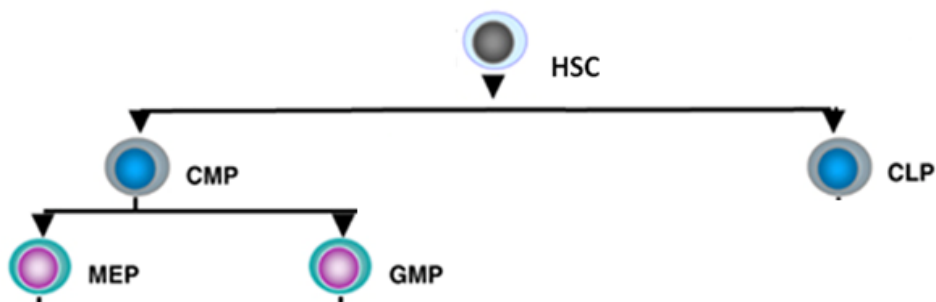
Resposta à molecula de origem bacteriana (p-valor: 5e-4)					
miRNA	Probabilidade posterior do miRNA	logFc do miRNA	Gene alvo	Probabilidade posterior do gene alvo	LogFC do gene alvo
hsa-miR-155-5p	1,00	1,80	CD36	1,00	-5,91
hsa-miR-3934-5p	1,00	4,23	TRIB1	1,00	-3,93
hsa-miR-142-5p	1,00	2,87	MAPK1	1,00	-1,01
hsa-miR-342-5p	1,00	4,77	PLCG2	1,00	-4,04
hsa-miR-342-5p	1,00	4,77	ADM	1,00	-4,64
hsa-miR-95-5p	1,00	10,70	TLR6	1,00	-3,09
hsa-miR-141-3p	1,00	3,74	MAPK14	1,00	-1,40
hsa-miR-99a-5p	0,99	2,49	TRIB1	1,00	-3,93
hsa-miR-99a-5p	0,99	2,49	SERPINE1	0,97	-2,45
hsa-let-7g-5p	0,99	2,27	LYN	1,00	-5,69

Figura 5.18. Vias metabólicas relacionadas à mielócitos e linfócitos. A) Primeira comparação realizada: Linfócitos vs. Mielócitos. B) As três vias metabólicas com menor p-valor relacionadas à comparação.

**Tabela 5.24. As 10 vias metabólicas inibidas com menor p-valor em Megacariócitos e Eritroblatos.**

Via Metabólica	P-valor
Resposta a ácido	2,00E-05
Regulação da produção de interleucina-6	4,00E-05
Regulação do tamanho do tubo	4,00E-05
Regulação do diâmetro de vaso sanguíneo	4,00E-05
Processo vascular no sistema circulatório	6,00E-05
Regulação do tamanho de vaso sanguíneo	6,00E-05
Regulação do diâmetro de tubo	7,00E-05
Resposta celular a lipídeos	7,00E-05
Regulação positiva da produção de citocinas	1,10E-04
Processo do sistema circulatório	1,10E-04

A)



B)

Regulação da produção de interleucina 6 (p-valor: 4e-05)					
miRNA	Probabilidade posterior do miRNA	logFc do miRNA	Gene alvo	Probabilidade posterior do gene alvo	LogFc do gene alvo
hsa-miR-125b-2-3p	0,99	3,39	PTAFR	0,99	-2,58
hsa-miR-335-3p	0,99	5,32	LILRB2	0,99	-5,57
hsa-miR-335-3p	0,99	5,32	TNFAIP3	0,99	-3,53
hsa-miR-335-3p	0,99	5,32	IL6R	0,99	-3,48
hsa-miR-181b-5p	0,99	1,70	ARRB2	1	-2,89
hsa-miR-181b-5p	0,99	1,70	IL1A	0,98	-3,30
hsa-miR-4516	0,98	2,73	MAPK13	0,99	-3,65
hsa-let-7c-5p	0,97	1,55	CEBPB	0,99	-2,97
hsa-let-7c-5p	0,97	1,55	IL10	0,94	-2,48
hsa-miR-106a-5p	0,95	2,26	IL10	0,94	-2,48

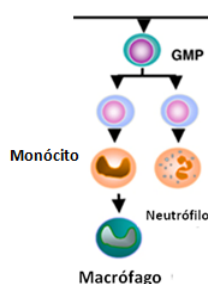
Resposta à molécula de origem bacteriana (p-valor: 1,3e-04) / Resposta inflamatória (p-valor: 1,5e-4)					
miRNA	Probabilidade posterior do miRNA	logFc do miRNA	Gene alvo	Probabilidade posterior do gene alvo	LogFc do gene alvo
hsa-miR-296-3p	0,99	3,82	ICAM1	0,99	-4,49
hsa-miR-296-3p	0,99	3,82	PELI1	0,99	-2,85
hsa-miR-125b-2-3p	0,99	3,39	PTAFR	0,99	-2,58
hsa-miR-335-3p	0,99	5,32	LILRB1	0,99	-5,10
hsa-miR-335-3p	0,99	5,32	LILRB2	0,99	-5,57
hsa-miR-335-3p	0,99	5,32	NOTCH1	0,99	-2,13
hsa-miR-335-3p	0,99	5,32	TNFAIP3	0,99	-3,52
hsa-miR-335-3p	0,99	5,32	FOXP1	0,97	-1,02
hsa-miR-335-3p	0,99	5,32	MAPK14	0,97	-0,76
hsa-miR-195-3p	0,99	2,54	B2M	0,99	-2,34

**Figura 5.19. Vias metabólicas relacionadas à GMP e MEP.** A) Segunda comparação realizada: GMP vs. MEP. B) As três vias metabólicas com menor p-valor relacionadas à comparação.

**Tabela 5.25. As 10 vias metabólicas inibidas com menor p-valor em neutrófilos.**

Via Metabólica	P-valor
Regulação positiva da diferenciação em célula espumosa derivada de macrófago	1,00E-05
Regulação da proliferação de célula epitelial	2,00E-05
Regulação da proteólise de ectodomínio de proteína de membrana	2,00E-05
Regulação da cascata ERK1 e ERK2	3,00E-05
Ativador da atividade receptora	6,00E-05
Pós-sinapse	6,00E-05
Regulador da atividade de receptor	7,00E-05
Regulação da diferenciação de célula mieloide	9,00E-05
Resposta a interferon gama	1,30E-04
Via de sinalização de receptor purínico	1,40E-04

A)



B)

Resposta à interferon gama					
miRNA	Probabilidade posterior do miRNA	logFc do miRNA	Gene alvo	Probabilidade posterior do gene alvo	LogFC do gene alvo
hsa-miR-2115-3p	1,00	3,79	KYNU	1,00	-5,18
hsa-miR-6876-3p	1,00	0,42	HSP90AB1	1,00	-3,12
hsa-miR-26b-5p	1,00	2,18	CYP27B1	1,00	-3,87
hsa-miR-26b-5p	1,00	2,18	SYNCRIP	1,00	-1,34
hsa-miR-26b-5p	1,00	2,18	CCL2	1,00	-5,37
hsa-miR-26b-5p	1,00	2,18	TRIM68	1,00	-1,04
hsa-miR-26b-5p	1,00	2,18	CCL7	1,00	-4,52
hsa-miR-192-5p	1,00	3,27	SYNCRIP	1,00	-1,34
hsa-miR-192-5p	1,00	3,27	PPARG	1,00	-4,06
hsa-miR-192-5p	1,00	3,27	TRIM68	1,00	-1,04

Sinapse imunológica					
miRNA	Probabilidade posterior do miRNA	logFc do miRNA	Gene alvo	Probabilidade posterior do gene alvo	LogFC do gene alvo
hsa-miR-143-5p	1,00	5,34	DUSP3	1,00	-0,89
hsa-miR-26b-5p	1,00	2,18	STOML2	1,00	-1,50
hsa-miR-192-5p	1,00	3,27	ALCAM	1,00	-1,10
hsa-miR-450a-1-3p	1,00	2,21	HAVCR2	1,00	-4,59
hsa-miR-335-5p	1,00	2,02	SCIMP	1,00	-4,17
hsa-miR-582-5p	1,00	2,29	CD81	1,00	-4,09
hsa-miR-16-5p	1,00	1,47	PDCC6IP	0,91	-0,37
hsa-miR-30c-1-3p	1,00	0,79	HAVCR2	1,00	-4,59
hsa-miR-101-5p	0,99	2,39	DUSP3	1,00	-0,89
hsa-miR-424-5p	0,98	1,67	SOCS6	1,00	-5,89

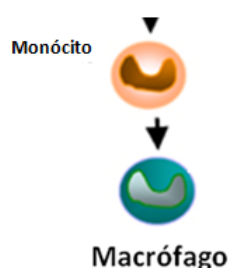
Migração de neutrófilos					
miRNA	Probabilidade posterior do miRNA	logFc do miRNA	Gene alvo	Probabilidade posterior do gene alvo	LogFC do gene alvo
hsa-miR-26b-5p	1,00	2,18	CCL2	1,00	-5,37
hsa-miR-26b-5p	1,00	2,18	CCL7	1,00	-4,52
hsa-miR-450a-1-3p	1,00	2,21	CCL22	1,00	-11,01
hsa-miR-335-5p	1,00	2,02	CXCL3	1,00	-2,07
hsa-miR-335-5p	1,00	2,02	CCL24	0,97	-4,89
hsa-miR-335-5p	1,00	2,02	IL1A	0,96	-2,06
hsa-miR-30c-1-3p	1,00	0,79	CCL22	1,00	-11,01
hsa-miR-181d-5p	1,00	1,25	CCL22	1,00	-11,01
hsa-miR-181d-5p	1,00	1,25	IL1A	0,96	-2,06
hsa-miR-142-3p	1,00	1,96	IL1A	0,96	-2,06

**Figura 5.20. Vias metabólicas relacionadas à neutrófilos e monócitos.** A) Terceira comparação realizada: Neutrófilos vs. Monócitos. B) As três vias metabólicas com menor p-valor relacionadas à comparação.

**Tabela 5.26. As 10 vias metabólicas com menor p-valor em monócitos.**

Vias Metabólicas	P-valor
Regulação positiva da movimentação de componente celular	7,00E-05
Regulação do transporte de lipídeos	1,20E-04
Inibidos da atividade de lípase	2,50E-04
Transdução de sinal por fosforilação de proteína	3,10E-04
Junção aderente	4,50E-04
Cascata MAPK	4,60E-04
Ligação a fator de crescimento	4,90E-04
Ligação a ácido graxo	5,20E-04
Regulação positiva de mobilidade celular	5,50E-04
Regulação positiva de migração celular	5,80E-04

**A)**



**B)**

Regulação Positiva de remodelação tecidual					
miRNA	Probabilidade posterior do miRNA	logFc do miRNA	Gene alvo	Probabilidade posterior do gene alvo	LogFC do gene alvo
hsa-miR-223-3p	0,98	1,60	TMEM64	0,94	-0,67
hsa-miR-335-5p	0,92	1,42	SPP1	1	-9,41
hsa-miR-335-5p	0,92	1,42	DCSTAMP	1	-5,72

Atividade de citocinas					
miRNA	Probabilidade posterior do miRNA	logFc do miRNA	Gene alvo	Probabilidade posterior do gene alvo	LogFC do gene alvo
hsa-miR-223-3p	0,98	1,60	CXCL2	0,99	3,79
hsa-miR-25-3p	0,93	1,00	TXLNA	0,99	-0,83
hsa-miR-335-5p	0,92	1,42	SPP1	1,00	-9,41
hsa-miR-335-5p	0,92	1,42	NRP1	0,99	-3,69
hsa-miR-335-5p	0,92	1,42	PPBP	0,99	5,15
hsa-miR-335-5p	0,92	1,42	CXCL2	0,991	3,79
hsa-miR-335-5p	0,92	1,42	NDP	0,99	-6,87
hsa-miR-335-5p	0,92	1,42	GDF15	0,98	-2,93
hsa-miR-335-5p	0,92	1,42	CXCL3	0,98	2,45

Produção de citocinas					
miRNA	Probabilidade posterior do miRNA	logFc do miRNA	Gene alvo	Probabilidade posterior do gene alvo	LogFC do gene alvo
hsa-miR-150-5p	0,98	2,93	S1PR3	0,99	3,51
hsa-miR-150-5p	0,98	2,93	EREG	0,99	6,40
hsa-miR-335-5p	0,92	1,42	EREG	0,99	6,40
hsa-miR-335-5p	0,92	1,42	LGALS9	0,99	-0,81

**Figura 5.21. Vias metabólicas relacionadas à monócitos e macrófagos.** A) Quarta comparação realizada: Monócitos vs. Macrófagos. B) As três vias metabólicas com menor p-valor relacionadas à comparação.

## 6. DISCUSSÃO

Os RNAs não codificadores pequenos (sncRNAs) são moléculas fundamentais para a manutenção do metabolismo. Estas moléculas estão envolvidas no controle da expressão gênica atuando na transcrição e tradução de mRNAs e, conseqüentemente, modulando o crescimento, multiplicação e diferenciação celular. Assim, diversos grupos de pesquisa relatam alterações no padrão de expressão destas moléculas em diferentes doenças como diabetes, câncer, Alzheimer, entre outras (Calin et al. 2002).

Atualmente, um dos métodos mais utilizados para avaliar o perfil de expressão de sncRNAs é o sequenciamento de alta vazão (HTS). A principal vantagem desta técnica em relação às anteriores reside no fato dela não utilizar sondas específicas e, portanto, não requerer conhecimento prévio das moléculas expressas. Outras duas grandes vantagens são a maior vazão e sensibilidade, permitindo a detecção de moléculas antes desconhecidas e moléculas pouco expressas (Meyerson et al. 2010).

Neste trabalho, utilizamos dados públicos produzidos pela tecnologia de HTS com duas finalidades: a detecção de snoRNAs e piRNAs diferencialmente expressos em pacientes fumantes e não fumantes de adenocarcinoma de pulmão; e na identificação do perfil de expressão de miRNAs em diferentes células sanguíneas.

### **6.1. sncRNAs diferencialmente expressos entre não fumantes e fumantes com adenocarcinoma de pulmão.**

O fumo expõe o indivíduo à diversas substâncias carcinogênicas, sendo considerado um dos maiores fatores de risco para o desenvolvimento do câncer de pulmão (Gumus et al. 2008) Contudo, alguns pacientes de câncer de pulmão nunca fizeram uso de tabaco, indicando que outros fatores podem estar envolvidos na aquisição desta doença (Huang et al. 2015).

Os snoRNAs e piRNAs estão diretamente relacionados à manutenção do metabolismo celular através da modificação de rRNAs e do controle de elementos transponíveis, respectivamente (Thorenoor e Slaby 2015, Le Thomas et al. 2014). Estudos recentes sugerem que estas moléculas também podem atuar diretamente

no controle da expressão de diversos genes, através de mecanismos de RNA de interferência (Ross et al. 2014, Falaleeva et al. 2017). Desta forma, o estudo destas moléculas é importante.

A fim de validar nossa metodologia, comparamos nossos resultados com os obtidos por Kim e colaboradores (2013). Ao todo, obtivemos 23 miRNAs diferencialmente expressos, dos quais 20 estão de acordo com os dados originais. Os outros quatro com expressão alterada miRNAs encontrados por Kim e colaboradores (2013), a saber, hsa-miR-139-5p, hsa-miR-127-3p, hsa-miR-134-5p e hsa-miR-889-3p, também foram encontrados em nossas análises porém foram removidos por não atingirem o “fold change” mínimo estabelecido.

Entre os outros três miRNAs encontrados em nossas análise mas não pelos autores do artigo original, todos foram relatados alterados em câncer de pulmão. O hsa-miR-301b também foi identificado como mais expresso em câncer de pulmão por outros autores (Edmonds e Eischen 2014; Wu et al. 2016) e relacionado com invasão de tecidos através da inibição de TP63 em câncer de pâncreas (Funamizu 2014). O hsa-miR-451a é um gene supressor tumoral que atua inibindo o gene *ESDN/DCBLD2* e inibindo a migração celular e invasão (Fukumoto et al. 2014) e intensifica a atuação do quimioterápico Tamoxifen através do recrutamento de macrófagos (Liu et al. 2015). Este miRNA também foi identificado como menos expresso em câncer de pulmão (Li et al. 2017), gástrico (Su et al. 2015) e carcinoma escamoso hipofaríngeo (Fukumoto et al. 2014). Em nosso trabalho, o hsa-miR-449b-5p foi identificado como mais expresso nas amostras tumorais, porém os estudos de Jeon e colaboradores (2012) e Yang e colaboradores (2009) relatam este miRNA menos expresso em linhagens de câncer de pulmão.

Após confirmada nossa metodologia, procedemos para análises exploratórias do perfil de expressão de snoRNAs e piRNAs em não fumantes e fumantes. As análises de componente principal e agrupamento hierárquico demonstraram que o perfil de expressão de snoRNAs e piRNAs é distinto entre não fumantes e fumantes para os dois tipos de amostras analisadas, tecido tumoral (Figura 5.5) e normal adjacente ao tumor (Figura 5.2). Estes dados indicam que pacientes não fumantes e fumantes possuem perfis de expressão distintos de snoRNAs e piRNAs.

Ao todo, nós encontramos 49 snoRNAs diferencialmente expressos entre as amostras normais de fumantes e não fumantes, e 55 snoRNAs ou piRNAs diferencialmente expressos entre as amostras tumorais de não fumantes e



fumantes. Entre os sncRNAs com padrão de expressão alterada, 28 estão presentes em ambas as análises e demonstrando a mesma tendência de expressão, como demonstrado pela análise de correlação de Pearson (Figura 5.8). O fato dos 28 genes estarem alterados de forma semelhante entre não fumantes e fumantes e em ambas as amostras normal e tumoral indicam que estas alterações devem ter ocorrido devido ao fumo e não devido a tumorigênese. Entre os 28 sncRNAs encontrados diferencialmente expressos em ambas análises, 11 apresentaram expressão constante entre as amostras normais e tumorais do mesmo grupo de tabagismo, reforçando que as alterações encontradas estão ligadas ao uso de tabaco.

Entre os 11 snoRNAs encontrados alterados em ambas análises, somente U60 (SNORD60) e U63 (SNORD63) estão mais expressos em amostras de pacientes não fumantes. Já foi demonstrado que U60 possui um papel na atenuação da vasoconstrição em ratos (Burghuber et al. 1986) e na regulação do colesterol da membrana plasmática (Brandis et al. 2013). Por sua vez, o U63 está localizado em uma região frequentemente deletada em mielodisplasias (Graubert et al. 2009). Assim, acreditamos que o fumo possa inibir a expressão de snoRNAs necessários para a manutenção celular.

Os outros nove snoRNAs restantes foram encontrados mais expressos em amostras de fumantes, sendo muitos destes já relatados como alterados em outros tipos de câncer. Um exemplo é o HBII-142 (SNORD66). Este snoRNAs é localizado na região 3q27 do genoma que está frequentemente amplificada em tumores (Thorenor e Slaby 2015). Este snoRNA também já foi sugerido como biomarcador por estar mais expresso no plasma de pacientes de câncer de pulmão (Liao et al. 2010). Outro snoRNA, U30 (SNORD30) também foi relatado como mais expresso em gliomas pediátricos (Jha et al. 2015) e níveis altos de expressão deste snoRNA estão relacionados com progressão mais rápida em mielomas múltiplos (Lopez-Corral et al. 2012). O HBI-100 (SCARNA3) foi relatado mais expresso em câncer de mama (Liang et al. 2015) e a super expressão de U59B (SNORD59B) correlacionado com tempo de sobrevida maior que dois anos em linfoma de células T (Valleron et al. 2012). Estes resultados sugerem que o uso de tabaco altera o perfil de expressão das células de tecido normal de forma semelhante ao encontrado em diversos tumores.

Diversos snoRNAs encontrados alterados em nosso estudo não possuem alvos conhecidos. Este é o caso dos snoRNAs HBII-420 (SNORD99) e ACA44 (SNORA44), ambos encontrados mais expressos em fumantes. Lan e colaboradores (2017) avaliaram a expressão destes snoRNAs em relação ao seu gene hospedeiro, o lncRNA SNHG12. Os autores relatam que o silenciamento do SNHG12 não altera os níveis de expressão dos snoRNA, indicando que estes genes possuem regiões promotoras próprias. Assim, apesar de já serem conhecidos há muito tempo e bastante estudados, ainda há informações a se descobrir sobre os snoRNAs. Estes estudos adicionais podem revelar novos meios de atuação e novos alvos e, conseqüentemente, revelar novas vias metabólicas afetadas pelo fumo.

O snoRNA U44 (SNORD44) foi encontrado mais expresso nas amostras normais de não fumantes quando comparadas com fumantes e constitutivamente expresso nas amostras normais e tumorais de não fumantes. Curiosamente, este snoRNA é frequentemente utilizado para normalização em experimentos de RT-qPCR (Krell et al. 2014). Nossos dados indicam que o uso deste snoRNA pode inserir erros caso analisadas amostras de não fumantes e fumantes. Alterações similares no padrão de expressão deste snoRNA já foram relatadas em câncer colorretal e de mama (Gee et al. 2011; Krell et al. 2014).

Ao todo, 33 snoRNAs foram encontrados constitutivamente expressos nas amostras normais e tumorais em não fumantes e fumantes (Tabela 4.1). Destes, 16 não apresentam diferenças significativas entre amostras normais de não fumantes e fumantes e amostras tumorais de fumantes e não fumantes, indicando que seus níveis de expressão não são alterados com o uso de tabaco. Entre eles, o U43 (SNORD43) também é frequentemente usado como parâmetro para normalização para experimentos envolvendo o perfil de expressão de miRNAs devido aos seus níveis estáveis, reforçando nossos achados (Patnaik et al. 2010). Este snoRNA também não apresentou alterações em comparações realizadas entre as amostras normais e tumorais de não fumantes e entre normais e tumorais de fumantes. Outro snoRNA encontrado constitutivamente expresso em amostras de não-fumantes e fumantes foi o U50 (SNORD50A). Este snoRNA foi recentemente relatado como capaz de interagir com sítios de poliadenilação na região 3' de mRNAs, alterando o perfil de adenilação e os níveis de expressão de diferentes transcritos (Huang et al. 2017). Portanto, a expressão deste snoRNA é um candidato à parâmetro para normalização de experimentos envolvendo amostras de pulmão. Estudos adicionais

deste e de outros snoRNAs com outros tecidos podem revelar outros candidatos para parâmetros de normalização.

Os piRNAs e as proteínas PIWI atuam em conjunto para controlar elementos transponíveis e manter a integridade do genoma (Le Thomas et al. 2014). Diferentes proteínas PIWI já foram encontradas alteradas em câncer de pulmão e associadas com a sobrevivência do paciente (Moisés et al. 2015; Navarro et al. 2015; Qu et al. 2015a). Especula-se que a expressão aberrante de elementos transponíveis aumente a taxa de deleções, rearranjos e duplicações frequentemente observadas nos genoma de células cancerosas (Qu et al. 2015b). Neste estudo, encontramos somente dois piRNAs diferencialmente expressos entre amostras tumorais de não fumantes e fumantes, hsa-piR-001168-4 e hsa-piR-010894-3, ambos mais expressos em não fumantes. Assim, o uso de tabaco parece estar relacionado a inibição destes piRNAs e contribuir para aumentar o número de alterações encontradas no genoma de células cancerosas causadas pela atuação desregulada de elementos transponíveis.

Tem aumentado consideravelmente a quantidade de pesquisas envolvendo piRNAs e snoRNAs em câncer. Apesar de serem conhecidos há mais tempo, encontramos apenas oito artigos relacionando snoRNAs e câncer publicados em 2004. A maior quantidade de artigos sobre os temas foi publicada em 2015 com 55 artigos. Por outro lado, desde 2004 até 2016, o número de publicações envolvendo piRNAs e câncer aumentou em 2,5 vezes, passando de 696 em 2004 para 1730 em 2016. Contudo, a quantidade de artigos investigando as mudanças no perfil de expressão destas moléculas ainda pouca e, até o momento, não encontramos artigos que avaliassem o perfil de expressão de snoRNAs e piRNAs entre não fumantes e fumantes.

A maioria dos estudos envolvendo amostras de pulmão indica a quantidade de indivíduos não fumantes e fumantes, porém não distingue os dois grupos no momento da análise (Mei et al. 2012; Xia et al. 2013; Chen et al. 2015; Zheng et al. 2015; Zhu et al. 2016). Porém, nosso trabalho e outros dois (Wang e Li 2010; Minicã et al. 2017) tem mostrado que o perfil de expressão de diversas moléculas é alterado entre esses dois grupos. Desta forma, consideramos que futuros estudos deveriam separar amostras provenientes de fumantes e não fumantes e analisá-los separadamente a fim de evitar a introdução de alterações causadas pelo uso de tabaco e não pela patologia estudada.

## 6.2. miRNAs diferencialmente expressos na árvore hematopoiética.

Os miRNAs é a família de sncRNAs mais estudada e eles atuam impedindo a tradução de mRNAs, diminuindo assim os níveis intracelulares da proteína correspondente. O mecanismo de regulação da expressão gênica por miRNAs está presente em diversos processo celulares tais como a replicação, ativação e diferenciação celular. Neste trabalho, utilizamos a tecnologia de HTS para avaliar o perfil de expressão de miRNAs entre diferentes tipos celulares ao longo da árvore hematopoiética (Figura 1.8).

Diversos grupos de pesquisa relataram o papel de diferentes miRNAs na diferenciação e ativação de linfócitos T. Em nossa abordagem, encontramos 83 miRNAs diferencialmente expressos entre linfócitos T CD4+ e os outros tipos celulares avaliados (Anexo 6). Os nível de expressão do miRNA hsa-miR-181a é um dos que mais se altera ao longo da diferenciação de linfócitos no timo (Belver et al. 2011). Neilson e colaboradores (2007) indicam que este miRNA está mais expresso em células T duplo positivas<sup>5</sup> quando comparado com células T duplo negativas ou a linfócitos T CD4 maduros. Isto ocorre porque a expressão do hsa-miR-181 está diretamente relacionada com a sensibilidade do receptor TCR, pois o aumento de sua expressão em células maduras diminui os níveis de expressão da fosfatase DUSP6 que por sua vez controla a ativação de ERK. A enzima ERK é responsável pela fosforilação da serina 59 de LCK, permitindo a transdução de sinal após a ativação do receptor TCR (Acuto et al. 2008). Desse modo o aumento dos níveis de miR-181 aumentam a sensibilidade do receptor TCR, enquanto a diminuição em células imaturas reduziu esta sensibilidade (Li et al. 2007). Em resumo, durante a seleção positiva, este miRNA é expresso a fim de identificar os linfócitos que possuem alta sensibilidade a antígenos, em seguida, sua expressão é reduzida para que antígenos fracos não seja capazes de desencadear uma resposta imune (Sonkoly et al. 2008). Em nossos resultados, três membros da famílias do miR-181 foram encontrados mais expressos: o miR-181a-5p apresentou probabilidade posterior de 0,97 e logFC de 1,94; o hsa-miR-181a-2-3p apresentou probabilidade

---

<sup>5</sup> Linfócitos duplo positivos são aqueles que apresentam ambos os receptores CD4 e CD8. Os duplo negativos não apresentam nenhum dos receptores (Abbas et al. 2015).

posterior de 0,95 e logFC de 1,85; e o hsa-miR-181b-2-3p apresentou probabilidade posterior de 0,97 e logFC de 2,52. Os membros desta família atuam inibindo diversas fosfatases, tais como DUSP5, DUSP6, SHP2 e PTPN22, que inibem a via de sinalização do receptor de antígenos de célula T (TCR) (Belver et al. 2011).

Os genes *DUSP5* e *SHP2* foram encontrados menos expressos em linfócitos, porém apresentaram probabilidade posterior igual a 0,9 (“fold change” 0,19 e 0,18, respectivamente), e por isso não foram considerados na análise de vias metabólicas enriquecidas. O gene *PTPN22* foi encontrado mais expresso em linfócitos, porém também não apresentou probabilidade posterior acima do limite estipulado para a análise de vias metabólicas (probabilidade posterior: 0,50 e logFC:0,47) e o gene *DUSP6* foi encontrado menos expresso e com probabilidade posterior alta, sendo considerado nas análises de enriquecimento de vias (probabilidade posterior: 1 e logFC: -4,15).

O miRNA miR-92 também já foi relatado mais expresso em linfócitos duplo positivos (Neilson et al. 2007). Neste estudo, os miRNAs hsa-miR-92a-1-5p e hsa-miR-92a-3p também se apresentaram mais expressos (probabilidade posterior de 0,85 e 0,80 e logFC de 2,05 e 1,39, respectivamente). Outro miRNA encontrado mais expresso em linfócitos T CD4+ pela nossa abordagem é o hsa-miR-142-5p que apresentou probabilidade posterior de 0,99 e logFC de 2,87. A transfecção deste miRNA em camundongos levou a um aumento de 30 a 40% na quantidade de linfócitos T circulantes, não afetando a produção de linfócitos B (Chen 2004). A ativação dos linfócitos T CD4+ em Th1 e Th17 também está relacionada com a expressão do miR-155. Camundongos “knockout” para miR-155 apresentam níveis reduzidos de interleucina 2 (IL-2), interleucina 17 (IL-17) e interferon- $\gamma$  (INF-  $\gamma$ ) (Rodriguez et al. 2007). O hsa-miR-155 atua inibindo o gene *SOCS1*, um inibidor da IL-2 (O’Connell et al. 2010). Em nossas análises, também encontramos o hsa-miR-155-5p diferencialmente mais expresso (probabilidade posterior: 0,99 e logFC: 1,79). O gene *SOCS1* não apresentou probabilidade posterior acima de 0,9 e também não foi incluído nas análises de GSEA. Este gene foi encontrado mais expresso em linfócitos (logFC: 0,64).

O hsa-miR-150 já foi relacionado a diferenciação de linfócitos T duplo positivos em linfócitos CD4+ ou CD8+, sendo encontrado mais expresso neste tipo celular (Neilson et al. 2007). Este miRNA atua inibindo o gene *NOTCH3* que é um receptor envolvido na sobrevivência, proliferação e diferenciação de células T

(Bellavia et al. 2003). Xiao e colaboradores (2007) demonstraram que o miR-150 e o gene *NOTCH3* possuem níveis de expressão inversos. Neste trabalho, encontramos ambas formas maduras do hsa-miR-150 mais expressas em células T CD4+ (hsa-miR-150-3p: probabilidade posterior: 0,99 e logFC: 5,54; hsa-miR-150-5p: probabilidade posterior: 0,99; logFC: 7,65). O gene *NOTCH3* foi encontrado menos expresso em linfócitos e foi incluído na análise de vias metabólicas enriquecidas (probabilidade posterior: 0,99 e logFC: -2,25).

O hsa-miR-150 já foi relatado como responsável por direcionar a diferenciação das células precursoras de eritroblatos e megacariócitos para a linhagem megacariocítica, diminuindo o número de eritroblatos (Bissels et al. 2012). Em nossas análises, o hsa-miR-150-5p apresentou probabilidade posterior de 0,83 e loFC de -2,47. Apesar de termos optado por unir as células da série vermelha, megacariócitos e eritroblatos, em um único grupo para realizar as comparações, o comportamento encontrado assemelha-se ao relatado em megacariócitos. Isso se justifica porque possuímos mais amostras de megacariócitos do que eritroblatos.

Outro miRNA cujo comportamento se altera ao longo da árvore hematopoética é o miR-223. Níveis intracelulares altos deste miRNA estimulam a diferenciação em granulócitos, tais como neutrófilos, enquanto os níveis baixos de miRNA estimulam a diferenciação em eritroblatos e monócitos (Utikal et al. 2015). Em nossas análises, o miRNA hsa-miR-223-3p foi encontrado em todos os tipos celulares derivados do progenitor mielóide comum, apresentando logFC de -1,82 em eritroblatos e megacariócitos, de 2,83 em neutrófilos e de 1,6 em monócitos. De acordo com Johnnidis e colaboradores (2008), este miRNA atua inibindo a expressão do gene *Mef2c*, um fator de transcrição que estimula a proliferação de progenitores mielóides. Por sua vez, a expressão do miR-223 é controlada por dois fatores de transcrição: *NFI-A*, que induz baixos níveis de expressão do miRNA, e *C/EBP $\alpha$*  que induz maior expressão de miR-223, por sua vez, também inibe *NFI-A* (Fazi et al. 2005).

A comparação entre neutrófilos e monócitos e macrófagos foi a que mais identificou genes diferencialmente expressos. Ao todo identificamos 244 miRNAs alterados, sendo que 123 encontravam-se mais expresso em neutrófilos. Muitos destes miRNAs já foram relatados pela literatura, como o hsa-miR-424-5p (probabilidade posterior: 0,98 e logFC: 1,66) e que atua inibindo o gene *NF1-A*, um inibidor da diferenciação de neutrófilos e granulócitos (Undi et al. 2013). Outros

miRNAs já relatado mais expresso em neutrófilos são o hsa-miR-29b-1-5p (probabilidade posterior 0,98 e logFC: 1,08), o hsa-miR-19a-3p (probabilidade posterior: 0,92 e logFC: 1.64), o hsa-miR-19b-3p (probabilidade posterior: 0,917 e logFC: 1,29). O miRNA miR-146a inibe o gene MUC5AC e, portanto, atua inibindo a secreção de muco. Em nossas análises, este gene foi encontrado menos expresso em neutrófilos (probabilidade posterior: 0,99 e logFC: -2,72) (Undi et al. 2013).

O miRNA hsa-miR-21-3p foi encontrado menos expresso em neutrófilos (logFC: -1,45), este miRNA é inibido pelo gene Gfi1, cuja expressão é necessária para a diferenciação de granulócitos e monócitos. Os possíveis alvos do hsa-miR-21 em neutrófilos ainda não foram identificados (Sayed e Abdellatif 2011).

A penúltima comparação realizada foi entre monócitos e macrófagos. O aumento da expressão do miRNA-146a já foi relacionado com o aumento da atividade fagocítica em monócitos e com a inibição da produção de citocinas inflamatórias (Pauley et al. 2011). Os monócitos analisados neste trabalho não haviam sido ativados e por isso encontramos este miRNA menos expresso em monócitos em nossas análises (probabilidade posterior: 0,99 e logFC: -3,7). O mesmo ocorreu com o miRNA-155-5p, encontrado menos expresso em monócitos (logFC: -2.85). A expressão deste miRNA se eleva em macrófagos após sua ativação por IFN- $\beta$  ou TNF- $\alpha$ . Entre os genes inibidos por este miRNA encontram-se I $\kappa$ B quinase epsilon (IKK $\epsilon$ ), domínio de morte associado ao Fas (FADD), e a super família de receptores TNFR (Sayed e Abdellatif 2011). O membro TNFRSF21 desta super família foi encontrado menos expresso nos monócitos e incluído na análise de GSEA (probabilidade posterior: 0,99 e logFC: -3,04).

Apesar da análise dos mRNAs diferencialmente expressos não ser o alvo principal deste trabalho, notamos que nas comparações entre GMP e MEP, neutrófilos e monócitos, monócitos e macrófagos e macrófagos M1 e macrófagos M2 foram encontrados RNAs não codificadores longos (lncRNA). A interação entre estes dois tipos de RNAs tem sido bastante estudada e revelam que tanto os miRNAs podem desestabilizar lncRNAs, como os lncRNAs podem agir como chamariz para miRNAs (Gorospe 2014).

Apesar de ser uma família extensivamente estudada, poucos alvos de miRNA foram validados experimentalmente. Isto ocorre devido a grande quantidade de miRNAs e possíveis alvos, tornando a determinação de alvos por experimentos de biologia molecular uma atividade laboriosa. Desta forma, poucos alvos de poucos

miRNAs foram validados experimentalmente. Assim, diversos grupos de pesquisa criaram diferentes programas de predição de alvos de miRNAs, cada um considerando diferentes aspectos da biologia de miRNAs, tais como o pareamento com alvos validados, o tamanho da região *seed*, conservação do par miRNA-alvo em outras espécies, entre outros. Porém, atualmente, as predições realizadas por estes programas não possuem muito pares miRNA-alvo em comum.

Dada esta incerteza, optamos por utilizar o banco de dados miRTarBase que possui somente interações miRNA-alvos validados experimentalmente. Ainda assim, nossa abordagem foi capaz de identificar mais de 800 alvos em quase todas as comparações realizadas, exceto entre Macrófagos M1 e M2. Esta última, por envolver dois tipos celulares muito similares, é esperado que apresentem poucos genes, miRNAs e mRNAs, diferencialmente expressos entre eles. Outra consequência dos poucos miRNAs diferencialmente expressos e poucos alvos encontrados na comparação entre macrófagos M2 e macrófagos M1 é o fato de não encontrarmos nenhuma via metabólica relacionada a miRNAs alterada. Todas as outras comparações revelaram vias metabólicas reprimidas por miRNAs relacionadas aos tipos celulares estudados.

Na comparação entre linfócitos e mielócitos, foram encontrados 152 vias alteradas, entre elas vias como regulação positiva da resposta a ferimentos (p-valor 3,9e-04) que é característica de células mielóides; regulação de coagulação sanguínea (p-valor 5,8e-04) que está relacionadas às plaquetas; regulação da via de sinalização de receptor Toll (p-valor 2,43) que não está presente em linfócitos; reconhecimento para fagocitose (p-valor 3,93) característico de células fagocíticas tais como macrófagos e neutrófilos; morfogênese de tecidos, que está relacionada ao papel de macrófagos (p-valor 1,17e-03), diferenciação de eritroblatos (p-valor: 3,15e-3), e diferenciação de granulócitos tais como neutrófilos (p-valor: 8,2e-3) e regulação da produção do fator de estímulo de formação de colônia de macrófagos granulócitos, que ocorre durante a proliferação e diferenciação de macrófagos (p-valor: 9,52e-03) apresentaram-se inibidas nos linfócitos T CD4+ indiferenciados quando comparados aos outros tipos celulares. Estas vias indicam a inibição de diferenciação em linfócitos B, eritroblatos, megacariócitos e granulócitos, tais como neutrófilos.

Na segunda comparação, envolvendo eritroblatos e megacarioócitos e granulócitos e monócitos, foram encontrados 173 vias metabólicas alteradas. As vias



encontradas também estão relacionadas aos outros tipos celulares estudados, tais como regulação da produção de interleucina-6 por monócitos e macrófagos (p-valor: 4e-05), resposta celular à molécula de origem bacteriana, relacionada à fagócitos (p-valor: 1,56e-03), regulação da produção do fator de necrose tumoral, presente em macrófagos e monócitos (p-valor: 1,69e-03), regeneração, relacionada à macrófagos e monócitos (p-valor 1,69e-03), regulação positiva do processo biosintético de espécies reativa à oxigênio que ocorre em neutrófilos (p-valor: 1,92e-03), regulação positiva da migração de leucócitos (p-valor: 3,74e-03) que está ativada em macrófagos e neutrófilos, ativação de linfócitos promovida por macrófagos (p-valor: 3,74e-03), ativação de granulócitos, tais como neutrófilos (p-valor: 4,05e-3).

Muitas vias relacionadas aos macrófagos foram encontradas inibidas na comparação entre neutrófilos e monócitos e macrófagos. Exemplos destas vias encontradas são resposta à interferon gama (p-valor 1,3e-04), sinapse imunológica (p-valor 2,2e-04), resposta celular à interferon gama (p-valor 1,66e-03), regulação de ativação de linfócito (p-valor 3,24e-03), regulação da citotoxicidade mediada por célula natural killer (p-valor 4,98e-03), ligação de antígeno (p-valor 7,01e-03) e migração de granulócito (p-valor 9,77e-03). Vias metabólicas diretamente relacionadas aos neutrófilos também foram encontradas nesta análise como migração (p-valor: 1,05e-03) e quimiotaxia (p-valor: 1,16e-03), sugerindo um papel para os miRNAs na migração de neutrófilos também.

A comparação entre monócitos e macrófagos também identificou vias metabólicas alteradas relacionadas à atividade de macrófagos, tais como regulação positiva de remodelação tecidual (p-valor: 1,48e-0), atividade de citocinas (p-valor: 1,77e-03) e produção de citocinas (p-valor: 3,13e-03).

Em conclusão, avaliamos nesta parte do estudo os miRNAs diferencialmente expressos entre diferentes tipos celulares hematopoéticos. Muitos dos miRNAs encontrados por nós já foram identificados por outros autores como necessários para a diferenciação e ativação das células estudadas. Com esta informação, procuramos por alvos validados diferencialmente menos expressos nestes mesmos tipos celulares a fim de identificar as vias metabólicas reprimidas por miRNAs. A maioria das comparações revelou vias metabólicas relacionadas aos outros tipos celulares estudados, indicando uma inibição da diferenciação destas células e, conseqüentemente, promoção da diferenciação do tipo celular estudado. Nossos

resultados reforçam os diferentes relatos da importância de miRNAs na diferenciação, ativação e atuação de diferentes tipos células.

## **7. CONCLUSÕES**

### **7.1. snoRNAs e piRNAs diferencialmente expressos entre não fumantes e fumantes com adenocarcinoma**

- Nosso estudo indica que o fumo modifica a expressão de diversos sncRNAs, alterando o perfil de expressão de diversos sncRNAs para um perfil similar ao encontrado em diversos tipos de câncer.
- Os snoRNAs encontrados diferencialmente expressos neste estudo podem auxiliar na distinção de amostras de não fumantes e fumantes.
- Nossos dados também apontam diferentes snoRNAs que devem ser considerados ao escolher o gene para atuar como parâmetro de normalização de expressão em diferentes experimentos.
- Grupos de pesquisa devem considerar examinar separadamente as amostras de não fumantes e fumantes a fim de evitar a introdução de fatores de confusão.
- Mais estudos envolvendo snoRNAs e piRNAs são necessários para a melhor compreensão do metabolismo destas moléculas e de seu papel no metabolismo celular e em diferentes patologias.

### **7.2. miRNAs diferencialmente expressos na árvore hematopoiética**

- microRNAs são fundamentais para a diferenciação e ativação de células do sistema hematopoiético
- A expressão de miRNAs ao longo do processo de diferenciação de diversas células do sistema hematopoiético é regulada por diversos

fatores de transcrição e inibidores a fim de garantir concentrações intracelulares adequadas para aquele tipo celular.

- miRNAs inibem diversos genes que promovem a diferenciação de diversos tipos celulares, auxiliando o processo de decisão final da diferenciação.

## REFERÊNCIAS BIBLIOGRÁFICAS

- Abbas AK, Lichtman AH, Pillai S. *Imunologia Celular e Molecular*. 8<sup>o</sup> ed. Rio de Janeiro: Elsevier; 2015.
- Abbott A. Europe to map the human epigenome. *Nature*. 2011;477(7366):518–518.
- Acuto O, Bartolo V Di, Michel F. Tailoring T-cell receptor signals by proximal negative feedback mechanisms. *Nat Rev Immunol*. 2008;8(9):699–712.
- Andrews S. FastQC A Quality Control tool for High Throughput Sequence Data. <http://www.bioinformatics.babraham.ac.uk/projects/fastqc/>. 2010.
- Apellaniz-Ruiz M, Gallego C, Ruiz-Pinto S, Carracedo A, Rodríguez-Antona C. Human genetics: international projects and personalized medicine. *Drug Metab Pers Ther*. 2016;31(1).
- Aravin AA, Sachidanandam R, Bourc'his D, Schaefer C, Pezic D, Toth KF, et al. A piRNA Pathway Primed by Individual Transposons Is Linked to De Novo DNA Methylation in Mice. *Mol Cell*. 2008;31(6):785–99.
- Auton A, Abecasis GR, Altshuler DM, Durbin RM, Abecasis GR, Bentley DR, et al. A global reference for human genetic variation. *Nature*. 2015;526(7571):68–74.
- Aziz N, Zhao Q, Bry L, Driscoll DK, Funke B, Gibson JS, et al. College of American Pathologists' Laboratory Standards for Next-Generation Sequencing Clinical Tests. *Arch Pathol Lab Med*. 2015;139(4):481–93.
- Baltimore D, Boldin MP, O'Connell RM, Rao DS, Taganov KD. MicroRNAs: new regulators of immune cell development and function. *Nat Immunol*. 2008;9(8):839–45.
- Bellavia D, Campese AF, Vacca A, Gulino A, Screpanti I. Notch3, another Notch in T cell development. *Semin Immunol*. 2003;15(2):107–12.
- Belver L, Papavasiliou FN, Ramiro AR. MicroRNA control of lymphocyte differentiation and function. *Curr Opin Immunol*. 2011;23(3):368–73.
- Bialous SA, Sarna L. Lung Cancer and Tobacco. *Nurs Clin North Am*. 2017;52(1):53–63.
- Bissels U, Bosio A, Wagner W. MicroRNAs are shaping the hematopoietic landscape. *Haematologica*. 2012;97(2):160–7.
- Brandis KA, Gale S, Jinn S, Langmade SJ, Dudley-Rucker N, Jiang H, et al. Box C/D Small Nucleolar RNA (snoRNA) U60 Regulates Intracellular Cholesterol Trafficking. *J Biol Chem*. 2013;288(50):35703–13.
- Burghuber OC, Strife R, Zirolli J, Mathias MM, Murphy RC, Reeves JT, et al. Hydrogen peroxide induced pulmonary vasoconstriction in isolated rat lungs is attenuated by U60,257, a leucotriene synthesis blocker. *Wien Klin Wochenschr*. 1986;98(4):117–9.
- Busch J, Ralla B, Jung M, Wotschofsky Z, Trujillo-Arribas E, Schwabe P, et al. Piwi-interacting RNAs as novel prognostic markers in clear cell renal cell carcinomas. *J Exp Clin Cancer Res*. 2015;34(1):61.
- Calin GA, Dumitru CD, Shimizu M, Bichi R, Zupo S, Noch E, et al. Frequent deletions and down-regulation of micro- RNA genes miR15 and miR16 at 13q14 in chronic lymphocytic leukemia. *Proc Natl Acad Sci U S A*. 2002;99(24):15524–9.
- Cao C, Chen J, Lyu C, Yu J, Zhao W, Wang Y, et al. Bioinformatics Analysis of the Effects of Tobacco Smoke on Gene Expression. te Pas MFW, organizador. *PLoS One*. 2015;10(12):e0143377.
- Chen C-Z. MicroRNAs Modulate Hematopoietic Lineage Differentiation. *Science* (80-). 2004;303(5654):83–6.
- Chen S, Li P, Li J, Wang Y, Du Y, Chen X, et al. MiR-144 Inhibits Proliferation and Induces Apoptosis and Autophagy in Lung Cancer Cells by Targeting TIGAR.

- Cell Physiol Biochem. 2015;35(3):997–1007.
- Cheng J, Guo J-M, Xiao B-X, Miao Y, Jiang Z, Zhou H, et al. piRNA, the new non-coding RNA, is aberrantly expressed in human cancer cells. *Clin Chim Acta*. 2011;412(17–18):1621–5.
- Chervitz SA, Deutsch EW, Field D, Parkinson H, Quackenbush J, Rocca-Serra P, et al. Data Standards for Omics Data: The Basis of Data Sharing and Reuse. In 2011. p. 31–69.
- Chu A, Robertson G, Brooks D, Mungall AJ, Birol I, Coope R, et al. Large-scale profiling of microRNAs for The Cancer Genome Atlas. *Nucleic Acids Res*. 2016;44(1):e3–e3.
- Collisson EA, Campbell JD, Brooks AN, Berger AH, Lee W, Chmielecki J, et al. Comprehensive molecular profiling of lung adenocarcinoma. *Nature*. 2014;511(7511):543–50.
- Darzacq X. Cajal body-specific small nuclear RNAs: a novel class of 2'-O-methylation and pseudouridylation guide RNAs. *EMBO J*. 2002;21(11):2746–56.
- Daugaard I, Venø MT, Yan Y, Kjeldsen TE, Lamy P, Hager H, et al. Small RNA sequencing reveals metastasis-related microRNAs in lung adenocarcinoma. *Oncotarget*. 2017;
- Delamarre L. Differential Lysosomal Proteolysis in Antigen-Presenting Cells Determines Antigen Fate. *Science* (80- ). 2005;307(5715):1630–4.
- Dong X-Y, Guo P, Boyd J, Sun X, Li Q, Zhou W, et al. Implication of snoRNA U50 in human breast cancer. *J Genet Genomics*. 2009;36(8):447–54.
- Dong X-Y, Rodriguez C, Guo P, Sun X, Talbot JT, Zhou W, et al. SnoRNA U50 is a candidate tumor-suppressor gene at 6q14.3 with a mutation associated with clinically significant prostate cancer. *Hum Mol Genet*. 2008;17(7):1031–42.
- Edmonds MD, Eischen CM. Differences in miRNA Expression in Early Stage Lung Adenocarcinomas that Did and Did Not Relapse. Hoheisel JD, organizador. *PLoS One*. 2014;9(7):e101802.
- Eisenberg E, Levanon EY. Human housekeeping genes, revisited. *Trends Genet*. 2013;29(10):569–74.
- ENCODE Project Consortium, Birney E, Stamatoyannopoulos J a, Dutta A, Guigó R, Gingeras TR, et al. Identification and analysis of functional elements in 1% of the human genome by the ENCODE pilot project. *Nature*. 2007;447(7146):799–816.
- Ender C, Krek A, Friedländer MR, Beitzinger M, Weinmann L, Chen W, et al. A Human snoRNA with MicroRNA-Like Functions. *Mol Cell*. 2008;32(4):519–28.
- Esteller M. Non-coding RNAs in human disease. *Nat Rev Genet*. 2011;12(12):861–74.
- Falaleeva M, Stamm S. Processing of snoRNAs as a new source of regulatory non-coding RNAs. *BioEssays*. 2013;35(1):46–54.
- Falaleeva M, Welden JR, Duncan MJ, Stamm S. C/D-box snoRNAs form methylating and non-methylating ribonucleoprotein complexes: Old dogs show new tricks. *BioEssays*. 2017;1600264.
- Fazi F, Rosa A, Fatica A, Gelmetti V, De Marchis ML, Nervi C, et al. A Minicircuitry Comprised of MicroRNA-223 and Transcription Factors NFI-A and C/EBP? Regulates Human Granulopoiesis. *Cell*. 2005;123(5):819–31.
- Fiedler K, Brunner C. Mechanisms Controlling Hematopoiesis. In: *Hematology - Science and Practice*. InTech; 2012.
- Fu A, Jacobs DI, Hoffman AE, Zheng T, Zhu Y. PIWI-interacting RNA 021285 is involved in breast tumorigenesis possibly by remodeling the cancer epigenome. *Carcinogenesis*. 2015;36(10):1094–102.
- Fukumoto I, Kinoshita T, Hanazawa T, Kikkawa N, Chiyomaru T, Enokida H, et al.

- Identification of tumour suppressive microRNA-451a in hypopharyngeal squamous cell carcinoma based on microRNA expression signature. *Br J Cancer*. 2014;111(2):386–94.
- Funamizu N. MicroRNA-301b promotes cell invasiveness through targeting TP63 in pancreatic carcinoma cells. *Int J Oncol*. 2014;
- Gall JG. Timeline: The centennial of the Cajal body. *Nat Rev Mol Cell Biol*. 2003;4(12):975–80.
- Gao L, Ma J, Mannoor K, Guarnera MA, Shetty A, Zhan M, et al. Genome-wide small nucleolar RNA expression analysis of lung cancer by next-generation deep sequencing. *Int J Cancer*. 2015;136(6):E623–9.
- Gee HE, Buffa FM, Camps C, Ramachandran A, Leek R, Taylor M, et al. The small-nucleolar RNAs commonly used for microRNA normalisation correlate with tumour pathology and prognosis. *Br J Cancer*. 2011;104(7):1168–77.
- Gentleman RC, Carey VJ, Bates DM, Bolstad B, Dettling M, Dudoit S, et al. Bioconductor: open software development for computational biology and bioinformatics. *Genome Biol*. 2004;5(10):R80.
- Gonzalez-Porta M, Calvo M, Sammeth M, Guigo R. Estimation of alternative splicing variability in human populations. *Genome Res*. 2012;22(3):528–38.
- Graubert TA, Payton MA, Shao J, Walgren RA, Monahan RS, Frater JL, et al. Integrated Genomic Analysis Implicates Haploinsufficiency of Multiple Chromosome 5q31.2 Genes in De Novo Myelodysplastic Syndromes Pathogenesis. Schrijver I, organizador. *PLoS One*. 2009;4(2):e4583.
- Griffiths-Jones S. Rfam: annotating non-coding RNAs in complete genomes. *Nucleic Acids Res*. 2004;33(Database issue):D121–4.
- Grimson A, Farh KK-H, Johnston WK, Garrett-Engele P, Lim LP, Bartel DP. MicroRNA targeting specificity in mammals: determinants beyond seed pairing. *Mol Cell*. 2007;27(1):91–105.
- Gumus ZH, Du B, Kacker A, Boyle JO, Bocker JM, Mukherjee P, et al. Effects of Tobacco Smoke on Gene Expression and Cellular Pathways in a Cellular Model of Oral Leukoplakia. *Cancer Prev Res*. 2008;1(2):100–11.
- Gyoba J, Shan S, Roa W, Bédard ELR. Diagnosing Lung Cancers through Examination of Micro-RNA Biomarkers in Blood, Plasma, Serum and Sputum: A Review and Summary of Current Literature. *Int J Mol Sci*. 2016;17(4).
- Han Li C, Chen Y. Small and Long Non-Coding RNAs: Novel Targets in Perspective Cancer Therapy. *Curr Genomics*. 2015;16(5):319–26.
- He Y, Jiang X, Chen J. The role of miR-150 in normal and malignant hematopoiesis. *Oncogene*. 2014;33(30):3887–93.
- Herbst RS, Heymach J V, Lippman SM. Lung cancer. *N Engl J Med*. 2008;359(13):1367–80.
- Hercus C. Novoalign: a short read aligner with qualities. 2008.
- Holley RW, Apgar J, Everett GA, Madison JT, Marquisee M, Merrill SH, et al. Structure of a Ribonucleic Acid. *Science (80- )*. 1965;147(3664):1462–5.
- Hsu S-D, Tseng Y-T, Shrestha S, Lin Y-L, Khaleel A, Chou C-H, et al. miRTarBase update 2014: an information resource for experimentally validated miRNA-target interactions. *Nucleic Acids Res*. 2014;42(Database issue):D78-85.
- Huang C, Shi J, Guo Y, Huang W, Huang S, Ming S, et al. A snoRNA modulates mRNA 3' end processing and regulates the expression of a subset of mRNAs. *Nucleic Acids Res*. 2017;45(15):8647–8660.
- Huang J-Y, Jian Z-H, Nfor ON, Ku W-Y, Ko P-C, Lung C-C, et al. The effects of pulmonary diseases on histologic types of lung cancer in both sexes: a population-based study in Taiwan. *BMC Cancer*. 2015;15:834.

- Huber W, Carey VJ, Gentleman R, Anders S, Carlson M, Carvalho BS, et al. Orchestrating high-throughput genomic analysis with Bioconductor. *Nat Methods*. 2015;12(2):115–21.
- Inamura K. Major Tumor Suppressor and Oncogenic Non-Coding RNAs: Clinical Relevance in Lung Cancer. *Cells*. 2017;6(2):12.
- Italiano JE, Lecine P, Shivdasani RA, Hartwig JH. Blood Platelets Are Assembled Principally at the Ends of Proplatelet Processes Produced by Differentiated Megakaryocytes. *J Cell Biol*. 1999;147(6):1299–312.
- Jacobs L, Nawrot TS, de Geus B, Meeusen R, Degraeuwe B, Bernard A, et al. Subclinical responses in healthy cyclists briefly exposed to traffic-related air pollution: an intervention study. *Environ Heal*. 2010;9(1):64.
- Jeon HS, Lee SY, Lee EJ, Yun SC, Cha EJ, Choi E, et al. Combining microRNA-449a/b with a HDAC inhibitor has a synergistic effect on growth arrest in lung cancer. *Lung Cancer*. 2012;76(2):171–6.
- Jha P, Agrawal R, Pathak P, Kumar A, Purkait S, Mallik S, et al. Genome-wide small noncoding RNA profiling of pediatric high-grade gliomas reveals deregulation of several miRNAs, identifies downregulation of snoRNA cluster HBII-52 and delineates H3F3A and TP53 mutant-specific miRNAs and snoRNAs. *Int J Cancer*. 2015;137(10):2343–53.
- Johnnidis JB, Harris MH, Wheeler RT, Stehling-Sun S, Lam MH, Kirak O, et al. Regulation of progenitor cell proliferation and granulocyte function by microRNA-223. *Nature*. 2008;451(7182):1125–9.
- Johnson SM, Grosshans H, Shingara J, Byrom M, Jarvis R, Cheng A, et al. RAS Is Regulated by the let-7 MicroRNA Family. *Cell*. 2005;120(5):635–47.
- Jorge NAN, Ferreira CG, Passeti F. Bioinformatics of Cancer ncRNA in High Throughput Sequencing: Present State and Challenges. *Front Genet*. 2012;3(December):287.
- Joyce DP, Kerin MJ, Dwyer RM. Exosome-encapsulated microRNAs as circulating biomarkers for breast cancer. *Int J cancer*. 2016;
- Kim SCS, Jung YY, Park JJ, Cho S, Seo C, Kim JJJ, et al. A high-dimensional, deep-sequencing study of lung adenocarcinoma in female never-smokers. *PLoS One*. 2013;8(2):e55596.
- Kishore S, Stamm S. The snoRNA HBII-52 regulates alternative splicing of the serotonin receptor 2C. *Science*. 2006;311(5758):230–2.
- Kozomara A, Griffiths-Jones S. miRBase: integrating microRNA annotation and deep-sequencing data. *Nucleic Acids Res*. 2011;39(Database issue):D152-7.
- Krell J, Frampton AE, Mirnezami R, Harding V, De Giorgio A, Roca Alonso L, et al. Growth Arrest-Specific Transcript 5 Associated snoRNA Levels Are Related to p53 Expression and DNA Damage in Colorectal Cancer. Calogero RA, organizador. *PLoS One*. 2014;9(6):e98561.
- Krishnan AR, Korrapati A, Zou AE, Qu Y, Wang XQ, Califano JA, et al. Smoking status regulates a novel panel of PIWI-interacting RNAs in head and neck squamous cell carcinoma. *Oral Oncol*. 2017;65:68–75.
- Kröger W, Mapiye D, Entfellner J-BD, Tiffin N. A meta-analysis of public microarray data identifies gene regulatory pathways deregulated in peripheral blood mononuclear cells from individuals with Systemic Lupus Erythematosus compared to those without. *BMC Med Genomics*. 2016;9(1):66.
- Krueger F. *Trim Gallore!* 2012.
- Kunz M, Wolf B, Schulze H, Atlán D, Walles T, Walles H, et al. Non-Coding RNAs in Lung Cancer: Contribution of Bioinformatics Analysis to the Development of Non-Invasive Diagnostic Tools. *Genes (Basel)*. 2016;8(1):8.



- Lan T, Ma W, Hong Z, Wu L, Chen X, Yuan Y. Long non-coding RNA small nucleolar RNA host gene 12 (SNHG12) promotes tumorigenesis and metastasis by targeting miR-199a/b-5p in hepatocellular carcinoma. *J Exp Clin Cancer Res*. 2017;36(1):11.
- Lander ES, Linton LM, Birren B, Nusbaum C, Zody MC, Baldwin J, et al. Initial sequencing and analysis of the human genome. *Nature*. 2001;409(6822):860–921.
- Larrea E, Sole C, Manterola L, Goicoechea I, Armesto M, Arestin M, et al. New Concepts in Cancer Biomarkers: Circulating miRNAs in Liquid Biopsies. *Int J Mol Sci*. 2016;17(5):627.
- Lee H-C, Gu W, Shirayama M, Youngman E, Conte D, Mello CC. C. elegans piRNAs Mediate the Genome-wide Surveillance of Germline Transcripts. *Cell*. 2012;150(1):78–87.
- Lee Y, Kim M, Han J, Yeom K-H, Lee S, Baek SH, et al. MicroRNA genes are transcribed by RNA polymerase II. *EMBO J*. 2004;23(20):4051–60.
- Letelier P, Riquelme I, Hernández AH, Guzmán N, Farías JG, Roa JC. Circulating MicroRNAs as Biomarkers in Biliary Tract Cancers. *Int J Mol Sci*. 2016;17(5).
- Lewis BP, Shih I, Jones-Rhoades MW, Bartel DP, Burge CB. Prediction of mammalian microRNA targets. *Cell*. 2003;115(7):787–98.
- Li C, Yin Y, Liu X, Xi X, Xue W, Qu Y. Non-small cell lung cancer associated microRNA expression signature: integrated bioinformatics analysis, validation and clinical significance. *Oncotarget*. 2017;
- Li L, Zhu T, Gao Y-F, Zheng W, Wang C-J, Xiao L, et al. Targeting DNA Damage Response in the Radio(Chemo)therapy of Non-Small Cell Lung Cancer. *Int J Mol Sci*. 2016;17(6):839.
- Li Q-J, Chau J, Ebert PJR, Sylvester G, Min H, Liu G, et al. miR-181a Is an Intrinsic Modulator of T Cell Sensitivity and Selection. *Cell*. 2007;129(1):147–61.
- Li Z, Rana TM. Therapeutic targeting of microRNAs: current status and future challenges. *Nat Rev Drug Discov*. 2014;13(8):622–38.
- Liang F, Qu H, Lin Q, Yang Y, Ruan X, Zhang B, et al. Molecular biomarkers screened by next-generation RNA sequencing for non-sentinel lymph node status prediction in breast cancer patients with metastatic sentinel lymph nodes. *World J Surg Oncol*. 2015;13(1):258.
- Liao J, Yu L, Mei Y, Guarnera M, Shen J, Li R, et al. Small nucleolar RNA signatures as biomarkers for non-small-cell lung cancer. *Mol Cancer*. 2010;9:198.
- Liu Z, Miao T, Feng T, Jiang Z, Li M, Zhou L, et al. miR-451a Inhibited Cell Proliferation and Enhanced Tamoxifen Sensitive in Breast Cancer via Macrophage Migration Inhibitory Factor. *Biomed Res Int*. 2015;2015:1–12.
- Liu Z, Yang G, Zhao T, Cao G, Xiong L, Xia W, et al. Small ncRNA expression and regulation under hypoxia in neural progenitor cells. *Cell Mol Neurobiol*. 2011;31(1):1–5.
- Lonsdale J, Thomas J, Salvatore M, Phillips R, Lo E, Shad S, et al. The Genotype-Tissue Expression (GTEx) project. *Nat Genet*. 2013;45(6):580–5.
- Lopez-Corral L, Mateos M V., Corchete LA, Sarasquete ME, de la Rubia J, de Arriba F, et al. Genomic analysis of high-risk smoldering multiple myeloma. *Haematologica*. 2012;97(9):1439–43.
- Lu C, Huang X, Zhang X, Roensch K, Cao Q, Nakayama KI, et al. miR-221 and miR-155 regulate human dendritic cell development, apoptosis, and IL-12 production through targeting of p27kip1, KPC1, and SOCS-1. *Blood*. 2011;117(16):4293–303.
- Lund E. Nuclear Export of MicroRNA Precursors. *Science (80- )*. 2004;303(5654):95–

- 8.
- Mannoor K, Liao J, Jiang F. Small nucleolar RNAs in cancer. *Biochim Biophys Acta*. 2012;1826(1):121–8.
- Martens JHA, Stunnenberg HG. BLUEPRINT: mapping human blood cell epigenomes. *Haematologica*. 2013;98(10):1487–9.
- Martin M. Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet Journal*. 2011;17:10–2.
- Mei Y-P, Liao J-P, Shen J, Yu L, Liu B-L, Liu L, et al. Small nucleolar RNA 42 acts as an oncogene in lung tumorigenesis. *Oncogene*. 2012;31(22):2794–804.
- Mendes R, Carreira B, Baptista P V., Fernandes AR. Non-small cell lung cancer biomarkers and targeted therapy - two faces of the same coin fostered by nanotechnology. *Expert Rev Precis Med Drug Dev*. 2016;1(2):155–68.
- Meyerson M, Gabriel S, Getz G. Advances in understanding cancer genomes through second-generation sequencing. *Nat Rev Genet*. 2010;11(10):685–96.
- Michel CI, Holley CL, Scruggs BS, Sidhu R, Brookheart RT, Listenberger LL, et al. Small nucleolar RNAs U32a, U33, and U35a are critical mediators of metabolic stress. *Cell Metab*. 2011;14(1):33–44.
- Minicã CC, Mbarek H, Pool R, Dolan C V, Boomsma DI, Vink JM. Pathways to smoking behaviours: biological insights from the Tobacco and Genetics Consortium meta-analysis. *Mol Psychiatry*. 2017;22(1):82–8.
- Moisés J, Navarro A, Tejero R, Viñolas N, Cordeiro A, Marrades RM, et al. PIWI proteins as prognostic markers in non small cell lung cancer. *Eur Respir J*. 2015;46(suppl 59).
- Müller S, Raulefs S, Bruns P, Afonso-Grunz F, Plötner A, Thermann R, et al. Next-generation sequencing reveals novel differentially regulated mRNAs, lncRNAs, miRNAs, sdRNAs and a piRNA in pancreatic cancer. *Mol Cancer*. 2015;14(1):94.
- National Cancer Institute. NCI Dictionary of Cancer Terms. 16.09. Bethesda, MD; 2016.
- Navarro A, Tejero R, Viñolas N, Cordeiro A, Marrades RM, Fuster D, et al. The significance of PIWI family expression in human lung embryogenesis and non-small cell lung cancer. *Oncotarget*. 2015;6(31):31544–56.
- Neilson JR, Zheng GXY, Burge CB, Sharp PA. Dynamic regulation of miRNA expression in ordered stages of cellular development. *Genes Dev*. 2007;21(5):578–89.
- O'Connell RM, Kahn D, Gibson WSJ, Round JL, Scholz RL, Chaudhuri AA, et al. MicroRNA-155 Promotes Autoimmune Inflammation by Enhancing Inflammatory T Cell Development. *Immunity*. 2010;33(4):607–19.
- O'Connell RM, Zhao JL, Rao DS. MicroRNA function in myeloid biology. *Blood*. 2011;118(11):2960–9.
- Patnaik SK, Kannisto E, Knudsen S, Yendamuri S. Evaluation of MicroRNA Expression Profiles That May Predict Recurrence of Localized Stage I Non-Small Cell Lung Cancer after Surgical Resection. *Cancer Res*. 2010;70(1):36–45.
- Pauley KM, Stewart CM, Gauna AE, Dupre LC, Kuklani R, Chan AL, et al. Altered miR-146a expression in Sjögren's syndrome and its functional role in innate immunity. *Eur J Immunol*. 2011;41(7):2029–39.
- Peng L, Song L, Liu C, Lv X, Li X, Jie J, et al. piR-55490 inhibits the growth of lung carcinoma by suppressing mTOR signaling. *Tumor Biol*. 2016;37(2):2749–56.
- Perez-Riverol Y, Alpi E, Wang R, Hermjakob H, Vizcaíno JA. Making proteomics data accessible and reusable: Current state of proteomics databases and repositories. *Proteomics*. 2015;15(5–6):930–50.

- Pierigè F, Serafini S, Rossi L, Magnani M. Cell-based drug delivery. *Adv Drug Deliv Rev.* 2008;60(2):286–95.
- Purtell L, Qi Y, Campbell L, Sainsbury A, Herzog H. Adult-onset deletion of the Prader-Willi syndrome susceptibility gene *Snord116* in mice results in reduced feeding and increased fat mass. *Transl Pediatr.* 2017;6(2):88–97.
- Qu X, Liu J, Zhong X, Li X, Zhang Q. PIWIL2 promotes progression of non-small cell lung cancer by inducing CDK2 and Cyclin A expression. *J Transl Med.* 2015a;13(1):301.
- Qu X, Liu J, Zhong X, Li X, Zhang Q, Siegel R, et al. PIWIL2 promotes progression of non-small cell lung cancer by inducing CDK2 and Cyclin A expression. *J Transl Med.* 2015b;13(1):301.
- Quinlan AR, Hall IM. BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics.* 2010;26(6):841–2.
- Rabellino EM. Human megakaryocytes. II. Expression of platelet proteins in early marrow megakaryocytes. *J Exp Med.* 1981;154(1):88–100.
- Reuter M, Berninger P, Chuma S, Shah H, Hosokawa M, Funaya C, et al. Miwi catalysis is required for piRNA amplification-independent LINE1 transposon silencing. *Nature.* 2011;480(7376):264–7.
- Roberts L. Controversial From the Start. *Science (80- ).* 2001;291(5507).
- Robinson MD, McCarthy DJ, Smyth GK. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics.* 2010;26(1):139–40.
- Rodriguez A, Vigorito E, Clare S, Warren M V., Couttet P, Soond DR, et al. Requirement of bic/microRNA-155 for Normal Immune Function. *Science (80- ).* 2007;316(5824):608–11.
- Rosenbloom KR, Armstrong J, Barber GP, Casper J, Clawson H, Diekhans M, et al. The UCSC Genome Browser database: 2015 update. *Nucleic Acids Res.* 2015;43(Database issue):D670-81.
- Roy S, LaFramboise WA, Nikiforov YE, Nikiforova MN, Routbort MJ, Pfeifer J, et al. Next-Generation Sequencing Informatics: Challenges and Strategies for Implementation in a Clinical Environment. *Arch Pathol Lab Med.* 2016;140(9):958–75.
- Rung J, Brazma A. Reuse of public genome-wide gene expression data. *Nat Rev Genet.* 2012;14(2):89–99.
- Saalberg Y, Wolff M. VOC breath biomarkers in lung cancer. *Clin Chim Acta.* 2016;459:5–9.
- Sai Lakshmi S, Agrawal S. piRNABank: a web resource on classified and clustered Piwi-interacting RNAs. *Nucleic Acids Res.* 2008;36(Database issue):D173-7.
- Saus E, Brunet-Vega A, Iraola-Guzmán S, Pegueroles C, Gabaldón T, Pericay C. Long Non-Coding RNAs As Potential Novel Prognostic Biomarkers in Colorectal Cancer. *Front Genet.* 2016;7:54.
- Sayed D, Abdellatif M. MicroRNAs in Development and Disease. *Physiol Rev.* 2011;91(3):827–87.
- Schmutz J, Wheeler J, Grimwood J, Dickson M, Yang J, Caoile C, et al. Quality assessment of the human genome sequence. *Nature.* 2004;429(6990):365–8.
- Siomi MC, Sato K, Pezic D, Aravin AA. PIWI-interacting small RNAs: the vanguard of genome defence. *Nat Rev Mol Cell Biol.* 2011;12(4):246–58.
- Sonkoly E, St?hle M, Pivarcsi A. MicroRNAs and immunity: Novel players in the regulation of normal immune function and inflammation. *Semin Cancer Biol.* 2008;18(2):131–40.
- Spira A, Beane J, Shah V, Liu G, Schembri F, Yang X, et al. Effects of cigarette

- smoke on the human airway epithelial cell transcriptome. *Proc Natl Acad Sci.* 2004;101(27):10143–8.
- Srivastava A, Sarkar H, Gupta N, Patro R. RapMap: a rapid, sensitive and accurate tool for mapping RNA-seq reads to transcriptomes. *Bioinformatics.* 2016;32(12):i192–200.
- Stepanov GA, Filippova JA, Komissarov AB, Kuligina E V, Richter VA, Semenov D V. Regulatory role of small nucleolar RNAs in human diseases. *Biomed Res Int.* 2015;2015:206849.
- Su Z, Zhao J, Rong Z, Geng W, Wang Z. MiR-451, a potential prognostic biomarker and tumor suppressor for gastric cancer. *Int J Clin Exp Pathol.* 2015;8(8):9154–60.
- Taft R, Pang K, Mercer T, Dinger ME, Mattick JS. Non-coding RNAs: regulators of disease. *J Pathol.* 2010;220:126–39.
- Tainsky M. Genomic and proteomic biomarkers for cancer: a multitude of opportunities. ... *Biophys Acta (BBA)-Reviews Cancer.* 2009;1796(2):176–93.
- Le Thomas A, Tóth K, Aravin AA, Ghildiyal M, Zamore P, Malone C, et al. To be or not to be a piRNA: genomic origin and processing of piRNAs. *Genome Biol.* 2014;15(1):204.
- Thorenor N, Slaby O. Small nucleolar RNAs functioning and potential roles in cancer. *Tumor Biol.* 2015;36(1):41–53.
- Tomczak K, Czerwińska P, Wiznerowicz M. Review The Cancer Genome Atlas (TCGA): an immeasurable source of knowledge. *Współczesna Onkol.* 2015;1A:68–77.
- Turro E, Astle WJ, Tavar? S. Flexible analysis of RNA-seq data using mixed effects models. *Bioinformatics.* 2014;30(2):180–8.
- Turro E, Su S-Y, Gonçalves Â, Coin LJM, Richardson S, Lewin A. Haplotype and isoform specific expression estimation using multi-mapping RNA-seq reads. *Genome Biol.* 2011;12(2):R13.
- Undi RB, Kandi R, Gutti RK. MicroRNAs as Haematopoiesis Regulators. *Adv Hematol.* 2013;2013:1–20.
- Utikal J, Abba M, Novak D, Moniuszko M, Allgayer H. Function and significance of MicroRNAs in benign and malignant human stem cells. *Semin Cancer Biol.* 2015;35:200–11.
- Valleron W, Ysebaert L, Berquet L, Fataccioli V, Quelen C, Martin A, et al. Small nucleolar RNA expression profiling identifies potential prognostic markers in peripheral T-cell lymphoma. *Blood.* 2012;120(19):3997–4005.
- Vavalà T, Mariniello A, Reale ML, Novello S. Gender differences in lung cancer. *Ital J Gender-Specific Med.* 2016;2(3):99–109.
- Venter JC, Adams MD, Myers EW, Li PW, Mural RJ, Sutton GG, et al. The Sequence of the Human Genome. *Science (80- ).* 2001;291(5507):1304–51.
- Wang G, Wang R, Strulovici-Barel Y, Salit J, Staudt MR, Ahmed J, et al. Persistence of Smoking-Induced Dysregulation of MiRNA Expression in the Small Airway Epithelium Despite Smoking Cessation. *Yildirim AO, organizador. PLoS One.* 2015;10(4):e0120824.
- Wang J, Li MD. Common and Unique Biological Pathways Associated with Smoking Initiation/Progression, Nicotine Dependence, and Smoking Cessation. *Neuropsychopharmacology.* 2010;35(3):702–19.
- Watson JD. The human genome project: past, present, and future. *Science.* 1990;248(4951):44–9.
- Weinstock GM. ENCODE: more genomic empowerment. *Genome Res.* 2007;17(6):667–8.

- Witten D, Tibshirani R, Gu SG, Fire A, Lui W-O. Ultra-high throughput sequencing-based small RNA discovery and discrete statistical biomarker analysis in a collection of cervical tumours and matched controls. *BMC Biol.* 2010;8:58.
- Wu D, Chen B, Cui F, He X, Wang W, Wang M. Hypoxia-induced microRNA-301b regulates apoptosis by targeting Bim in lung cancer. *Cell Prolif.* 2016;49(4):476–83.
- Xia Y, Chen Q, Zhong Z, Xu C, Wu C, Liu B, et al. Down-Regulation of MiR-30c Promotes the Invasion of Non-Small Cell Lung Cancer by Targeting MTA1. *Cell Physiol Biochem.* 2013;32(2):476–85.
- Xiao C, Calado DP, Galler G, Thai T-H, Patterson HC, Wang J, et al. MiR-150 Controls B Cell Differentiation by Targeting the Transcription Factor c-Myb. *Cell.* 2007;131(1):146–59.
- Xue W, Dahlman JE, Tammela T, Khan OF, Sood S, Dave A, et al. Small RNA combination therapy for lung cancer. *Proc Natl Acad Sci U S A.* 2014;111(34):E3553-61.
- Yang X, Feng M, Jiang X, Wu Z, Li Z, Aau M, et al. miR-449a and miR-449b are direct transcriptional targets of E2F1 and negatively regulate pRb-E2F1 activity through a feedback loop by targeting CDK6 and CDC25A. *Genes Dev.* 2009;23(20):2388–93.
- Zheng D, Zhang J, Ni J, Luo J, Wang J, Tang L, et al. Small nucleolar RNA 78 promotes the tumorigenesis in non-small cell lung cancer. *J Exp Clin Cancer Res.* 2015;34(1):49.
- Zhou L, Li X, Liu Q, Zhao F, Wu J. Small RNA transcriptome investigation based on next-generation sequencing technology. *J Genet Genomics.* 2011;38(11):505–13.
- Zhu W, Zhou K, Zha Y, Chen D, He J, Ma H, et al. Diagnostic Value of Serum miR-182, miR-183, miR-210, and miR-126 Levels in Patients with Early-Stage Non-Small Cell Lung Cancer. Zheng SG, organizador. *PLoS One.* 2016;11(4):e0153046.

## **ANEXOS**

# ANEXO 1: ARTIGO PUBLICADO

RESEARCH ARTICLE

## snoRNA and piRNA expression levels modified by tobacco use in women with lung adenocarcinoma

Natasha Andressa Nogueira Jorge<sup>1</sup>, Gabriel Wajnberg<sup>1</sup>, Carlos Gil Ferreira<sup>2</sup>, Benilton de Sa Carvalho<sup>3</sup>, Fabio Passetti<sup>1\*</sup>

**1** Laboratory of Functional Genomics and Bioinformatics, Oswaldo Cruz Institute, Fundação Oswaldo Cruz, Rio de Janeiro, RJ, Brazil, **2** D'Or Institute for Research and Education, Rio de Janeiro, RJ, Brazil, **3** Department of Statistics, State University of Campinas, Campinas, SP, Brazil

\* [passetti@fiocruz.br](mailto:passetti@fiocruz.br)



### Abstract

Lung cancer is one of the most frequent types of cancer worldwide. Most patients are diagnosed at advanced stage and thus have poor prognosis. Smoking is a risk factor for lung cancer, however most smokers do not develop lung cancer while 20% of women with lung adenocarcinoma are non-smokers. Therefore, it is possible that these two groups present differences besides the smoking status, including differences in their gene expression signature. The altered expression patterns of non-coding RNAs in complex diseases make them potential biomarkers for diagnosis and treatment. We analyzed data from differentially and constitutively expressed PIWI-interacting RNAs and small nucleolar RNAs from publicly available small RNA high-throughput sequencing data in search of an expression pattern of non-coding RNA that could differentiate these two groups. Here, we report two sets of differentially expressed small non-coding RNAs identified in normal and tumoral tissues of women with lung adenocarcinoma, that discriminate between smokers and non-smokers. Our findings may offer new insights on metabolic alterations caused by tobacco and may be used for early diagnosis of lung cancer.

### OPEN ACCESS

**Citation:** Nogueira Jorge NA, Wajnberg G, Ferreira CG, de Sa Carvalho B, Passetti F (2017) snoRNA and piRNA expression levels modified by tobacco use in women with lung adenocarcinoma. *PLoS ONE* 12(8): e0183410. <https://doi.org/10.1371/journal.pone.0183410>

**Editor:** Bernard Mari, Institut de Pharmacologie Moléculaire et Cellulaire, FRANCE

**Received:** April 27, 2017

**Accepted:** August 3, 2017

**Published:** August 17, 2017

**Copyright:** © 2017 Nogueira Jorge et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

**Data Availability Statement:** All relevant data are within the paper and its Supporting Information files.

**Funding:** NANJ and GW are supported by Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES). FP acknowledges the support of Fundação Carlos Chagas Filho de Amparo à Pesquisa do Estado do Rio de Janeiro (FAPERJ) and Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq). BSC is supported by Fundação de Amparo à

### Introduction

Lung cancer is one of the leading causes of death from cancer in both men and women worldwide [1]. The most common type of lung cancer is non-small-cell lung cancer (NSCLC), accounting for 85% of the cases [2]. Lung cancer is often detected when the disease is at an advanced clinical stage and thus has a poor prognosis [3–5] and with a high mortality rate.

Smoking is one of the risk factors for this disease [6]. There are many well-known carcinogens in tobacco that bind to DNA and create somatic mutations such as the ones observed in the KRAS gene in lung cancer and in TP53 in several cancer types [7]. However, most smokers do not develop lung cancer while 20% of women with lung adenocarcinoma are non-smokers [8]. Therefore, other factors than smoking status may contribute to the development of lung cancer [9].

Pesquisa do Estado de São Paulo (FAPESP 13/00506-1). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

**Competing interests:** The authors have declared that no competing interests exist.

On the other hand, cigarette smoke is known to change gene expression in the transcriptome of MSK-Leuk1 cell line [6], epithelial cells [10] and buccal mucosa [11] and has been reported to alter biological pathways related to signal transduction, asthma and cell proliferation [6,12,13]. Thus, many studies have searched for differences in gene expression that may reveal changes in the genetic profile between smokers and non-smokers [6, 12] but none have evaluated the expression patterns of small nucleolar RNA (snoRNA) or PIWI-interacting RNA (piRNA).

piRNA ranges from 26 to 31 nucleotides long and is one of the least investigated classes of sncRNAs, and many aspects of its biogenesis and mechanism are still unknown. The major role of piRNA is to silence transposable elements [14,15] in germline cells (reviewed by [12, 13]), thus acting in stem cell division, apoptosis, epigenetic control of transposons and telomeres, and translational control [16]. However, piRNAs are also expressed in somatic tissues [17–19], where they act as gene expression regulators by inducing histone modifications and DNA methylation [20,21]. Importantly, these molecules are altered in several types of cancer [22,19,23,24]. Thus, a better understanding of the expression pattern of these molecules could contribute to lung cancer biology, early detection and survival.

On the other hand, snoRNAs are an abundant sncRNA class that comprises 60 to 300 nucleotides long sncRNA molecules, which associate with the nucleolar enzymes methylase and pseudouridine synthase to form the small ribonucleoprotein responsible for rRNA methylation and pseudouridylation, respectively [25]. However, some snoRNAs do not have an identified target rRNA. Different groups have been reporting 20 to 24 nucleotides long small RNAs derived from the further processing of snoRNAs. These new RNAs seems to act as miRNAs [26] suggesting other functions for these molecules [27]. Several snoRNAs have been found to be altered under hypoxia [28] and oxidative stress conditions [29]. Additionally, SNORA42 is amplified and up-regulated in NSCLC and its levels inversely correlate with survival in NSCLC [30].

One of the most common approaches to study sncRNA is to produce a large-scale profile with techniques like microarray, then validate the findings using strategies such as RT-qPCR [31]. However, these approaches require the use of probes or primers, which means that some sequence fragments must be known *a priori*, making the identification of truly novel genes a complicated task. Recently, the high-throughput sequencing (HTS) technology was used to study sncRNA. The technology is highly precise [32] and sensitive, allowing for the inference at gene expression level, detection of chromosome rearrangements, mutations, novel transcripts, isoforms, and low expressed genes besides the identification of novel genes [33,34]. For instance, Müller and collaborators [24] used high-throughput sequencing technology to evaluate the coding and non-coding transcriptome of six pancreatic cancer samples. Besides several deregulated miRNAs, the authors found that the snoRNAs HBII-296B and U104 as well as piRNA piR-017061 were differentially expressed in tumors when compared to normal cells. However, the role of piRNAs and snoRNAs in cancer is still unknown and warrants further investigation.

There are many commercially available HTS platforms, each with its specifications, such as throughput, read size, error frequency, cost, and the number of sequenced reads [32]. The vast amount of data produced can be stored in public databases and made available to the global scientific community. Different datasets can be combined, after quality control, normalization, and evaluation of the information regarding each experiment, to improve the detection of weaker signals and generate new knowledge, without the burden of one single research center generating all data. Rung and Brazma [35] performed a survey on how often public data from ArrayExpress were mentioned in published papers in 2011 and found that almost one in four papers used data that were available in that public database to answer new biological questions



different from the question raised in the original study that collected the data. For instance, Kröger and colleagues (2016) used several public microarray experiments to identify new altered genetic pathways in blood mononucleolar cells of systematic lupus erythematosus patients [36], while Gonzalez-Porta and colleagues (2012) used public RNA-Seq data from HapMap to evaluate different splicing patterns in Caucasians and Yorubas [37]. Also, Cao and colleagues (2015) [11] used public microarray data to identify more than 300 genes differentially expressed between smokers and non-smokers.

In this article we detected, for the first time, differentially and constitutively expressed piRNAs and snoRNAs in smoker versus non-smoker women with lung adenocarcinoma. Our data were collected from publicly available datasets from samples of patients and may offer new insights in sncRNA biology and on the effects of tobacco use on molecules and cancer biology.

## Material and methods

We obtained two datasets of small RNA-sequencing from lung adenocarcinoma samples belonging to female patients from The Cancer Genome Atlas [38] (TCGA) and from the work published by Kim and collaborators [39]. It is important to stress that throughout this study we use the term 'normal tissue' to describe and discuss data obtained from the normal tissue adjacent to tumors.

Kim and collaborators [39] used the Illumina Genome analyzer Ix to sequence the small RNAs (ranging from 22 to 30 nucleotides long) present in six matched primary lung adenocarcinoma tumors and normal tissues from never-smoker women [NCBI GEO:GSE37764]. Although the authors only assessed the expression of miRNAs, they mention the possibility of the raw data containing sncRNAs other than microRNAs.

The 36-nucleotide length reads from the data published by Kim and collaborators were investigated for quality and adaptor presence. The adaptors were removed with Cutadapt [40] and only the reads longer than 15 nucleotides were kept. Also, any read with more than 10% of its bases with quality lower than 20 was removed using the software FASTQ Quality Filter from FASTX-Toolkit (Gordon and Hannon, unpublished). The remaining reads were aligned to the human genome using Novoalign (version 3.02.13, <http://www.novocraft.com>) returning reads that aligned at a single location in the human reference genome.

The TCGA project is a collaboration between the National Cancer Institute (NCI) and the National Human Genome Research Institute (NHGRI) that generated and made publicly available genomic, transcriptomic and methylomic data from 33 types of cancer. We downloaded (January 2016) paired normal tissue and tumor miRNA-Seq data from samples belonging to the same patient, obtaining a total of 6 paired samples from 6 women (dbgap data access committee Project ID 43224–3).

TCGA data were generated using either the Illumina Genome Analyzer or the Illumina HiSeq. The files downloaded in the BAM file format were converted to FASTQ using BEDtools [41]. Converted files were aligned to the human genome using Novoalign with the same configuration as described above, aiming for the detection of sncRNAs other than miRNA.

The piRNA annotation file was obtained from the piRNAbank database [42], and overlapping annotations that were in the same orientation were grouped into clusters (S1 File). This modified GTF file has 2,049 piRNA clusters and 11,710 piRNAs without any overlap. The annotation file for the snoRNA was obtained from UCSC [43]. The piRNA and snoRNA annotation files were merged into a single file (S2 File). Any piRNA regions that overlapped with snoRNA annotations were considered as snoRNA, and the ones that overlapped with miRNAs annotations were discarded as annotation errors.

The raw counts for each annotated piRNA and snoRNA present in the samples sequenced using the same technology were obtained using BEDtools [41] (S3 File). The identification of the differentially expressed piRNAs and snoRNAs was performed using R version 3.4.0 and following the Bioconductor package EdgeR version 3.18.1 manual [44]. In short, we used the TMM methodology [45] to normalize the count per million (CPM) values obtained from the raw counts, and the negative binomial test to identify the differentially expressed genes. In this step, only small RNAs with at least 50 reads mapped in at least half of the samples, and those with counts greater than 0 in all but two samples were considered. We called differentially expressed all snoRNAs and piRNAs with FDR lower than 0.01 and logarithm of fold change greater than 2 or lower than -2. Aiming to validate the efficacy of our strategy, we first used it to investigate the differentially expressed miRNAs in Kim and collaborators and TCGA data using an annotation file obtained from MiRBase version 20 [46].

According to Eisenberg and Levanon, housekeeping genes are constitutively expressed in all cell types under normal conditions [47]. Therefore, to identify only the snoRNAs and piRNAs altered by tobacco usage and not by tumor status, we performed a data dispersion analysis on the samples from non-smokers and smokers according to the procedures described in [47]. Briefly, we kept only the sncRNAs with more than 1 CPM in all samples, whose expression did not differ more than 2 times from the average log<sub>2</sub> normalized counts, and those with standard deviation lower than 1. These sncRNAs were considered putative housekeeping genes.

## Results

### Methodology validation

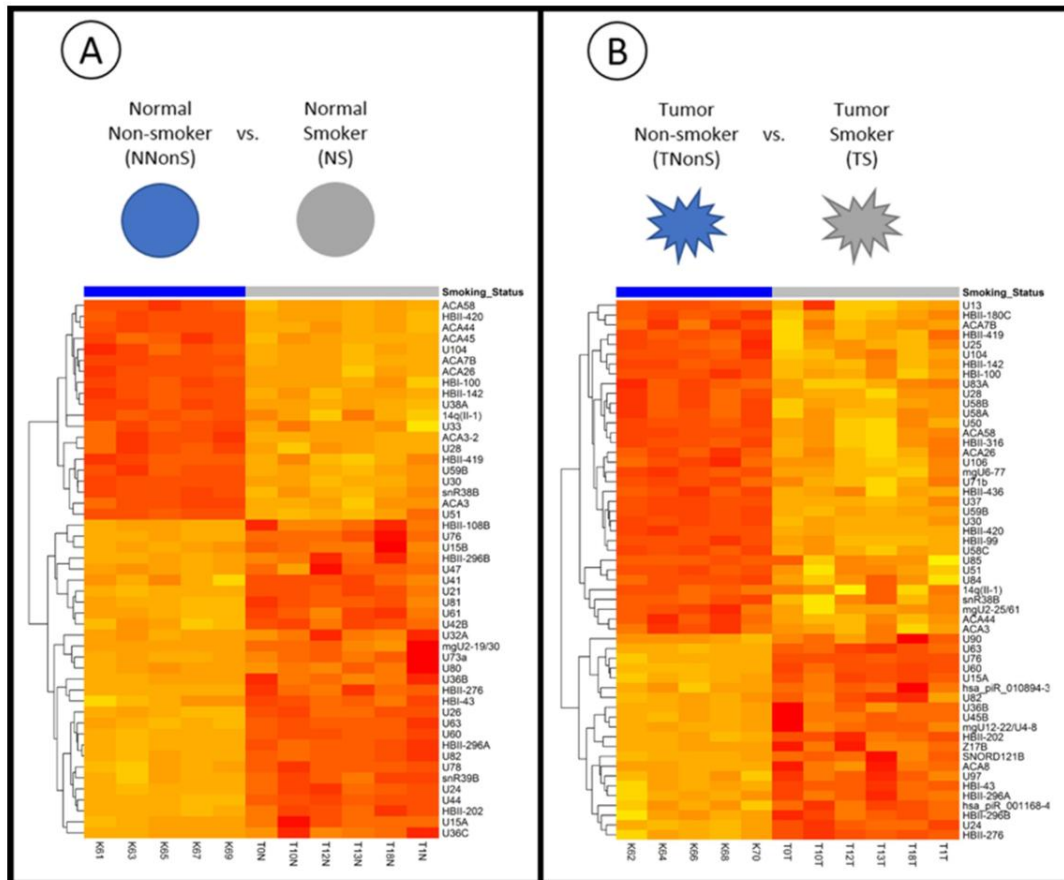
Kim et al. [39] pointed out 40 differentially expressed miRNAs in their work. Using our approach, we obtained 23 miRNAs that were differentially expressed between normal tissue and tumor samples belonging to non-smokers. Twenty of these were similar to those found by the original authors. Four other miRNAs found by Kim et al. [39] were also found by us but removed either by the false discovery rate (FDR) or logarithm of fold change filter (S4 File).

The same approach described above was adopted between normal tissue and tumor samples using smokers' data. We found 23 differentially expressed miRNAs (S5 File). Only 5 miRNAs were found in both analyses. Out of those, 2 were found up-regulated in tumors: hsa-miR-183-5p, hsa-miR-210-3p, while 3 were down-regulated in tumors: hsa-miR-144-5p, hsa-miR-451a, and hsa-miR-30c-2-3p. The comparison between normal tissue samples from non-smokers and smokers revealed very distinct expression profiles with 130 differentially expressed miRNAs (S6 File). Although this distinction was not as clear when tumor samples were compared, we still found 135 miRNAs that were differentially expressed (S7 File).

### Comparison between lung tissues of smokers and non-smokers

Aiming to detect differences in the expression levels of piRNA and snoRNA between lung samples belonging to smoker and non-smoker women, we performed two differential expression analysis involving four groups of samples: normal non-smoker (NNonS) versus normal smoker (NS) samples (Fig 1A), and tumor non-smoker (TNonS) versus tumor smoker (TS) samples (Fig 1B).

First, we analyzed the differentially expressed genes in normal tissue samples from non-smokers as compared to smokers. After applying our FDR and log<sub>2</sub>FC criteria, we detected 49 differentially expressed snoRNA between NNonS and NS samples (S8 File). According to our analysis, 29 snoRNAs are more expressed in NNonS and 20 are down-regulated (Fig 1A). U60 (SNORD60) is the most up-regulated snoRNA (log<sub>2</sub>FC = 6.22) and HBII-420 (SNORD99) is the most down-regulated snoRNA (log<sub>2</sub>FC = -5.25). The magnitudes of fold-change are similar



**Fig 1. Differential expression comparisons performed in this study and heatmaps.** A) Non-smoker normal versus Smoker normal samples. B) Non-smoker tumor versus Smoker tumor samples. Heatmap of Log<sub>2</sub> of Normalized CPM counts for differentially expressed genes. Data below the blue bar are from non-smokers, and those below the gray bar are from smokers. The figure shows four sets of genes whose expressions are very different according to the smoking status.

<https://doi.org/10.1371/journal.pone.0183410.g001>

between the up- and down-regulated genes. Table 1 shows the logFC and logCPM of the 10 greatest changes found in this analysis. No piRNA was found differentially expressed between NNoS and NS.

The differential expression analysis of the Non-smokers tumor (TNonS) and Smokers tumor (TS) samples allowed the identification of 55 piRNA or snoRNA that presented altered levels of expression according to smoking status (S9 File and Fig 1B). In this analysis, 34 genes were found up-regulated in TNonS and 21 down-regulated. Again, U60 is the most up-regulated gene (logFC = 5.63), while U30 is the most down-regulated gene (logFC = -6.72). The magnitudes of fold changes are similar between the two groups as well. Table 2 shows the 10

**Table 1. Top 10 differentially expressed snoRNAs between normal tissue samples from non-smokers and smokers.**

Gene	logFC	logCPM	FDR
U60	6.22	18.78	2.31E-57
U76	5.84	11.88	6.63E-30
U15A	5.50	11.34	5.55E-24
U44	4.23	12.19	6.63E-30
HBII-296A	3.82	11.00	9.94E-21
HBII-142	-3.44	15.29	4.53E-22
snR38B	-3.85	15.37	1.52E-21
ACA7B	-4.21	9.70	1.21E-20
U30	-5.23	16.87	7.16E-36
HBII-420	-5.25	13.92	2.01E-30

logFC: logarithm of Fold Change; logCPM: logarithm of CPM; FDR: false discovery rate; Positive logFC indicates genes more expressed in normal tissue samples from non-smokers.

<https://doi.org/10.1371/journal.pone.0183410.t001>

greatest logFC changes. In this analysis, two piRNAs were found up-regulated in TNonS: has-piR-010894-3 (logFC = 3.43) and has-piR-001168-4 (logFC = 2.95).

### Discriminative expression profile between groups

Next, we sought to evaluate if the differentially expressed sncRNAs found in both comparisons can distinguish non-smokers from smokers by performing principal component analysis (PCA). This analysis showed distinct groups from non-smokers and smokers normal samples (Fig 2A) and tumor samples (Fig 2B). The greatest distinction was found in the comparison between non-smokers and smokers tumor, where the non-smoker samples showed a similar pattern while the smokers samples are dispersed.

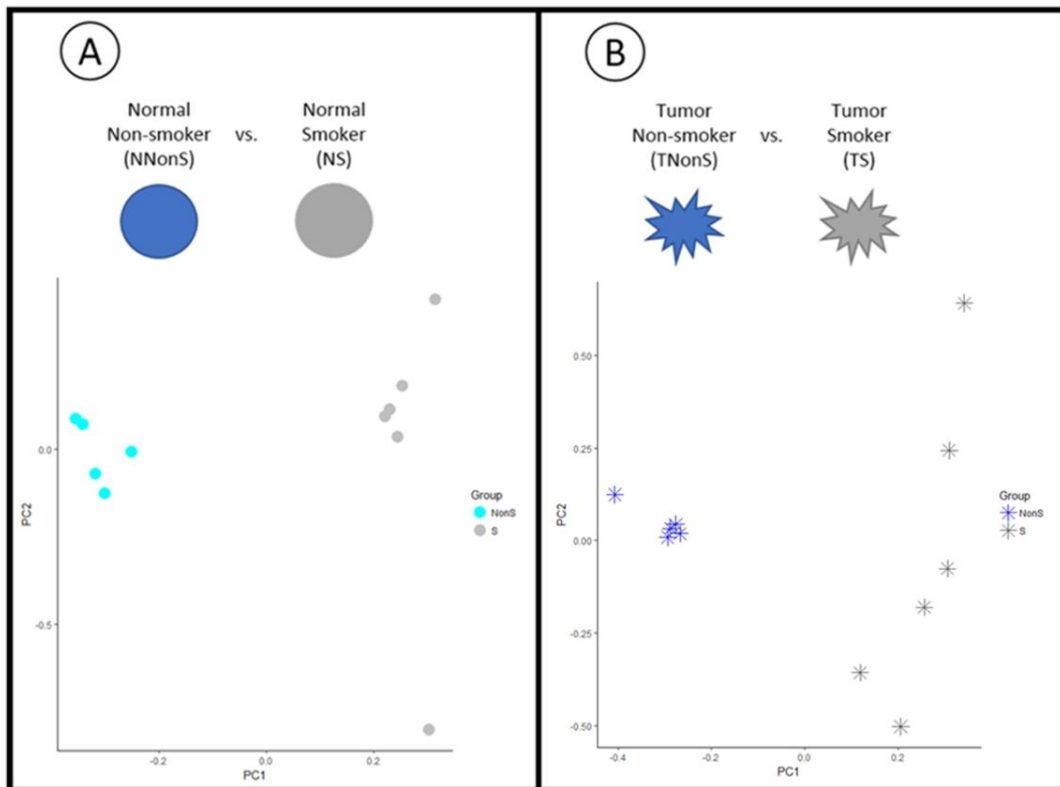
According to our analysis, U60 (SNORD60) is up-regulated in the NNonS when compared to its expression in the NS and the same gene is also up-regulated in TNonS samples when compared to its expression in TS smokers. The same trend is found in U30 (SNORD30) that is down-regulated in NNonS when compared to its expression in NS and in TNonS when

**Table 2. Top 10 differentially expressed snoRNA between tumor samples from smokers and non-smokers.**

Gene	logFC	logCPM	FDR
U60	5.63	19.02	1.69E-28
U76	4.00	11.35	6.78E-22
U15A	3.89	10.75	2.34E-12
U63	3.01	13.33	2.67E-12
HBII-142	-3.00	13.91	2.34E-12
U58C	-3.60	9.94	5.95E-13
U37	-3.69	9.45	5.46E-13
HBII-99	-4.15	8.07	3.21E-15
HBII-420	-4.80	13.20	7.43E-30
U30	-6.72	17.81	5.38E-35

logFC: logarithm of Fold Change; logCPM: logarithm of CPM; FDR: false discovery rate; Positive logFC indicates genes more expressed in non-smoker tumor samples.

<https://doi.org/10.1371/journal.pone.0183410.t002>



**Fig 2. PCA analysis of the differentially expressed piRNAs/snoRNAs.** A) Non-smoker normal (NNonS) versus Smoker normal (NS) samples. B) Non-smoker tumor (TNonS) versus Smoker tumor (TS) samples. The dots represent the normal samples, and stars the tumor samples. Light blue indicates Non-smokers normal samples, dark blue the Non-smokers tumor samples, light gray the Smokers normal samples and dark gray the Smokers tumor samples. Both analysis show distinct expression patterns between non-smokers and smokers.

<https://doi.org/10.1371/journal.pone.0183410.g002>

compared to its expression in TS. In fact, more than half of the differentially expressed snoRNAs were detected in both analyses (28 out of 49 normal samples and 28 out of 55 tumor samples) (Table 3). The complete list of differential expressed sncRNAs is provided as supplemental material (S8 and S9 Files). Additional comparisons between Normal and Tumor samples from Non-smokers and Normal and Tumor samples from Smokers are on S10 and S11 Files, respectively.

Based on this result, we investigated if the expression pattern of these snoRNAs presented the same trend, regardless of the sample pathological status. In this analysis, only snoRNAs and piRNAs that were differentially expressed in at least one set of samples were considered. According to our results, there was not a single sncRNA that showed opposite expression between the Normal and Tumor analysis (Pearson coefficient = 0.80) (Fig 3). The principal component analysis with all differentially expressed piRNAs or snoRNAs also shows clearly

**Table 3. snoRNAs differentially expressed in both normal tissue and tumor comparisons of non-smokers' and smokers' samples.**

Gene	Normal		Tumor	
	logFC	logCPM	logFC	logCPM
<b>U60</b>	<b>6.22</b>	<b>18.78</b>	<b>5.63</b>	<b>19.02</b>
U76	5.84	11.88	4	11.35
U15A	5.5	11.34	3.89	10.75
HBII-296B	4.55	9.25	2.55	8.61
U36B	3.93	8.9	2.65	7.92
HBII-296A	3.82	11	2.68	10.41
HBII-202	3.72	10.82	2.22	9.74
U24	3.49	12.99	2.76	12.15
HBII-276	3.47	10.62	2.13	10.11
U82	3.1	12.44	2.05	11.71
<b>U63</b>	<b>2.63</b>	<b>13.61</b>	<b>3.01</b>	<b>13.33</b>
HBI-43	2.24	10.87	3.38	10.32
<b>U28</b>	<b>-2.06</b>	<b>10.96</b>	<b>-2.33</b>	<b>10.89</b>
<b>U51</b>	<b>-2.17</b>	<b>10.76</b>	<b>-3.03</b>	<b>10.96</b>
<b>U104</b>	<b>-2.21</b>	<b>16.35</b>	<b>-2.27</b>	<b>15.16</b>
<b>HBII-419</b>	<b>-2.27</b>	<b>13.59</b>	<b>-2.34</b>	<b>13.21</b>
<b>U59B</b>	<b>-2.66</b>	<b>10.53</b>	<b>-2.78</b>	<b>9.88</b>
ACA26	-3.04	7.84	-4.23	9.84
ACA3	-3.33	7.33	-2.2	7.47
<b>HBII-142</b>	<b>-3.44</b>	<b>15.29</b>	<b>-3</b>	<b>13.91</b>
ACA58	-3.75	9.25	-4.66	9.44
snR38B	-3.85	15.37	-2.05	13.25
<b>HBI-100</b>	<b>-3.89</b>	<b>10.9</b>	<b>-3.18</b>	<b>9.98</b>
ACA44	-3.96	8.77	-2.81	8.15
ACA7B	-4.21	9.7	-2.17	8.92
14q(l1-1)	-5.01	11.04	-3.57	9.52
<b>U30</b>	<b>-5.23</b>	<b>16.87</b>	<b>-6.72</b>	<b>17.81</b>
<b>HBII-420</b>	<b>-5.25</b>	<b>13.92</b>	<b>-4.8</b>	<b>13.2</b>

logFC: logarithm of Fold Change; logCPM: logarithm of CPM; Positive logFC indicates genes up-regulated in normal tissue samples if compared to tumor samples. A total of 28 genes were found to be differentially expressed in both analysis. All genes found presented the same expression pattern (up- or down-regulated) regardless of the pathological status. The snoRNAs in bold face type showed low dispersion in samples from non-smokers and smokers.

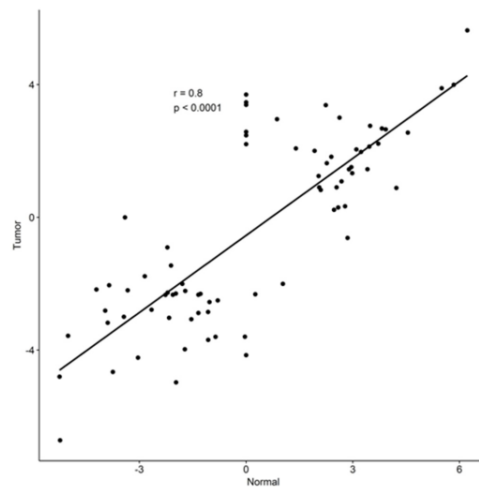
<https://doi.org/10.1371/journal.pone.0183410.t003>

distinct expression differences between non-smokers and smokers (Fig 4). The normal and tumor samples from smokers also show distinct patterns, while the non-smokers samples present the same trend.

### sncRNA expression dispersion analysis between non-smokers and smokers

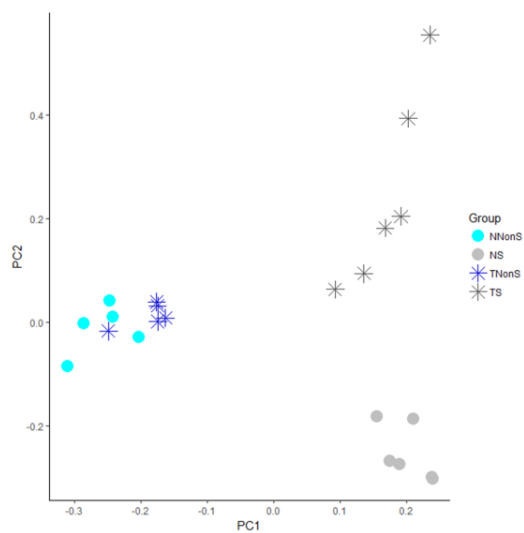
We investigated patterns of data dispersion for samples belonging to non-smokers and smokers, regardless of their pathological status. We performed two data dispersion analysis: NNonS and TNonS and NS and TS.

After applying our constitutive expression criteria, we found 179 snoRNA or piRNAs to be constitutively expressed in samples from non-smokers (S12 File) and 33 genes identified as constitutive in samples from smokers (S13 File). The log2 of the normalized CPM of each sample, and the variance and the standard deviation of the 10 snoRNAs that presented the lowest



**Fig 3. Scatterplot of the log<sub>2</sub> fold change for the 28 snoRNAs shared between analysis.**

<https://doi.org/10.1371/journal.pone.0183410.g003>



**Fig 4. PCA analysis of the differentially expressed piRNAs/snoRNAs.** The dots represent the normal samples, and the stars the tumor samples. Light blue indicates Non-smokers normal samples, dark blue the Non-smokers tumor samples, light gray the Smokers normal samples and dark gray the Smokers tumor samples. This PCA shows 3 distinct groups that correspond to Smokers normal samples, Smokers tumor samples and Non-smokers samples.

<https://doi.org/10.1371/journal.pone.0183410.g004>

**Table 4. Top 10 constitutively expressed snoRNA in samples from non-smokers.**

Gene	K61	K62	K63	K64	K65	K66	K67	K68	K69	K70	Var	SD
U51	8.45	8.48	8.54	8.58	8.79	8.37	8.57	8.76	8.53	8.35	0.02	0.15
U57	12.20	12.38	12.34	12.30	12.07	11.92	12.37	12.36	12.25	11.67	0.05	0.23
HBII-429	9.42	9.37	9.57	9.40	9.81	9.60	9.34	8.92	9.72	9.42	0.06	0.25
HBII-95	5.89	5.62	5.37	5.94	5.87	5.94	5.77	6.12	5.38	5.63	0.06	0.25
U21	13.34	13.27	14.15	13.56	13.60	13.52	13.54	13.68	13.77	13.30	0.07	0.26
U35B	4.98	4.85	4.84	4.14	4.83	4.45	4.93	4.82	4.96	4.74	0.07	0.26
U37	6.65	6.37	5.98	5.95	6.25	6.13	6.21	6.73	6.32	6.67	0.08	0.28
U75	9.08	9.38	8.77	9.23	8.66	8.88	9.01	8.94	8.90	8.42	0.08	0.28
U95	12.08	12.11	12.30	11.66	12.48	11.85	12.26	11.87	12.36	12.45	0.08	0.28
ACA8	5.96	5.27	4.95	5.11	5.50	5.64	5.45	5.52	5.36	5.68	0.08	0.29

SD: standard deviation; Var: variance. Columns 2 to 11 represent each of the non-smokers samples analyzed in this study. Samples ending in odd numbers are from the normal tissue adjacent to tumor and those ending in even numbers represent tumor samples. The numbers below each sample corresponds to log2 of the normalized CPM counts. Overall, the non-smokers samples presented the lowest expression variation.

<https://doi.org/10.1371/journal.pone.0183410.t004>

standard deviation in non-smokers and smokers are shown in the Tables 4 and 5, respectively. All constitutive snoRNAs found in the smokers were also found in the non-smokers samples. However, the lowest variations and standard deviations were found in the non-smokers samples. The only exception is U25, whose expression pattern is more uniform for smokers (Table 6). A total of 11 snoRNAs found to be constitutively expressed in both analyses had expression levels changed in the non-smokers versus smokers comparisons: 2 are up-regulated in non-smokers and 9 are up-regulated in smokers (Boldface type in Table 3).

## Discussion

Small noncoding RNAs, such as snoRNAs and piRNAs, are involved in fundamental biological pathways, and have been considered as potential lung cancer biomarkers [25,30,48]. In this study, we compared the expression pattern of piRNAs, one of the least studied sncRNA classes, and snoRNAs, one of the most studied and well-known sncRNA classes, obtained from

**Table 5. Top 10 constitutively expressed snoRNA in samples from smokers.**

Gene	T0N	T0T	T10N	T10T	T12N	T12T	T13N	T13T	T18N	T18T	T1N	T1T	Var	SD
U59B	11.50	11.05	11.56	11.58	11.53	11.03	11.90	11.49	11.74	10.87	11.08	11.23	0.10	0.32
U25	13.24	13.95	12.68	12.93	12.97	12.98	13.14	13.57	13.13	13.52	12.62	12.57	0.17	0.41
U57	12.71	12.29	12.37	12.40	12.28	11.78	12.25	12.91	12.67	11.72	11.69	11.88	0.17	0.41
U48	12.81	13.26	13.23	13.56	13.49	13.46	13.34	12.75	13.38	14.38	13.39	12.89	0.19	0.43
U27	14.04	14.30	13.27	14.65	14.07	13.93	14.05	14.04	13.93	13.40	12.96	13.92	0.21	0.46
U21	11.90	11.95	12.52	11.94	12.28	12.87	11.28	12.62	12.36	12.85	13.21	12.03	0.29	0.53
HBII-420	14.95	14.53	15.03	14.01	14.24	14.78	14.67	14.85	14.49	14.90	16.25	14.67	0.30	0.55
U20	13.15	13.07	13.00	11.70	12.94	12.96	12.70	13.61	12.90	11.75	12.98	12.22	0.33	0.57
U43	13.73	12.91	12.84	12.45	13.74	13.25	13.43	12.72	14.26	13.67	14.02	12.59	0.36	0.60
snR39B	12.67	12.55	12.00	11.33	12.42	13.40	12.90	12.97	12.19	12.85	11.79	13.31	0.38	0.62

SD: standard deviation; Var: variance. Columns 2 to 13 represent each of the smokers samples analyzed in this study. Samples ending with "N" are from the normal tissue adjacent to tumor and those ending with "T" are from tumor samples. The variances and standard deviation for smokers were greater than the non-smokers, with exception of U25.

<https://doi.org/10.1371/journal.pone.0183410.t005>



**Table 6. snoRNAs found constitutively expressed in samples from both smoker and non-smoker patients.**

Gene	Non-smoker			Smoker		
	Mean	Var	SD	Mean	Var	SD
U59B	7.78	0.09	0.30	11.38	0.10	0.32
U25	10.34	0.27	0.52	13.11	0.17	0.41
U57	12.19	0.05	0.23	12.25	0.17	0.41
U48	11.47	0.18	0.43	13.33	0.19	0.43
U27	14.34	0.14	0.37	13.88	0.21	0.46
U21	13.57	0.07	0.26	12.32	0.29	0.53
HBII-420	8.87	0.20	0.45	14.78	0.30	0.55
U20	10.69	0.11	0.33	12.75	0.33	0.57
U43	12.68	0.29	0.53	13.30	0.36	0.60
snR39B	13.19	0.23	0.48	12.53	0.38	0.62
U63	13.91	0.14	0.38	11.90	0.38	0.62
U104	13.66	0.37	0.61	16.78	0.42	0.65
HBII-336	11.03	0.44	0.66	11.49	0.43	0.66
HBII-419	11.27	0.23	0.48	14.34	0.44	0.66
SNORD119	12.97	0.29	0.53	12.55	0.45	0.67
U28	8.88	0.24	0.49	11.96	0.46	0.68
U42A	9.70	0.41	0.64	12.09	0.46	0.68
U38A	8.92	0.40	0.64	12.10	0.48	0.69
U59A	9.03	0.15	0.38	10.89	0.49	0.70
HBII-55	9.94	0.21	0.46	11.12	0.52	0.72
U31	13.95	0.17	0.42	15.06	0.53	0.73
HBII-142	11.65	0.25	0.50	15.74	0.55	0.74
ACA45	10.14	0.35	0.59	12.69	0.57	0.75
HBII-210	11.67	0.16	0.40	11.67	0.58	0.76
U52	11.03	0.34	0.58	11.64	0.60	0.78
U95	12.14	0.08	0.28	13.54	0.64	0.80
HBII-251	11.78	0.22	0.47	12.24	0.65	0.81
HBI-100	7.22	0.32	0.57	11.54	0.67	0.82
U60	19.47	0.53	0.73	14.43	0.71	0.84
U51	8.54	0.02	0.15	11.78	0.82	0.91
U50	8.63	0.17	0.41	11.53	0.88	0.94
HBII-295	11.43	0.29	0.54	13.22	0.90	0.95
U30	11.75	0.14	0.37	18.49	0.90	0.95

SD: standard deviation; Var: variance. The mean corresponds to the average of the log<sub>2</sub> normalized CPM expression values of each smoking status group. A total of 33 genes were found to be constitutively expressed in both non-smoker and smoker groups.

<https://doi.org/10.1371/journal.pone.0183410.t006>

publicly available datasets from lung adenocarcinoma samples belonging to matched smoker and non-smoker women.

We obtained 5 matched samples from normal and tumor tissues belonging to non-smoker women from the work of Kim and colleagues [39]. Additionally, we also used TCGA data from 6 matched samples from normal tissue and lung adenocarcinoma belonging to smoker women. To validate our methodology, we first evaluated the profile of differentially expressed miRNAs and compared our results with the original publication. We found 20 miRNAs shared with the work of Kim and colleagues [39] while 4 other miRNAs found by the authors were removed by our FDR and logFC filters. For smokers from TCGA, 23 miRNAs were found

differentially expressed. Several papers confirm our findings [49–53]. Principal component analysis and multidimensional analysis showed clear differences between the normal tissue and tumor samples from smokers and non-smokers (SFile 6 and 7).

In total, we identified 49 differentially expressed snoRNAs or piRNAs between NNonS and NS samples, 20 down-regulated and 29 up-regulated (Fig 1A). A total of 55 snoRNAs or piRNAs with altered expression were also identified between TNonS and TS samples, 34 down-regulated and 21 up-regulated (Fig 1B). The changes in expression profile are also confirmed by principal component analysis and multidimensional analysis (SFile 8 and 9). Twenty-eight snoRNAs were found in both analysis and showed the same expression pattern in both normal tissue and tumor samples (Fig 3), thus further indicating that this alteration in expression profile is related to smoking. One of these is the snoRNA U15A (SNORD15A), which is up-regulated in samples from non-smokers. Curiously, this snoRNA was reported as having an 'miRNA-like' function and being capable of silencing the reporter gene [54] as well as being related to the regulation of chromatin structure in fibroblasts [55]. SNORD15A may have other putative roles beyond the modification of ribosomal RNAs and our results suggest that it may be involved in a mechanism in which smoking could alter its cell expression profile.

Out of the 28 sncRNAs altered between non-smokers and smokers, the data dispersion analysis showed that 11 snoRNAs presented a constant expression across normal tissue and tumor samples that belong to individuals in the same smoking status group. Only two of them were up-regulated in the non-smoker samples: U60 (SNORD60) and U63 (SNORD63). U60 has been reported as an attenuator of pulmonary vasoconstriction in rats [56] and as a key factor in the regulation of plasma membrane cholesterol [57], while U63 (SNORD63) is located in a chromosomal region frequently deleted in myelodysplastic syndromes [58]. Thus, we hypothesize that smoking may inhibit the expression of snoRNAs that are important for cell maintenance.

Interestingly, many of the remaining 9 snoRNAs that were found up-regulated in smokers were reported as altered in different cancer types. One example is the HBII-142 (SNORD66), located on a chromosome region frequently amplified in tumors [25]. SNORD66 was already suggested as a biomarker candidate for lung cancer due to its being up-regulated in the plasma of lung cancer patients [59]. For instance, U30 (SNORD30) was reported as up-regulated in pediatric gliomas [60] and seems to correlate with the shorter time to progression in multiple myeloma [61]. Another differentially expressed snoRNA was HBI-100 (SCARNA3), which showed to be up-regulated in breast cancer samples [62]. In peripheral T-cell lymphoma, U59B (SNORD59B) correlated with long-term survival [63]. Taken together, our findings suggest that tobacco use modifies the expression profile of several snoRNAs towards a pattern like that observed in the malignant phenotype.

Many snoRNAs still do not have an identified target and their expression patterns are still unknown. In the work of Lan and collaborators [64], authors assess the expression of HBII-420 (SNORD99) and ACA44 (SNORA44). Both genes are located in introns of the SNHG12 long non-coding RNA, and knocking down the host expression does not alter the snoRNAs expression [64]. Further studies on snoRNAs targets may reveal new biological pathways affected by smoking. These findings can be used to better understand different tobacco-related pathologies and improve their treatment.

We found that U44 was up-regulated in samples from non-smokers and constitutively expressed in samples from non-smokers. Interestingly, this is an snoRNA frequently used for normalization in qRT-PCR experiments [65]. According to our findings, the use of this snoRNA may create bias between non-smokers and smokers. Similar alterations in the expression of this snoRNA were already reported in colorectal and breast cancer [65,66].

Of the 33 snoRNAs found to be constitutively expressed, 16 were not significantly different between smokers and non-smokers, suggesting that they are not modified by tobacco use. Among those, U43 (SNORD43) is frequently used as normalization parameter for miRNA expression profile experiments due to its stable levels, thus reinforcing our findings [67]. More studies on the expression levels of these snoRNAs may reveal other candidates for normalization parameters.

As for piRNAs, it is known that PIWI proteins and piRNAs protect integrity and stability of the genome by regulating transposable elements. However, PIWI proteins have been described to be differently expressed in NSCLC and to be associated with patient survival [68–70]. It is speculated that aberrant transposable elements could increase the number of deletions, rearrangements, and duplications that are frequently observed in the genome of cancer cells [71].

Although snoRNAs and piRNAs are receiving more attention from research groups, there are still few reports assessing changes in their expression under different biological conditions and, until the submission of this paper, we could not find any report that has evaluated the expression patterns of this many snoRNAs and piRNAs in smokers' and non-smokers' samples.

To our knowledge, in many lung studies searching for genes playing a role in the disease, data between non-smoker and smoker samples were not separated [30,48–50,53]. However, here we show that these two groups have distinct gene expression profiles. This is line with data generated elsewhere [12,13]. Therefore, data from these two groups of individuals should be considered separately in future studies to avoid the introduction or errors and confusing factors that may lead to misleading results.

Our results show that smoking modifies the expression of many sncRNAs, thus changing their expression profile towards one that is more like to that reported in different types of cancer. Here we report distinct sets of sncRNAs that can be used to distinguish smokers and non-smokers and should be considered when analyzing data from these two groups. Further studies about sncRNA targets should reveal new affected biological pathways.

## Conclusions

The identification of a molecular signature in lung cancer that permit the discrimination of tumors from smokers and non-smokers is still a challenge. In this article, we report several snoRNAs and piRNAs that are differentially expressed in lung adenocarcinoma samples from smoker and from non-smoker women. We believe that these sets of constitutively and differentially expressed snoRNAs and piRNAs can be used in the future to improve the molecular diagnosis and treatment of lung cancer patients. Our findings highlight the importance of studying sncRNAs in cancer biology and their application as potential biomarkers in the era of precision medicine.

## Supporting information

### S1 File. Overlapping piRNA clusters.

(TXT)

### S2 File. snoRNA piRNA annotation file in BED format.

(BED)

### S3 File. Raw snoRNA and piRNA counts.

(TSV)

### S4 File. Non-smoker Normal vs Tumor miRNA analysis.

(PDF)

**S5 File. Smoker Normal vs Tumor miRNA analysis.**  
(PDF)

**S6 File. Normal Non-smoker vs Smoker miRNA analysis.**  
(PDF)

**S7 File. Tumor Non-smoker vs Smoker miRNA analysis.**  
(PDF)

**S8 File. Normal Non-smoker vs Smoker snoRNA piRNA analysis.**  
(PDF)

**S9 File. Tumor Non-smoker vs Smoker snoRNA piRNA analysis.**  
(PDF)

**S10 File. Non-smoker Normal vs Tumor snoRNA piRNA analysis.**  
(PDF)

**S11 File. Smoker Normal vs Tumor snoRNA piRNA analysis**  
(PDF)

**S12 File. Non-smoker variance analysis.**  
(PDF)

**S13 File. Smoker variance analysis.**  
(PDF)

### Acknowledgments

The authors acknowledge Plataforma de Bioinformática da Fiocruz RPT04A/RJ and Novocraft support, the University of Pittsburgh, the Christiana Healthcare, the Johns Hopkins, and the ABS–Indiana University–Purdue University Indiana tissue source sites and Canada’s Michael Smith Genome Sciences Centre sequencing. All authors acknowledge Otacilio da Cruz Moreira, Mariana Caldas Waghbi, Teca Calcagno Galvão, Rui Manuel Reis, Renato José da Silva-Oliveira, and Adriana Cruvinel-Carlioni for comments and assistance.

### Author Contributions

**Conceptualization:** Natasha Andressa Nogueira Jorge, Fabio Passetti.

**Data curation:** Natasha Andressa Nogueira Jorge, Fabio Passetti.

**Formal analysis:** Natasha Andressa Nogueira Jorge, Gabriel Wajnberg, Benilton de Sa Carvalho.

**Funding acquisition:** Benilton de Sa Carvalho, Fabio Passetti.

**Investigation:** Natasha Andressa Nogueira Jorge, Carlos Gil Ferreira, Fabio Passetti.

**Methodology:** Natasha Andressa Nogueira Jorge, Gabriel Wajnberg, Benilton de Sa Carvalho, Fabio Passetti.

**Resources:** Natasha Andressa Nogueira Jorge.

**Software:** Natasha Andressa Nogueira Jorge, Gabriel Wajnberg, Benilton de Sa Carvalho.

**Supervision:** Carlos Gil Ferreira, Benilton de Sa Carvalho, Fabio Passetti.

**Writing – original draft:** Natasha Andressa Nogueira Jorge, Gabriel Wajnberg, Carlos Gil Ferreira, Benilton de Sa Carvalho, Fabio Passetti.

**Writing – review & editing:** Natasha Andressa Nogueira Jorge, Gabriel Wajnberg, Carlos Gil Ferreira, Benilton de Sa Carvalho, Fabio Passetti.

## References

1. Saalberg Y, Wolff M. VOC breath biomarkers in lung cancer. *Clin Chim Acta*. 2016; 459: 5–9. <https://doi.org/10.1016/j.cca.2016.05.013> PMID: 27221203
2. Mendes R, Carreira B, Baptista P V., Fernandes AR. Non-small cell lung cancer biomarkers and targeted therapy—two faces of the same coin fostered by nanotechnology. *Expert Rev Precis Med Drug Dev*. 2016; 1: 155–168. <https://doi.org/10.1080/23808993.2016.1159914>
3. Herbst RS, Heymach J V, Lippman SM. Lung cancer. *N Engl J Med*. Massachusetts Medical Society; 2008; 359: 1367–80. <https://doi.org/10.1056/NEJMra0802714> PMID: 18815398
4. Li L, Zhu T, Gao Y-F, Zheng W, Wang C-J, Xiao L, et al. Targeting DNA Damage Response in the Radio(Chemo)therapy of Non-Small Cell Lung Cancer. *Int J Mol Sci. Multidisciplinary Digital Publishing Institute*; 2016; 17: 839. <https://doi.org/10.3390/ijms17060839> PMID: 27258253
5. Gyoba J, Shan S, Roa W, Bédard ELR. Diagnosing Lung Cancers through Examination of Micro-RNA Biomarkers in Blood, Plasma, Serum and Sputum: A Review and Summary of Current Literature. *Int J Mol Sci*. 2016; 17: 10.3390/ijms17040494
6. Gumus ZH, Du B, Kacker A, Boyle JO, Bocker JM, Mukherjee P, et al. Effects of Tobacco Smoke on Gene Expression and Cellular Pathways in a Cellular Model of Oral Leukoplakia. *Cancer Prev Res*. 2008; 1: 100–111. <https://doi.org/10.1158/1940-6207.CAPR-08-0007> PMID: 19138943
7. Bialous SA, Sama L. Lung Cancer and Tobacco. *Nurs Clin North Am*. 2017; 52: 53–63. <https://doi.org/10.1016/j.cnur.2016.10.003> PMID: 28189166
8. Vavalà T, Mariniello A, Reale ML, Novello S. Gender differences in lung cancer. *Ital J Gender-Specific Med*. 2016; 2: 99–109. <https://doi.org/10.1723/2625.26992>
9. Huang J-Y, Jian Z-H, Nfor ON, Ku W-Y, Ko P-C, Lung C-C, et al. The effects of pulmonary diseases on histologic types of lung cancer in both sexes: a population-based study in Taiwan. *BMC Cancer*. 2015; 15: 834. <https://doi.org/10.1186/s12885-015-1847-z> PMID: 26526071
10. Spira A, Beane J, Shah V, Liu G, Schembri F, Yang X, et al. Effects of cigarette smoke on the human airway epithelial cell transcriptome. *Proc Natl Acad Sci*. 2004; 101: 10143–10148. <https://doi.org/10.1073/pnas.0401422101> PMID: 15210990
11. Cao C, Chen J, Lyu C, Yu J, Zhao W, Wang Y, et al. Bioinformatics Analysis of the Effects of Tobacco Smoke on Gene Expression. *PLoS One*. 2015; 10: e0143377. <https://doi.org/10.1371/journal.pone.0143377> PMID: 26629988
12. Wang J, Li MD. Common and Unique Biological Pathways Associated with Smoking Initiation/Progression, Nicotine Dependence, and Smoking Cessation. *Neuropsychopharmacology*. 2010; 35: 702–719. <https://doi.org/10.1038/npp.2009.178> PMID: 19890259
13. Minică CC, Mbarek H, Pool R, Dolan C V, Boomsma DI, Vink JM. Pathways to smoking behaviours: biological insights from the Tobacco and Genetics Consortium meta-analysis. *Mol Psychiatry*. 2017; 22: 82–88. <https://doi.org/10.1038/mp.2016.20> PMID: 27021816
14. Le Thomas A, Tóth K, Aravin AA, Ghildiyal M, Zamore P, Malone C, et al. To be or not to be a piRNA: genomic origin and processing of piRNAs. *Genome Biol. BioMed Central*; 2014; 15: 204. <https://doi.org/10.1186/gb4154> PMID: 24467990
15. Siomi MC, Sato K, Pezic D, Aravin AA. PIWI-interacting small RNAs: the vanguard of genome defence. *Nat Rev Mol Cell Biol*. Nature Publishing Group, a division of Macmillan Publishers Limited. All Rights Reserved.; 2011; 12: 246–58. <https://doi.org/10.1038/nrm3089> PMID: 21427766
16. Klattenhoff C, Theurkauf W. Biogenesis and germline functions of piRNAs. *Development*. 2008; 135: 3–9. <https://doi.org/10.1242/dev.006486> PMID: 18032451
17. Yan Z, Hu HYH, Jiang X, Maierhofer V, Neb E, He L, et al. Widespread expression of piRNA-like molecules in somatic tissues. *Nucleic Acids Res*. 2011; 39: 6596–607. <https://doi.org/10.1093/nar/gkr298> PMID: 21546553
18. Yang Q, Hua J, Wang L, Xu B, Zhang H, Ye N, et al. MicroRNA and piRNA profiles in normal human testis detected by next generation sequencing. *PLoS One*. 2013; 8: e66809. <https://doi.org/10.1371/journal.pone.0066809> PMID: 23826142

19. Fu A, Jacobs DI, Hoffman AE, Zheng T, Zhu Y. PIWI-interacting RNA 021285 is involved in breast tumorigenesis possibly by remodeling the cancer epigenome. *Carcinogenesis*. 2015; 36: 1094–102. <https://doi.org/10.1093/carcin/bgv105> PMID: 26210741
20. Rajasethupathy P, Antonov I, Sheridan R, Frey S, Sander C, Tuschl T, et al. A role for neuronal piRNAs in the epigenetic control of memory-related synaptic plasticity. *Cell*. 2012; 149: 693–707. <https://doi.org/10.1016/j.cell.2012.02.057> PMID: 22541438
21. Ross RJ, Weiner MM, Lin H. PIWI proteins and PIWI-interacting RNAs in the soma. *Nature*. 2014; 505: 353–9. <https://doi.org/10.1038/nature12987> PMID: 24429634
22. Busch J, Ralla B, Jung M, Wotschovsky Z, Trujillo-Arribas E, Schwabe P, et al. Piwi-interacting RNAs as novel prognostic markers in clear cell renal cell carcinomas. *J Exp Clin Cancer Res. BioMed Central*; 2015; 34: 61. <https://doi.org/10.1186/s13046-015-0180-3> PMID: 26071182
23. Cheng J, Guo J-M, Xiao B-X, Miao Y, Jiang Z, Zhou H, et al. piRNA, the new non-coding RNA, is aberrantly expressed in human cancer cells. *Clin Chim Acta. Elsevier B.V.*; 2011; 412: 1621–5. <https://doi.org/10.1016/j.cca.2011.05.015> PMID: 21616063
24. Müller S, Raulefs S, Bruns P, Afonso-Grunz F, Plötner A, Thermann R, et al. Next-generation sequencing reveals novel differentially regulated mRNAs, lncRNAs, miRNAs, sdRNAs and a piRNA in pancreatic cancer. *Mol Cancer. BioMed Central*; 2015; 14: 94. <https://doi.org/10.1186/s12943-015-0358-5> PMID: 25910082
25. Thorenor N, Slaby O. Small nucleolar RNAs functioning and potential roles in cancer. *Tumor Biol*. 2015; 36: 41–53. <https://doi.org/10.1007/s13277-014-2818-8> PMID: 25420907
26. Taft RJ, Glazov EA, Lassmann T, Hayashizaki Y, Carninci P, Mattick JS. Small RNAs derived from snoRNAs. *RNA*. 2009; 15: 1233–1240. <https://doi.org/10.1261/ma.1528909> PMID: 19474147
27. Stepanov GA, Filippova JA, Komissarov AB, Kuligina E V, Richter VA, Semenov D V. Regulatory role of small nucleolar RNAs in human diseases. *Biomed Res Int*. 2015; 2015: 206849. <https://doi.org/10.1155/2015/206849> PMID: 26060813
28. Liu Z, Yang G, Zhao T, Cao G, Xiong L, Xia W, et al. Small ncRNA expression and regulation under hypoxia in neural progenitor cells. *Cell Mol Neurobiol*. 2011; 31: 1–5. <https://doi.org/10.1007/s10571-010-9556-2> PMID: 20886369
29. Michel CI, Holley CL, Scruggs BS, Sidhu R, Brookheart RT, Listenberger LL, et al. Small nucleolar RNAs U32a, U33, and U35a are critical mediators of metabolic stress. *Cell Metab*. 2011; 14: 33–44. <https://doi.org/10.1016/j.cmet.2011.04.009> PMID: 21723502
30. Mei Y-P, Liao J-P, Shen J, Yu L, Liu B-L, Liu L, et al. Small nucleolar RNA 42 acts as an oncogene in lung tumorigenesis. *Oncogene*. 2012; 31: 2794–804. <https://doi.org/10.1038/onc.2011.449> PMID: 21986946
31. Tainsky M. Genomic and proteomic biomarkers for cancer: a multitude of opportunities. . . . *Biophys Acta (BBA)-Reviews Cancer*. 2009; 1796: 176–193. <https://doi.org/10.1016/j.bbcan.2009.04.004>. *Genomic*
32. Zhou L, Li X, Liu Q, Zhao F, Wu J. Small RNA transcriptome investigation based on next-generation sequencing technology. *J Genet Genomics*. 2011; 38: 505–513. <https://doi.org/10.1016/j.jgg.2011.08.006> PMID: 22133681
33. Meyerson M, Gabriel S, Getz G. Advances in understanding cancer genomes through second-generation sequencing. *Nat Rev Genet. Nature Publishing Group*; 2010; 11: 685–696. <https://doi.org/10.1038/nrg2841> PMID: 20847746
34. Jorge NAN, Ferreira CG, Passetti F. Bioinformatics of Cancer ncRNA in High Throughput Sequencing: Present State and Challenges. *Front Genet*. 2012; 3: 287. <https://doi.org/10.3389/fgene.2012.00287> PMID: 23251139
35. Rung J, Brazma A. Reuse of public genome-wide gene expression data. *Nat Rev Genet*. 2012; 14: 89–99. <https://doi.org/10.1038/nrg3394> PMID: 23269463
36. Kröger W, Mapiye D, Entfellner J-BD, Tiffin N. A meta-analysis of public microarray data identifies gene regulatory pathways deregulated in peripheral blood mononuclear cells from individuals with Systemic Lupus Erythematosus compared to those without. *BMC Med Genomics*. 2016; 9: 66. <https://doi.org/10.1186/s12920-016-0227-0> PMID: 27846842
37. Gonzalez-Porta M, Calvo M, Sammeth M, Guigo R. Estimation of alternative splicing variability in human populations. *Genome Res*. 2012; 22: 528–538. <https://doi.org/10.1101/gr.121947.111> PMID: 22113879
38. The Cancer Genome Atlas [Internet]. Recuperado: <http://cancergenome.nih.gov/>
39. Kim SCS, Jung YY, Park JJ, Cho S, Seo C, Kim JJ, et al. A high-dimensional, deep-sequencing study of lung adenocarcinoma in female never-smokers. *PLoS One*. 2013; 8: e55596. <https://doi.org/10.1371/journal.pone.0055596> PMID: 23405175

40. Martin M. Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet Journal*. 2011 17: 10–12.
41. Quinlan AR, Hall IM. BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics*. 2010; 26: 841–2. <https://doi.org/10.1093/bioinformatics/btq033> PMID: 20110278
42. Sai Lakshmi S, Agrawal S. piRNABank: a web resource on classified and clustered Piwi-interacting RNAs. *Nucleic Acids Res*. 2008; 36: D173–7. <https://doi.org/10.1093/nar/gkm696> PMID: 17881367
43. Rosenbloom KR, Armstrong J, Barber GP, Casper J, Clawson H, Diekhans M, et al. The UCSC Genome Browser database: 2015 update. *Nucleic Acids Res*. 2015; 43: D670–81. <https://doi.org/10.1093/nar/gku1177> PMID: 25428374
44. Robinson MD, McCarthy DJ, Smyth GK. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics*. 2010; 26: 139–40. <https://doi.org/10.1093/bioinformatics/btp616> PMID: 19910308
45. Robinson MD, Oshlack A. A scaling normalization method for differential expression analysis of RNA-seq data. *Genome Biol*. 2010; 11: R25. <https://doi.org/10.1186/gb-2010-11-3-r25> PMID: 20196867
46. Kozomara A, Griffiths-Jones S. miRBase: integrating microRNA annotation and deep-sequencing data. *Nucleic Acids Res*. 2011; 39: D152–7. <https://doi.org/10.1093/nar/gkq1027> PMID: 21037258
47. Eisenberg E, Levanon EY. Human housekeeping genes, revisited. *Trends Genet*. 2013; 29: 569–574. <https://doi.org/10.1016/j.tig.2013.05.010> PMID: 23810203
48. Zheng D, Zhang J, Ni J, Luo J, Wang J, Tang L, et al. Small nucleolar RNA 78 promotes the tumorigenesis in non-small cell lung cancer. *J Exp Clin Cancer Res*. 2015; 34: 49. <https://doi.org/10.1186/s13046-015-0170-5> PMID: 25975345
49. Zhu W, Zhou K, Zha Y, Chen D, He J, Ma H, et al. Diagnostic Value of Serum miR-182, miR-183, miR-210, and miR-126 Levels in Patients with Early-Stage Non-Small Cell Lung Cancer. Zheng SG, organizer. *PLoS One*. 2016; 11: e0153046. <https://doi.org/10.1371/journal.pone.0153046> PMID: 27093275
50. Chen S, Li P, Li J, Wang Y, Du Y, Chen X, et al. MiR-144 Inhibits Proliferation and Induces Apoptosis and Autophagy in Lung Cancer Cells by Targeting TIGAR. *Cell Physiol Biochem*. 2015; 35: 997–1007. <https://doi.org/10.1159/000369755> PMID: 25660220
51. Iwaya T, Yokobori T, Nishida N, Kogo R, Sudo T, Tanaka F, et al. Downregulation of miR-144 is associated with colorectal cancer progression via activation of mTOR signaling pathway. *Carcinogenesis*. 2012; 33: 2391–2397. <https://doi.org/10.1093/carcin/bgs288> PMID: 22983984
52. Pan X, Wang R, Wang Z-X. The Potential Role of miR-451 in Cancer Diagnosis, Prognosis, and Therapy. *Mol Cancer Ther*. 2013; 12: 1153–1162. <https://doi.org/10.1158/1535-7163.MCT-12-0802> PMID: 23814177
53. Xia Y, Chen Q, Zhong Z, Xu C, Wu C, Liu B, et al. Down-Regulation of MiR-30c Promotes the Invasion of Non-Small Cell Lung Cancer by Targeting MTA1. *Cell Physiol Biochem*. 2013; 32: 476–485. <https://doi.org/10.1159/000354452> PMID: 23988701
54. Brameier M, Herwig A, Reinhardt R, Walter L, Gruber J. Human box C/D snoRNAs with miRNA like functions: expanding the range of regulatory RNAs. *Nucleic Acids Res*. 2011; 39: 675–686. <https://doi.org/10.1093/nar/gkq776> PMID: 20846955
55. Dupuis-Sandoval F, Poirier M, Scott MS. The emerging landscape of small nucleolar RNAs in cell biology. *Wiley Interdiscip Rev RNA*. John Wiley & Sons, Inc.; 2015; 6: 381–397. <https://doi.org/10.1002/wrna.1284> PMID: 25879954
56. Burghuber OC, Strife R, Zirrolli J, Mathias MM, Murphy RC, Reeves JT, et al. Hydrogen peroxide induced pulmonary vasoconstriction in isolated rat lungs is attenuated by U60,257, a leucotriene synthesis blocker. *Wien Klin Wochenschr*. 1986; 98: 117–9. PMID: 3518246
57. Brandis KA, Gale S, Jinn S, Langmade SJ, Dudley-Rucker N, Jiang H, et al. Box C/D Small Nucleolar RNA (snoRNA) U60 Regulates Intracellular Cholesterol Trafficking. *J Biol Chem*. 2013; 288: 35703–35713. <https://doi.org/10.1074/jbc.M113.488577> PMID: 24174535
58. Graubert TA, Payton MA, Shao J, Walgren RA, Monahan RS, Frater JL, et al. Integrated Genomic Analysis Implicates Haploinsufficiency of Multiple Chromosome 5q31.2 Genes in De Novo Myelodysplastic Syndromes Pathogenesis. Schrijver I, organizer. *PLoS One*. 2009; 4: e4583. <https://doi.org/10.1371/journal.pone.0004583> PMID: 19240791
59. Liao J, Yu L, Mei Y, Guamera M, Shen J, Li R, et al. Small nucleolar RNA signatures as biomarkers for non-small-cell lung cancer. *Mol Cancer*. 2010; 9: 198. <https://doi.org/10.1186/1476-4598-9-198> PMID: 20663213
60. Jha P, Agrawal R, Pathak P, Kumar A, Purkait S, Mallik S, et al. Genome-wide small noncoding RNA profiling of pediatric high-grade gliomas reveals deregulation of several miRNAs, identifies downregulation of snoRNA cluster HBII-52 and delineates H3F3A and TP53 mutant-specific miRNAs and snoRNAs. *Int J Cancer*. 2015; 137: 2343–2353. <https://doi.org/10.1002/ijc.29610> PMID: 25994230

61. Lopez-Corral L, Mateos M V., Corchete LA, Sarasquete ME, de la Rubia J, de Arriba F, et al. Genomic analysis of high-risk smoldering multiple myeloma. *Haematologica*. 2012; 97: 1439–1443. <https://doi.org/10.3324/haematol.2011.060780> PMID: 22331267
62. Liang F, Qu H, Lin Q, Yang Y, Ruan X, Zhang B, et al. Molecular biomarkers screened by next-generation RNA sequencing for non-sentinel lymph node status prediction in breast cancer patients with meta-static sentinel lymph nodes. *World J Surg Oncol*. 2015; 13: 258. <https://doi.org/10.1186/s12957-015-0642-2> PMID: 26311227
63. Valleron W, Ysebaert L, Berquet L, Fataccioli V, Quelen C, Martin A, et al. Small nucleolar RNA expression profiling identifies potential prognostic markers in peripheral T-cell lymphoma. *Blood*. 2012; 120: 3997–4005. <https://doi.org/10.1182/blood-2012-06-438135> PMID: 22990019
64. Lan T, Ma W, Hong Z, Wu L, Chen X, Yuan Y. Long non-coding RNA small nucleolar RNA host gene 12 (SNHG12) promotes tumorigenesis and metastasis by targeting miR-199a/b-5p in hepatocellular carcinoma. *J Exp Clin Cancer Res*. 2017; 36: 11. <https://doi.org/10.1186/s13046-016-0486-9> PMID: 28073380
65. Krell J, Frampton AE, Mizezami R, Harding V, De Giorgio A, Roca Alonso L, et al. Growth Arrest-Specific Transcript 5 Associated snoRNA Levels Are Related to p53 Expression and DNA Damage in Colorectal Cancer. Calogero RA, organizador. *PLoS One*. 2014; 9: e98561. <https://doi.org/10.1371/journal.pone.0098561> PMID: 24926850
66. Gee HE, Buffa FM, Camps C, Ramachandran A, Leek R, Taylor M, et al. The small-nucleolar RNAs commonly used for microRNA normalisation correlate with tumour pathology and prognosis. *Br J Cancer*. 2011; 104: 1168–77. <https://doi.org/10.1038/sj.bjc.6606076> PMID: 21407217
67. Patnaik SK, Kannisto E, Knudsen S, Yendamuri S. Evaluation of MicroRNA Expression Profiles That May Predict Recurrence of Localized Stage I Non-Small Cell Lung Cancer after Surgical Resection. *Cancer Res*. 2010; 70: 36–45. <https://doi.org/10.1158/0008-5472.CAN-09-3153> PMID: 20028859
68. Qu X, Liu J, Zhong X, Li X, Zhang Q. PIWIL2 promotes progression of non-small cell lung cancer by inducing CDK2 and Cyclin A expression. *J Transl Med*. 2015; 13: 301. <https://doi.org/10.1186/s12967-015-0666-y> PMID: 26373553
69. Navarro A, Tejero R, Viñolas N, Cordeiro A, Marrades RM, Fuster D, et al. The significance of PIWI family expression in human lung embryogenesis and non-small cell lung cancer. *Oncotarget*. 2015; 6: 31544–56. <https://doi.org/10.18632/oncotarget.3003> PMID: 25742785
70. Moisés J, Navarro A, Tejero R, Viñolas N, Cordeiro A, Marrades RM, et al. PIWI proteins as prognostic markers in non small cell lung cancer. *Eur Respir J*. 2015; 46.
71. Qu X, Liu J, Zhong X, Li X, Zhang Q, Siegel R, et al. PIWIL2 promotes progression of non-small cell lung cancer by inducing CDK2 and Cyclin A expression. *J Transl Med*. *BioMed Central*; 2015; 13: 301. <https://doi.org/10.1186/s12967-015-0666-y> PMID: 26373553



## ANEXO 2: snoRNAs diferencialmente expressos em amostras normais

Gene	K61	K63	K65	K67	K69	T0N	T10N	T12N	T13N	T18N	T1N	logFC	logCPM	PValue	FDR
U60	782588,47	1751491,46	643885,83	740903,64	961957,52	14115,21	10499,46	13575,76	15010,69	18041,40	7180,28	6,22	18,78	1,57E-59	2,31E-57
U30	4563,00	5173,93	6496,29	6356,90	6104,03	145027,97	142326,08	387509,14	268303,58	253362,17	97019,28	-5,23	16,87	9,73E-38	7,16E-36
HBII-420	1282,35	449,67	647,65	661,49	648,21	27825,47	22165,54	17438,47	27088,82	15433,01	58468,01	-5,25	13,92	4,11E-32	2,01E-30
U44	7641,76	8126,84	9757,99	11200,34	10093,17	694,19	453,68	337,52	695,86	347,79	512,88	4,23	12,19	2,20E-31	6,63E-30
U76	9814,37	11249,88	4639,06	6240,55	6966,23	231,40	194,43	150,01	49,70	0,00	256,44	5,84	11,88	2,25E-31	6,63E-30
U15A	8490,90	8062,13	3389,91	2983,28	3326,17	231,40	0,00	37,50	149,11	130,42	170,96	5,50	11,34	2,27E-25	5,55E-24
HBII-142	3346,54	6401,53	8422,39	6365,81	6968,50	66121,65	43294,09	90417,55	62627,36	59645,31	87616,53	-3,44	15,29	2,16E-23	4,53E-22
snR38B	4026,51	5385,43	5584,89	4229,80	6270,34	99327,09	43164,47	85429,77	126248,81	58341,12	29148,52	-3,85	15,37	8,26E-23	1,52E-21
HBII-296A	3879,76	4953,10	3155,47	4025,72	3466,94	231,40	388,87	337,52	347,93	217,37	85,48	3,82	11,00	6,09E-22	9,94E-21
ACA7B	58,42	68,05	115,02	109,78	98,20	1851,17	1166,61	1050,06	2634,33	1521,56	1795,07	-4,21	9,70	8,20E-22	1,21E-20
U82	10857,92	13595,00	8100,03	9883,92	10086,36	1272,68	1685,10	1275,07	1540,83	999,89	512,88	3,10	12,44	1,40E-19	1,87E-18
U24	16056,95	23461,16	12191,81	13608,45	14059,60	2603,21	777,74	487,53	2037,87	1565,04	1025,75	3,49	12,99	6,12E-19	7,50E-18
ACA44	74,31	38,70	70,33	55,36	66,41	983,44	972,17	525,03	1192,90	782,52	1367,67	-3,96	8,77	8,07E-19	9,12E-18
HBII-202	2901,18	3608,74	2728,34	3988,19	3837,58	636,34	129,62	150,01	397,63	86,95	170,96	3,72	10,82	4,31E-17	4,52E-16
ACA58	87,86	86,07	40,30	102,74	144,74	1851,17	972,17	1012,56	1192,90	956,41	1538,63	-3,75	9,25	4,95E-17	4,85E-16
U63	15353,62	22542,46	26543,43	22125,76	27738,33	4338,69	3305,39	4200,23	4622,50	3825,65	1795,07	2,63	13,61	3,44E-16	3,16E-15
HBI-100	137,39	207,49	411,01	157,16	283,24	2313,97	2333,21	2587,64	3330,19	1999,77	8975,35	-3,89	10,90	5,68E-16	4,92E-15
U38A	532,29	522,40	825,68	539,52	651,04	3702,35	3435,01	4650,26	4026,04	4260,38	6581,92	-2,85	11,34	6,38E-16	5,21E-15
HBII-276	2953,52	2728,07	2805,27	3170,00	2878,90	57,85	259,25	600,03	49,70	347,79	170,96	3,47	10,62	7,47E-15	5,78E-14
U21	18703,44	29080,10	19602,43	19770,18	21695,60	3355,25	3888,69	4500,25	2584,62	3521,33	7094,80	2,40	13,58	1,15E-13	8,37E-13
HBII-296B	1540,78	1272,97	833,01	981,45	667,50	115,70	64,81	0,00	49,70	0,00	85,48	4,55	9,25	1,20E-13	8,37E-13
ACA3-2	141,13	38,70	63,01	79,29	36,33	809,89	842,55	675,04	646,16	782,52	854,80	-3,41	8,45	1,35E-13	9,00E-13
14q(II-1)	51,41	214,16	91,58	153,41	112,95	694,19	1685,10	8437,97	497,04	2999,66	9830,15	-5,01	11,04	1,57E-13	1,01E-12
U78	6042,56	7989,40	3205,29	5349,65	4024,33	925,59	388,87	1012,56	994,09	825,99	769,32	2,68	11,54	2,43E-13	1,49E-12

U59B	427,14	248,19	475,48	456,48	397,89	2545,37	2009,16	2662,65	3976,34	2304,08	1624,11	-2,66	10,53	1,84E-12	1,08E-11
U36B	761,28	1185,57	654,24	723,42	638,56	0,00	64,81	37,50	99,41	86,95	0,00	3,93	8,90	2,57E-12	1,45E-11
U104	13998,84	23766,06	40352,90	28543,64	33484,78	138780,25	106809,37	108381,05	170982,65	104553,19	148221,53	-2,21	16,35	9,56E-12	5,20E-11
U26	5026,12	4269,91	5829,59	6826,51	6799,35	983,44	907,36	1500,08	1789,35	956,41	940,27	2,27	11,71	1,75E-11	9,19E-11
ACA26	41,12	67,38	68,14	59,11	85,14	404,94	388,87	412,52	894,68	478,21	598,36	-3,04	7,84	2,06E-11	1,04E-10
ACA45	1441,71	2683,37	2177,40	1276,54	1660,25	10644,25	5573,79	7162,90	9791,74	9303,28	8975,35	-2,21	12,41	3,52E-11	1,72E-10
ACA3	70,10	40,03	41,03	29,09	28,95	404,94	453,68	300,02	894,68	217,37	256,44	-3,33	7,33	8,14E-11	3,86E-10
U61	6554,29	6616,36	8235,57	12710,98	13425,58	578,49	1101,80	2850,16	646,16	434,73	2735,35	2,78	12,34	3,21E-10	1,43E-09
U81	2994,18	4290,59	3448,52	4655,78	4971,09	289,25	453,68	750,04	646,16	478,21	1709,59	2,55	11,19	3,22E-10	1,43E-09
HBII-419	2586,67	3337,87	4499,12	4899,27	4836,00	22040,55	9462,48	23963,84	31661,61	18215,30	11026,86	-2,27	13,59	3,31E-10	1,43E-09
snR39B	17798,22	26530,16	13624,11	17471,38	18903,55	5727,07	2722,08	4950,28	7952,68	3130,07	2649,87	2,05	13,44	1,85E-09	7,76E-09
U73a	879,05	501,05	808,10	699,02	424,57	173,55	64,81	37,50	99,41	43,47	0,00	3,23	8,71	3,59E-09	1,46E-08
U42B	1485,64	823,29	1002,25	1988,23	1469,53	173,55	129,62	375,02	149,11	173,89	341,92	2,59	9,69	4,50E-09	1,74E-08
HBII-108B	956,15	428,99	475,48	1048,54	792,95	0,00	129,62	150,01	49,70	0,00	85,48	3,41	8,84	5,18E-09	1,95E-08
U51	629,96	594,45	698,20	632,87	573,28	2776,76	3499,82	3037,67	4026,04	2477,98	940,27	-2,17	10,76	1,11E-08	4,06E-08
U15B	1290,30	665,84	971,48	758,14	532,98	57,85	64,81	150,01	149,11	0,00	256,44	2,98	9,04	3,93E-08	1,41E-07
U28	1154,30	443,00	663,77	937,35	597,12	2718,91	5314,54	2062,62	2684,03	3130,07	3162,74	-2,06	10,96	5,07E-08	1,78E-07
HBI-43	5935,08	3497,99	2773,76	1704,40	1711,33	404,94	518,49	825,05	596,45	1173,78	341,92	2,24	10,87	6,04E-08	2,06E-07
U36C	681,37	964,07	474,02	438,65	465,44	231,40	0,00	112,51	49,70	43,47	0,00	2,96	8,61	1,10E-07	3,58E-07
U41	799,13	1246,28	567,06	1532,22	4146,36	173,55	129,62	150,01	99,41	304,31	598,36	2,85	9,93	1,22E-07	3,90E-07
U47	908,49	513,06	896,01	1042,44	841,76	57,85	518,49	0,00	49,70	43,47	85,48	2,89	9,04	7,77E-07	2,38E-06
U33	205,16	100,74	174,37	112,59	101,60	462,79	259,25	525,03	497,04	391,26	1624,11	-2,11	8,30	9,90E-06	2,80E-05
U32A	419,19	160,79	695,27	481,81	461,46	231,40	64,81	0,00	49,70	130,42	0,00	2,47	8,26	3,69E-05	9,69E-05
U80	758,94	518,39	291,59	463,04	380,30	231,40	129,62	150,01	49,70	86,95	0,00	2,10	8,40	5,15E-05	1,31E-04
mgU2-19/30	463,59	500,38	358,26	420,35	376,32	57,85	64,81	37,50	149,11	260,84	0,00	2,04	8,24	7,16E-05	1,75E-04

### ANEXO 3: snoRNAs e piRNAs diferencialmente expressos em amostras tumorais

Gene	K62	K64	K66	K68	K70	T0T	T10T	T12T	T13T	T18T	T1T	logFC	logCPM	PValue	FDR
U30	3212,44	4628,22	3651,12	5559,13	2749,54	649423,86	227655,27	440893,34	552802,19	486680,09	147937,56	-6,72	17,81	3,22E-37	5,38E-35
HBII-420	586,06	324,28	651,24	667,08	778,51	19334,48	11472,18	18097,53	16094,53	20878,04	14834,88	-4,80	13,20	8,90E-32	7,43E-30
U60	1374383,79	1270126,57	1742409,52	696181,90	641741,89	11057,66	49190,05	19873,68	21536,12	13119,92	23450,95	5,63	19,02	3,04E-30	1,69E-28
U76	9133,01	4647,45	4615,44	3690,63	4276,96	372,24	371,40	288,02	397,65	269,59	290,88	4,00	11,35	1,62E-23	6,78E-22
HBII-99	35,96	30,54	28,36	18,00	29,61	306,55	275,11	648,05	711,59	703,92	421,27	-4,15	8,07	9,61E-17	3,21E-15
U37	99,57	63,35	75,27	119,01	121,34	1160,51	1389,32	1464,12	1925,48	1078,35	451,37	-3,69	9,45	1,96E-14	5,46E-13
U58C	109,71	170,44	115,09	212,39	100,95	1620,33	797,83	1560,13	3055,66	1332,96	1975,98	-3,60	9,94	2,49E-14	5,95E-13
U15A	6911,99	2595,78	3377,85	2285,05	2043,35	153,27	591,49	120,01	188,36	104,84	210,64	3,89	10,75	1,26E-13	2,34E-12
HBII-142	2720,42	2313,36	4186,73	3567,34	3181,02	32581,77	28597,90	31322,65	11678,48	33488,74	15396,58	-3,00	13,91	1,26E-13	2,34E-12
U63	18888,66	20295,31	24019,69	17031,38	18081,95	2583,77	3920,35	1512,13	1004,60	1737,34	3881,74	3,01	13,33	1,60E-13	2,67E-12
U59B	242,17	209,28	195,81	270,98	219,87	1729,81	2132,12	1344,11	1569,69	1273,05	1364,13	-2,78	9,88	4,07E-13	6,17E-12
U24	8056,15	9648,22	8083,29	10369,09	5966,48	963,44	660,27	1440,12	1716,19	1872,13	842,55	2,76	12,15	7,04E-13	9,79E-12
HBI-43	5247,53	1262,82	2129,36	1668,01	2090,91	131,38	302,62	288,02	104,65	344,47	240,73	3,38	10,32	2,63E-12	3,38E-11
HBI-100	183,16	201,74	230,17	121,45	215,01	2233,43	1416,83	1632,14	586,02	3235,05	1233,73	-3,18	9,98	4,65E-11	5,55E-10
ACA58	42,10	43,74	59,45	59,51	48,54	656,89	412,67	1992,17	3892,83	434,34	341,03	-4,66	9,44	6,55E-11	7,29E-10
SNORD121B	478,50	355,58	285,81	262,13	178,61	21,90	13,76	24,00	0,00	59,91	20,06	3,70	7,47	2,74E-10	2,86E-09
U45B	397,98	634,99	237,26	357,65	284,42	0,00	68,78	48,00	20,93	29,95	40,12	3,39	7,75	9,07E-10	8,91E-09
U50	318,39	429,87	460,89	544,41	385,37	2233,43	1801,98	4632,39	6467,11	1183,19	2648,01	-2,88	10,89	2,24E-09	2,08E-08
ACA26	122,31	61,09	96,54	32,04	131,05	1379,47	412,67	2064,17	4562,56	1347,94	240,73	-4,23	9,84	4,11E-09	3,61E-08
HBII-180C	127,23	69,38	116,18	89,41	63,58	1072,92	385,16	1032,09	816,24	449,31	331,00	-2,85	8,60	4,74E-09	3,96E-08
U13	60,85	38,08	36,00	61,64	82,51	1423,26	13,76	3984,34	3139,38	1332,96	662,00	-4,97	9,89	6,17E-09	4,83E-08
U58B	228,34	517,35	358,35	629,55	338,29	3722,38	1678,18	2424,21	2992,87	1902,09	1424,31	-2,50	10,50	6,37E-09	4,83E-08
hsa_piR_010894-3	433,94	219,83	973,59	339,34	317,91	43,79	68,78	24,00	41,86	0,00	60,18	3,47	7,98	7,92E-09	5,75E-08
mgU6-77	169,95	142,16	251,44	253,89	247,05	919,65	880,36	1464,12	1653,40	1976,97	621,88	-2,55	9,57	1,01E-08	7,01E-08
U71b	38,11	27,15	50,73	28,99	102,90	197,07	797,83	960,08	1988,27	644,01	120,36	-3,98	8,73	1,88E-08	1,21E-07
HBII-316	100,49	134,99	116,72	153,19	87,36	919,65	591,49	1536,13	2155,70	554,15	270,82	-3,07	9,17	1,89E-08	1,21E-07

U104	13107,31	11522,66	12682,36	14415,84	7868,11	94548,44	66879,75	39723,36	24026,69	75095,06	44193,70	-2,27	15,16	2,02E-08	1,25E-07
U51	429,64	390,65	355,08	485,82	390,71	1926,88	8418,43	1320,11	1255,75	1962,00	5135,54	-3,03	10,96	2,68E-08	1,60E-07
ACA44	96,19	30,17	76,91	38,15	111,15	437,93	632,76	624,05	209,29	733,88	331,00	-2,81	8,15	4,84E-08	2,79E-07
HBII-276	3006,23	1401,96	1590,48	2061,06	1277,94	416,03	261,36	432,04	523,23	404,38	521,58	2,13	10,11	6,45E-08	3,47E-07
HBII-202	1398,62	1333,71	1199,95	1928,62	1449,27	262,76	261,36	120,01	481,37	464,29	280,85	2,22	9,74	9,65E-08	5,04E-07
U28	361,41	704,38	465,80	861,17	591,65	2561,87	3108,77	2712,23	6111,32	1887,11	1695,13	-2,33	10,89	1,29E-07	6,54E-07
HBII-296A	5229,71	1931,38	1867,01	2203,27	1158,06	175,17	605,25	312,03	104,65	434,34	651,97	2,68	10,41	1,78E-07	8,70E-07
U84	1483,14	851,06	968,69	1032,06	715,42	7291,48	20949,79	4272,36	1339,47	4717,78	34715,02	-3,60	12,80	1,82E-07	8,70E-07
HBII-436	94,04	63,35	48,54	85,75	54,36	350,34	261,36	168,01	627,88	299,54	391,18	-2,33	7,69	3,62E-07	1,68E-06
hsa_piR_001168-4	547,34	141,40	323,44	163,26	619,80	65,69	13,76	96,01	20,93	29,95	60,18	2,96	7,71	6,29E-07	2,84E-06
HBII-419	2309,84	3844,66	3093,14	3698,87	1867,65	29560,07	8666,03	17641,49	11992,41	17717,88	4222,78	-2,34	13,21	7,06E-07	3,10E-06
HBII-296B	1532,00	388,76	380,71	596,29	478,08	43,79	68,78	216,02	125,58	89,86	150,46	2,55	8,61	1,40E-06	5,99E-06
U106	420,42	309,20	199,08	146,17	412,55	897,75	990,40	1968,17	2281,28	2036,88	682,06	-2,31	9,85	1,53E-06	6,39E-06
U25	1464,70	1564,86	1986,46	2634,77	976,05	12896,95	5405,95	5184,44	6613,62	8012,74	3470,50	-2,00	12,15	2,00E-06	8,16E-06
U58A	393,07	739,07	608,16	1048,84	580,00	3809,97	1884,52	3384,29	3767,25	1407,85	1996,04	-2,00	10,78	2,29E-06	9,11E-06
mgU12-22/U4-8	209,90	194,95	210,54	249,62	163,57	0,00	55,02	24,00	20,93	14,98	80,24	2,47	7,02	3,01E-06	1,17E-05
U36B	476,35	501,13	291,81	437,60	346,54	0,00	96,29	24,00	188,36	44,93	40,12	2,65	7,92	3,81E-06	1,44E-05
U83A	108,18	286,58	179,99	291,43	338,29	1072,92	1788,23	2064,17	1276,68	599,08	431,30	-2,32	9,54	4,69E-06	1,74E-05
U97	303,33	200,60	283,62	143,73	379,55	21,90	82,53	48,00	20,93	59,91	110,33	2,08	7,39	2,19E-05	7,64E-05
mgU2-25/61	137,68	86,35	85,09	51,27	212,59	394,13	1554,38	456,04	397,65	299,54	260,79	-2,29	8,43	5,61E-05	1,80E-04
14q(II-1)	64,54	58,82	130,36	85,75	211,62	175,17	481,45	4584,39	83,72	2336,42	210,64	-3,57	9,52	5,89E-05	1,86E-04
ACA7B	254,16	94,65	311,44	76,90	126,68	1532,74	275,11	1104,09	669,73	524,20	601,82	-2,17	8,92	6,09E-05	1,86E-04
ACA3	115,25	30,54	47,45	28,99	102,41	197,07	261,36	360,03	125,58	599,08	240,73	-2,20	7,47	6,11E-05	1,86E-04
U85	37,19	36,95	37,63	29,91	45,14	43,79	247,60	72,01	104,65	104,84	421,27	-2,22	6,65	7,02E-05	2,09E-04
Z17B	112,48	90,88	66,00	73,85	55,33	0,00	13,76	0,00	20,93	29,95	10,03	2,58	5,81	7,93E-05	2,28E-04
snR38B	4048,05	3574,67	5181,05	2906,67	2273,41	16947,78	35173,08	7536,64	3871,90	18736,32	6790,54	-2,05	13,25	8,10E-05	2,29E-04
U82	8460,59	6273,02	4618,16	6182,88	2920,87	1051,02	2104,61	600,05	376,73	284,56	3791,47	2,05	11,71	3,69E-04	9,78E-04
ACA8	46,41	35,45	53,45	51,57	61,15	0,00	13,76	24,00	0,00	14,98	10,03	2,21	5,26	1,95E-03	4,40E-03
U90	121,70	126,32	93,81	134,58	428,57	65,69	27,51	144,01	41,86	0,00	20,06	2,00	6,93	3,64E-03	7,42E-03

## ANEXO 4: snoRNAs constitutivamente expressos em não fumantes

Genes	K61	K62	K63	K64	K65	K66	K67	K68	K69	K70	Var	SD
U51	8,45	8,48	8,54	8,58	8,79	8,37	8,57	8,76	8,53	8,35	0,02	0,15
U57	12,20	12,38	12,34	12,30	12,07	11,92	12,37	12,36	12,25	11,67	0,05	0,23
HBII-429	9,42	9,37	9,57	9,40	9,81	9,60	9,34	8,92	9,72	9,42	0,06	0,25
HBII-95	5,89	5,62	5,37	5,94	5,87	5,94	5,77	6,12	5,38	5,63	0,06	0,25
U21	13,34	13,27	14,15	13,56	13,60	13,52	13,54	13,68	13,77	13,30	0,07	0,26
U35B	4,98	4,85	4,84	4,14	4,83	4,45	4,93	4,82	4,96	4,74	0,07	0,26
U37	6,65	6,37	5,98	5,95	6,25	6,13	6,21	6,73	6,32	6,67	0,08	0,28
U75	9,08	9,38	8,77	9,23	8,66	8,88	9,01	8,94	8,90	8,42	0,08	0,28
U95	12,08	12,11	12,30	11,66	12,48	11,85	12,26	11,87	12,36	12,45	0,08	0,28
ACA8	5,96	5,27	4,95	5,11	5,50	5,64	5,45	5,52	5,36	5,68	0,08	0,29
mgh28S-2411	11,48	12,00	11,48	12,16	12,15	12,05	12,01	12,25	11,97	11,57	0,08	0,29
U15B	9,48	8,86	8,70	8,81	9,27	8,98	8,83	8,77	8,42	8,95	0,09	0,29
U38B	9,69	9,60	9,17	9,56	10,27	9,62	9,92	9,87	9,85	9,86	0,08	0,29
U59B	7,89	7,65	7,28	7,67	8,24	7,51	8,10	7,92	8,00	7,52	0,09	0,3
HBII-276	10,68	11,28	10,73	10,42	10,80	10,53	10,90	10,84	10,86	10,06	0,11	0,32
HBII-85-24	5,90	6,18	5,69	5,89	6,09	5,13	6,07	5,72	5,98	5,48	0,1	0,32
HBII-85-23	4,57	4,71	4,87	5,02	4,32	4,27	4,90	4,43	4,55	3,95	0,11	0,33
U20	10,71	10,54	11,16	11,11	10,63	10,23	10,56	10,52	11,11	10,37	0,11	0,33
HBII-316	6,20	6,38	6,56	7,04	6,67	6,77	6,91	7,09	6,97	6,19	0,11	0,34
mgU2-19/30	8,01	8,68	8,29	7,70	7,83	7,97	7,98	7,75	7,92	7,40	0,12	0,34
U17b	8,12	8,01	8,01	7,98	8,33	8,49	7,76	8,02	7,76	8,86	0,12	0,34
U47	8,98	8,76	8,32	8,86	9,15	9,05	9,29	8,80	9,08	9,58	0,11	0,34
U88	5,30	5,57	6,16	5,74	5,80	5,70	5,16	5,43	5,30	5,04	0,12	0,34
snR38C	7,60	8,41	7,45	8,11	8,41	8,50	8,19	8,34	7,97	7,84	0,13	0,36
U101	10,15	10,49	10,42	10,15	9,93	10,19	10,14	10,11	10,16	9,18	0,13	0,36
U74	11,15	11,11	11,15	10,58	11,04	11,33	10,53	10,66	10,61	11,60	0,13	0,36
U81	10,70	11,21	11,39	11,03	11,10	11,32	11,45	11,18	11,64	10,44	0,13	0,36
U87	7,18	7,04	6,99	6,65	7,76	6,81	6,70	6,81	7,16	6,51	0,13	0,36
SNORD127	8,28	7,61	7,94	8,26	8,72	8,10	8,31	8,20	8,66	8,80	0,13	0,37
U27	13,88	14,26	14,27	14,51	14,37	14,61	14,52	14,97	14,38	13,64	0,14	0,37
U30	11,30	11,38	11,66	12,14	12,01	11,73	11,90	12,28	11,94	11,17	0,14	0,37
hsa-mir-1201	6,40	6,66	6,93	7,61	7,11	6,68	7,28	6,74	7,18	6,51	0,15	0,38
mgU12-22/U4-8	7,86	7,44	8,05	7,57	8,44	7,62	8,05	7,80	7,98	7,10	0,14	0,38
U59A	8,66	8,93	9,29	8,85	9,56	8,88	9,47	9,06	9,31	8,33	0,15	0,38
U63	13,05	13,93	13,78	14,27	14,04	14,45	13,70	13,89	14,12	13,89	0,14	0,38
U84	9,91	10,26	10,58	9,70	10,27	9,82	10,07	9,85	10,31	9,23	0,15	0,38
HBII-234	5,92	5,95	5,13	6,45	6,12	6,15	5,94	6,17	5,73	6,49	0,15	0,39
HBII-85-15	5,17	4,74	4,52	5,24	5,27	4,88	5,73	5,20	5,18	4,43	0,15	0,39
U26	11,44	11,01	11,38	11,87	11,86	11,83	12,00	12,22	12,10	11,38	0,15	0,39

HBII-210	10,69	11,71	11,55	12,00	11,97	12,00	11,65	11,91	11,87	11,40	0,16	0,4
SNORD121B	7,91	8,63	8,39	8,44	8,24	8,06	7,84	7,87	7,84	7,22	0,16	0,4
U24	13,12	12,70	13,84	13,20	12,92	12,88	13,00	13,18	13,14	12,29	0,16	0,4
U36B	8,72	8,62	9,53	8,93	8,70	8,09	8,76	8,61	8,68	8,18	0,16	0,4
U83	7,74	7,19	7,59	7,03	7,50	7,38	7,02	6,52	7,21	7,88	0,16	0,4
snR38B	11,12	11,71	11,71	11,77	11,79	12,24	11,31	11,34	11,98	10,89	0,17	0,41
U50	8,39	8,04	8,14	8,71	8,56	8,75	9,09	8,92	9,31	8,33	0,17	0,41
HBII-85-29	5,04	5,68	5,28	5,39	4,87	4,38	5,67	5,26	5,76	5,29	0,18	0,42
U31	13,46	13,68	13,68	14,22	14,01	14,38	13,94	14,62	14,24	13,33	0,17	0,42
HBII-82B	7,90	7,22	7,82	6,93	7,91	6,85	7,83	7,65	7,92	7,96	0,19	0,43
U48	10,90	12,06	11,17	11,48	11,15	11,92	11,03	11,41	12,09	11,53	0,18	0,43
U53	6,77	7,62	7,49	7,26	6,74	6,40	7,38	6,63	7,32	6,64	0,19	0,43
U96a	8,37	7,40	7,28	8,05	8,03	8,51	7,92	8,13	7,65	8,52	0,19	0,43
Z17B	6,66	6,54	6,37	6,47	5,65	5,94	5,91	6,04	5,51	5,53	0,18	0,43
HBII-202	10,65	10,18	11,14	10,35	10,76	10,13	11,23	10,75	11,27	10,24	0,19	0,44
U105	8,07	8,11	8,22	8,03	8,92	8,70	8,54	8,81	8,57	7,49	0,19	0,44
U36C	8,56	8,80	9,23	8,58	8,23	7,85	8,04	7,83	8,23	8,18	0,19	0,44
U42B	9,69	9,13	9,01	9,68	9,32	9,23	10,22	9,79	9,89	8,87	0,19	0,44
U45B	8,69	8,37	7,96	9,28	8,10	7,79	8,38	8,32	8,16	7,89	0,19	0,44
U80	8,72	8,46	8,34	7,53	7,53	7,46	8,12	7,72	7,94	8,31	0,2	0,44
HBII-420	9,47	8,92	8,13	8,31	8,69	9,25	8,63	9,22	8,71	9,35	0,2	0,45
U23	4,53	4,38	4,35	4,77	4,25	5,03	4,67	4,21	3,46	3,91	0,2	0,45
U76	12,41	12,89	12,78	12,15	11,53	12,07	11,87	11,68	12,13	11,81	0,2	0,45
HBII-239	9,99	9,22	10,17	9,47	10,73	9,51	10,05	9,81	10,36	9,71	0,21	0,46
HBII-55	9,74	10,07	10,51	10,07	10,06	9,85	10,14	9,99	10,17	8,77	0,21	0,46
U18B	7,44	6,92	6,32	7,39	7,31	7,25	6,89	7,19	6,62	7,94	0,21	0,46
U73a	8,93	8,68	8,29	7,75	9,00	8,03	8,71	8,57	8,10	9,11	0,21	0,46
U97	8,09	7,97	8,45	7,61	7,95	8,05	7,29	7,00	7,46	8,31	0,21	0,46
HBII-108B	9,05	8,69	8,06	9,18	8,24	8,67	9,30	9,36	9,00	8,25	0,22	0,47
HBII-251	11,12	11,24	11,78	12,10	12,10	12,07	11,84	12,02	12,45	11,08	0,22	0,47
U105B	5,86	5,45	5,44	5,73	6,67	6,43	6,67	6,52	6,11	6,13	0,22	0,47
U49A	7,39	7,55	6,99	6,70	6,65	6,99	6,14	6,30	6,30	6,49	0,22	0,47
ACA61	6,86	7,14	6,88	6,33	7,15	7,25	6,26	6,94	6,72	7,93	0,23	0,48
HBII-419	10,49	10,90	11,02	11,87	11,48	11,49	11,52	11,69	11,60	10,61	0,23	0,48
HBII-99	4,26	4,90	3,32	4,90	4,15	4,73	4,51	4,01	4,57	4,63	0,23	0,48
snR39B	13,27	13,08	14,02	13,12	13,08	12,99	13,36	13,29	13,57	12,13	0,23	0,48
U13	6,33	5,66	5,41	5,22	6,40	5,07	6,31	5,78	5,77	6,11	0,23	0,48
U55	6,75	6,03	5,91	6,05	6,71	5,81	5,41	5,63	5,59	6,58	0,23	0,48
U58C	7,61	6,51	6,49	7,38	7,23	6,75	7,40	7,57	7,37	6,40	0,23	0,48
HBII-85-20	4,61	4,87	3,44	4,49	4,54	4,38	4,51	3,93	4,97	3,79	0,24	0,49
U28	9,32	8,23	8,11	9,43	8,72	8,76	9,14	9,59	8,59	8,95	0,24	0,49
U50B	8,44	8,28	8,93	7,59	7,81	7,71	8,15	7,95	8,30	7,20	0,24	0,49
U82	12,55	12,78	13,05	12,58	12,33	12,07	12,54	12,43	12,67	11,26	0,24	0,49
U94	6,03	5,18	6,00	5,43	6,83	6,06	5,50	5,32	5,69	5,51	0,24	0,49
ACA58	5,61	5,12	5,75	5,42	4,68	5,79	5,95	5,73	6,54	5,34	0,25	0,5
HBII-142	10,86	11,14	11,96	11,14	12,39	11,93	11,90	11,64	12,13	11,38	0,25	0,5
mgh18S-121	9,42	8,88	8,83	10,14	9,16	8,53	9,57	9,21	9,99	9,38	0,25	0,5

mgU6-77	7,73	7,14	6,48	7,12	8,02	7,87	7,92	7,82	7,11	7,69	0,25	0,5
SNORD123	6,58	7,53	7,87	6,79	8,08	7,10	7,76	7,35	7,76	7,83	0,25	0,5
U8	8,92	7,17	7,76	7,75	7,91	8,03	7,37	7,64	7,40	7,39	0,25	0,5
U25	9,62	10,25	9,95	10,58	10,55	10,86	10,63	11,20	10,07	9,67	0,27	0,52
U54	6,40	6,78	6,24	6,61	6,65	5,70	6,84	6,46	6,62	5,21	0,27	0,52
HBII-85-1	4,85	5,05	4,49	4,86	4,29	3,78	5,02	4,08	5,02	3,65	0,28	0,53
SNORD119	12,99	13,43	13,70	12,79	12,89	12,53	13,19	12,73	13,55	11,89	0,29	0,53
U35A	4,92	4,33	3,54	4,20	4,96	4,42	4,99	4,78	4,01	5,25	0,28	0,53
U43	12,19	11,64	12,48	13,20	12,89	13,05	13,01	13,24	12,90	12,16	0,29	0,53
HBII-295	11,64	10,07	11,07	11,70	11,55	11,76	11,85	11,84	11,56	11,30	0,29	0,54
hsa-mir-1259	7,07	6,95	5,91	7,90	6,33	6,88	7,00	7,09	6,39	6,65	0,29	0,54
U56	6,96	6,48	5,53	6,69	7,36	6,46	7,36	7,16	6,63	6,96	0,3	0,54
ACA33	4,48	4,98	3,74	5,24	4,36	5,05	4,06	3,74	4,16	4,83	0,3	0,55
U65	7,27	6,82	7,50	7,46	8,25	7,11	6,67	7,01	6,66	8,11	0,31	0,55
ACA57	6,54	8,17	8,34	7,78	8,04	8,26	7,71	7,50	8,03	7,13	0,32	0,56
HBII-240	7,24	6,32	5,96	7,24	5,64	6,15	6,37	6,83	5,85	6,20	0,31	0,56
HBII-296A	11,07	12,08	11,59	10,88	10,97	10,77	11,24	10,94	11,12	9,92	0,31	0,56
HBII-436	6,23	6,28	5,87	5,95	7,01	5,50	7,05	6,26	6,80	5,51	0,32	0,56
U61	11,83	11,83	12,01	12,13	12,35	11,68	12,90	11,84	13,08	11,23	0,31	0,56
U83B	7,73	6,63	7,21	8,59	8,28	7,87	8,09	7,94	8,14	7,81	0,31	0,56
HBI-100	6,25	7,25	7,02	7,62	8,03	7,75	6,56	6,76	7,51	7,49	0,32	0,57
U16	6,57	5,81	5,34	5,07	5,73	4,88	4,60	5,20	5,01	5,59	0,32	0,57
ACA48	6,19	6,99	6,00	5,93	7,04	6,35	6,13	5,70	5,48	7,13	0,33	0,58
U52	11,75	12,22	11,13	10,67	10,57	11,26	10,46	11,11	10,48	10,70	0,34	0,58
U58B	8,66	7,56	7,61	8,98	8,81	8,38	9,14	9,13	8,70	8,15	0,34	0,58
U64	6,43	7,06	6,87	5,84	5,67	5,83	5,87	5,73	5,16	6,22	0,34	0,58
U71b	5,15	4,98	4,74	4,73	5,58	5,56	4,65	4,69	4,68	6,43	0,34	0,58
ACA45	9,64	10,98	10,71	9,98	10,43	9,99	9,58	9,17	10,06	10,85	0,35	0,59
HBII-13	9,14	9,15	9,42	9,19	9,75	8,89	9,68	9,22	9,82	7,77	0,34	0,59
HBII-296B	9,74	10,31	9,63	8,57	9,05	8,47	9,20	9,06	8,75	8,64	0,35	0,59
U29	9,16	8,06	8,48	9,00	9,45	9,01	9,81	9,78	9,37	9,85	0,35	0,59
U33	6,83	6,42	5,97	6,41	6,79	6,36	6,08	6,41	6,03	8,00	0,35	0,59
U85	5,94	4,95	5,56	5,17	6,61	5,13	4,74	4,74	4,78	5,24	0,36	0,6
U93	4,88	5,18	5,94	6,05	6,27	6,11	5,30	4,39	5,71	5,76	0,37	0,6
U104	12,92	13,41	13,86	13,46	14,65	13,53	14,07	13,65	14,40	12,68	0,37	0,61
U22	7,32	6,65	5,69	6,38	6,87	6,43	6,20	6,56	5,41	7,22	0,37	0,61
U44	12,05	12,68	12,31	13,89	12,60	13,87	12,72	13,27	12,67	12,70	0,38	0,61
U90	7,72	6,66	6,93	6,95	7,69	6,45	6,98	6,91	7,51	8,49	0,38	0,61
U18A	8,68	8,09	8,11	9,16	8,81	8,69	9,72	9,13	10,01	8,75	0,38	0,62
SNORD124	5,14	5,67	5,29	5,72	5,07	4,42	5,88	4,82	5,91	4,02	0,4	0,63
ACA42	3,78	4,83	4,06	4,93	5,11	4,38	4,90	5,47	4,36	5,85	0,4	0,64
U38A	8,20	9,46	8,35	10,16	9,04	9,56	8,34	8,53	8,71	8,81	0,4	0,64
U42A	9,41	8,46	8,71	10,11	9,84	9,93	10,11	9,97	10,20	10,26	0,41	0,64
ACA44	5,36	6,32	4,59	4,88	5,48	6,16	5,06	5,09	5,42	6,54	0,42	0,65
snR38A	6,53	7,52	6,58	8,58	6,94	7,82	6,88	7,44	6,76	6,88	0,42	0,65
U15A	12,20	12,48	12,30	11,31	11,07	11,62	10,81	10,99	11,06	10,74	0,42	0,65
U58A	9,76	8,35	7,90	9,49	9,58	9,15	9,73	9,87	9,25	8,92	0,42	0,65

HBII-135	8,72	8,44	8,27	7,70	7,44	6,98	9,22	8,15	7,97	7,56	0,43	0,66
HBII-180C	4,62	6,72	5,20	6,08	6,11	6,76	6,33	6,32	5,96	5,73	0,44	0,66
HBII-336	10,22	9,71	11,37	10,66	11,58	11,19	11,03	11,21	11,47	11,89	0,44	0,66
hsa_piR_010894-3	8,42	8,49	8,44	7,75	7,95	9,83	7,31	8,24	8,25	8,06	0,43	0,66
U106	8,78	8,44	8,26	8,24	6,86	7,54	8,30	7,03	8,51	8,43	0,44	0,66
U45A	9,26	9,12	9,47	10,79	9,51	9,21	10,20	9,77	10,70	8,92	0,43	0,66
ACA16	6,52	6,15	4,69	5,34	5,98	6,37	6,22	6,33	6,04	7,15	0,45	0,67
ACA18	5,49	5,36	4,78	5,75	5,58	5,74	4,87	5,04	5,04	7,12	0,45	0,67
ACA21	4,18	4,73	3,87	4,23	4,85	3,78	3,93	3,03	4,74	5,34	0,44	0,67
HBI-43	11,68	12,09	11,09	10,27	10,78	10,96	10,00	10,54	10,11	10,77	0,45	0,67
mgh28S-2409	9,15	8,16	7,46	9,33	8,86	8,98	9,57	9,76	9,11	9,24	0,46	0,68
mgU2-25/61	6,96	6,83	6,90	6,40	7,88	6,31	6,06	5,52	6,41	7,47	0,47	0,69
U83A	8,10	6,49	6,25	8,13	8,13	7,39	7,90	8,02	7,51	8,15	0,5	0,71
U46	7,72	7,04	5,91	6,56	6,95	5,82	6,85	6,87	6,27	8,15	0,54	0,73
U60	18,73	20,12	20,06	20,24	18,64	20,63	18,76	19,24	19,24	19,03	0,53	0,73
14q(II-1)	4,83	5,74	7,06	5,84	5,86	6,93	6,53	6,26	6,18	7,47	0,58	0,76
14q(II-3)	5,02	5,61	6,06	5,57	5,34	7,29	6,25	6,42	5,82	7,28	0,59	0,76
ACA7	5,83	7,59	5,98	6,92	7,04	8,55	6,92	6,91	7,14	7,18	0,58	0,76
U34	8,65	8,66	8,58	8,22	8,54	7,52	7,64	7,67	8,26	10,17	0,58	0,76
ACA26	4,51	6,66	5,39	5,90	5,44	6,49	5,15	4,84	5,78	6,78	0,6	0,77
U79	9,16	8,56	8,35	7,89	6,90	6,79	8,41	8,07	8,01	8,92	0,6	0,77
14q(0)	5,53	6,55	7,20	5,92	5,41	7,22	5,80	5,69	5,83	7,44	0,6	0,78
ACA3-2	6,29	5,63	4,59	4,51	5,32	4,48	5,57	5,28	4,55	6,77	0,63	0,79
HBII-95B	6,77	5,49	4,76	4,42	4,83	4,38	4,95	5,10	3,87	5,44	0,63	0,79
U45C	7,07	6,24	5,76	8,70	6,76	7,35	6,30	7,12	6,61	6,55	0,65	0,8
E2	6,16	4,64	4,41	6,33	6,00	6,04	4,32	5,18	5,38	6,39	0,65	0,81
hsa_piR_020828-2	9,01	9,04	9,94	7,78	8,51	8,91	7,80	7,24	8,19	9,21	0,66	0,81
U18C	7,00	6,73	5,00	8,08	6,04	7,05	6,99	7,14	6,39	6,88	0,65	0,81
ACA3	5,28	6,58	4,64	4,90	4,70	5,47	4,13	4,69	4,22	6,42	0,71	0,84
ACA50	5,35	3,83	3,95	5,19	4,92	3,73	4,62	4,37	4,89	6,52	0,71	0,84
ACA63	5,43	4,98	7,08	5,31	7,13	6,41	7,14	5,61	6,89	5,79	0,71	0,84
U17a	7,55	7,07	6,61	7,12	8,26	7,58	6,07	6,81	6,74	8,92	0,7	0,84
14q(II-17)	4,61	4,50	5,54	5,02	4,98	6,54	5,58	5,57	5,56	7,27	0,72	0,85
HBII-166	6,07	4,11	5,29	4,31	5,90	3,93	3,70	3,85	4,48	4,81	0,72	0,85
HBII-289	11,35	8,81	10,76	10,35	12,06	10,76	10,86	10,61	11,51	11,04	0,74	0,86
U32A	7,86	7,64	6,65	7,81	8,79	7,83	8,18	8,48	8,22	9,99	0,75	0,86
14q(I-8)	3,22	4,45	4,91	4,75	4,29	6,10	4,93	5,20	4,70	6,29	0,77	0,88
U78	11,71	13,10	12,28	13,98	10,99	12,99	11,65	12,75	11,34	12,39	0,84	0,92
14q(II-26)	3,44	5,33	4,93	5,07	4,78	7,14	4,89	5,67	5,19	5,95	0,89	0,94
HBII-85-22	4,77	5,68	5,41	4,29	3,42	3,55	4,67	4,01	5,75	3,16	0,89	0,94
hsa_piR_009294	10,98	11,77	12,88	12,50	13,62	13,50	13,38	12,73	13,67	11,60	0,89	0,94
U3	6,51	4,49	6,46	6,93	6,62	5,45	4,44	4,65	5,64	6,06	0,88	0,94
U102	10,26	8,67	9,58	9,90	10,57	11,31	9,19	8,68	9,80	8,28	0,89	0,95
ACA7B	5,02	7,72	5,41	6,53	6,19	8,18	6,04	6,10	5,98	6,73	0,93	0,96
14q(II-14)	3,61	4,14	4,82	4,42	5,07	6,26	4,79	5,29	5,42	6,99	0,99	0,99
hsa_piR_018165	7,70	7,93	7,04	7,04	5,53	6,59	5,42	5,42	5,82	7,56	0,98	0,99



## ANEXO 5: snoRNAs constitutivamente expressos em fumantes

Gene	T0N	T0T	T10N	T10T	T12N	T12T	T13N	T13T	T18N	T18T	T1N	T1T	Var	SD
U59B	11,50	11,05	11,56	11,58	11,53	11,03	11,90	11,49	11,74	10,87	11,08	11,23	0,1	0,32
U25	13,24	13,95	12,68	12,93	12,97	12,98	13,14	13,57	13,13	13,52	12,62	12,57	0,17	0,41
U57	12,71	12,29	12,37	12,40	12,28	11,78	12,25	12,91	12,67	11,72	11,69	11,88	0,17	0,41
U48	12,81	13,26	13,23	13,56	13,49	13,46	13,34	12,75	13,38	14,38	13,39	12,89	0,19	0,43
U27	14,04	14,30	13,27	14,65	14,07	13,93	14,05	14,04	13,93	13,40	12,96	13,92	0,21	0,46
U21	11,90	11,95	12,52	11,94	12,28	12,87	11,28	12,62	12,36	12,85	13,21	12,03	0,29	0,53
HBII-420	14,95	14,53	15,03	14,01	14,24	14,78	14,67	14,85	14,49	14,90	16,25	14,67	0,3	0,55
U20	13,15	13,07	13,00	11,70	12,94	12,96	12,70	13,61	12,90	11,75	12,98	12,22	0,33	0,57
U43	13,73	12,91	12,84	12,45	13,74	13,25	13,43	12,72	14,26	13,67	14,02	12,59	0,36	0,6
snR39B	12,67	12,55	12,00	11,33	12,42	13,40	12,90	12,97	12,19	12,85	11,79	13,31	0,38	0,62
U63	12,27	11,63	12,28	12,46	12,18	11,20	12,12	10,85	12,47	11,32	11,23	12,73	0,38	0,62
U104	17,27	16,82	17,30	16,55	16,87	15,92	17,33	15,43	17,25	16,75	17,59	16,24	0,42	0,65
HBII-336	12,08	11,69	10,78	12,76	11,98	10,98	11,22	10,43	12,02	11,32	11,00	11,58	0,43	0,66
HBII-419	14,61	15,15	13,80	13,61	14,70	14,75	14,90	14,43	14,73	14,67	13,85	12,86	0,44	0,66
SNORD119	13,00	12,25	12,70	13,11	12,27	12,45	12,61	13,67	12,57	10,82	12,67	12,45	0,45	0,67
U28	11,59	11,62	12,97	12,13	11,16	12,05	11,34	13,45	12,19	11,44	12,04	11,54	0,46	0,68
U42A	11,87	12,52	11,61	11,70	11,46	12,97	11,71	13,60	11,66	11,92	12,59	11,46	0,46	0,68
U38A	12,04	10,90	12,34	11,37	12,33	11,98	11,92	13,31	12,63	11,57	13,10	11,76	0,48	0,69
U59A	11,21	10,72	11,31	11,46	11,60	11,01	11,36	10,72	10,82	10,01	9,16	11,29	0,49	0,7
HBII-55	12,04	10,85	11,90	10,98	11,21	9,81	12,01	10,66	11,95	10,29	11,00	10,80	0,52	0,72
U31	15,83	15,19	16,15	15,06	15,28	14,09	15,84	14,40	15,77	14,12	14,70	14,35	0,53	0,73
HBII-142	16,20	15,29	15,99	15,33	16,61	15,58	15,88	14,39	16,44	15,58	16,84	14,72	0,55	0,74
ACA45	13,56	12,18	13,04	11,64	12,95	11,67	13,20	12,50	13,76	12,01	13,55	12,17	0,57	0,75
HBII-210	11,85	11,05	12,25	12,32	10,99	11,16	11,67	11,47	11,30	11,24	11,16	13,63	0,58	0,76
U52	11,99	11,79	11,70	12,15	12,12	11,20	12,41	9,72	11,97	12,52	11,00	11,14	0,6	0,78
U95	13,33	13,61	13,60	14,44	13,36	12,87	13,48	12,38	13,14	13,69	13,08	15,54	0,64	0,8
HBII-251	13,07	11,81	12,90	11,38	12,92	11,43	12,97	11,05	13,24	11,63	12,86	11,67	0,65	0,81
HBI-100	11,36	11,42	11,78	10,99	11,48	11,31	11,65	10,07	11,54	12,21	13,55	11,08	0,67	0,82
U60	13,97	13,73	13,95	16,11	13,88	14,92	13,82	15,27	14,71	14,23	13,23	15,33	0,71	0,84
U51	11,62	11,21	12,37	13,56	11,72	11,01	11,92	11,17	11,85	11,49	10,29	13,14	0,82	0,91
U50	10,50	11,42	12,07	11,34	11,18	12,82	10,79	13,54	10,72	10,76	11,08	12,18	0,88	0,94
HBII-295	13,54	12,37	14,96	12,56	13,67	12,31	13,01	12,36	14,02	12,76	14,72	12,39	0,9	0,95
U30	17,33	19,60	17,71	18,32	18,71	19,39	17,98	19,95	18,52	19,45	16,98	17,99	0,9	0,95

## ANEXO 6: miRNAs diferencialmente expressos em linfócitos

feature_id	posterior_probability	eta1_0
hsa-miR-146b-3p	0,999869	3,05742
hsa-miR-342-5p	0,999739	5,04651
hsa-miR-193b-5p	0,999318	3,81967
hsa-miR-1287-5p	0,999212	1,27364
hsa-miR-95-3p	0,998955	4,9137
hsa-miR-338-5p	0,998522	-3,0869
hsa-miR-1271-5p	0,99818	2,33906
hsa-miR-150-3p	0,997138	6,03672
hsa-miR-1275	0,997112	3,03411
hsa-miR-574-3p	0,997095	-4,06859
hsa-miR-3934-5p	0,996616	2,61296
hsa-miR-150-5p	0,99633	7,21831
hsa-miR-345-5p	0,995244	-3,06912
hsa-miR-145-5p	0,995007	-3,12762
hsa-miR-342-3p	0,994858	3,83414
hsa-miR-550a-5p	0,994478	-2,43206
hsa-miR-425-3p	0,9941	-2,97963
hsa-miR-2110	0,992713	1,7676
hsa-miR-574-5p	0,991372	-2,07666
hsa-miR-223-3p	0,989607	-4,07139
hsa-miR-361-3p	0,988879	2,84325
hsa-miR-296-3p	0,988761	1,55128
hsa-miR-1468-5p	0,986798	2,53059
hsa-miR-126-5p	0,986086	-4,12525
hsa-miR-548j-5p	0,985921	-1,5908
hsa-miR-505-5p	0,985566	2,10168
hsa-miR-99a-5p	0,985454	2,15293
hsa-miR-584-5p	0,984335	-4,36535
hsa-miR-766-5p	0,983865	1,68396
hsa-miR-363-5p	0,983149	2,5477
hsa-miR-126-3p	0,982982	-4,11052
hsa-miR-625-3p	0,97601	1,44179
hsa-miR-486-5p	0,975515	-2,65001
hsa-miR-185-3p	0,973376	-2,0776
hsa-miR-195-5p	0,971662	3,10778
hsa-miR-4518	0,97102	1,69605
hsa-miR-223-5p	0,9701	-3,06026
hsa-miR-193b-3p	0,968917	3,40213
hsa-miR-874-3p	0,965563	2,62886
hsa-miR-200c-3p	0,962511	1,3484
hsa-miR-92a-1-5p	0,96034	1,74616

hsa-miR-455-5p	0,959204	2,55911
hsa-miR-199b-5p	0,954176	-2,90949
hsa-miR-30d-5p	0,944011	1,66738
hsa-miR-141-3p	0,94244	3,3578
hsa-miR-194-5p	0,937106	2,03014
hsa-miR-557	0,933134	1,57353
hsa-miR-4778-5p	0,929115	2,36217
hsa-miR-548b-5p	0,928916	1,77239
hsa-miR-146b-5p	0,925996	3,34925
hsa-miR-125b-5p	0,92556	2,80879
hsa-miR-625-5p	0,922447	1,4885
hsa-let-7g-3p	0,920718	2,59447
hsa-miR-155-5p	0,913685	1,59197
hsa-miR-6769b-3p	0,908984	1,60543

**ANEXO 7: miRNAs diferencialmente expressos em megacariócitos e eritroblatos**

feature_id	posterior_probability	etal_1
hsa-miR-193a-5p	0,999494	-4,47219
hsa-miR-296-3p	0,999155	3,15907
hsa-miR-1249-3p	0,99779	-3,53205
hsa-miR-582-5p	0,995581	-3,85943
hsa-miR-338-5p	0,995391	-3,81237
hsa-miR-1275	0,994314	2,94745
hsa-miR-618	0,993322	-3,74614
hsa-miR-335-3p	0,991247	5,02157
hsa-miR-941	0,989099	-2,59448
hsa-miR-3690	0,985986	-4,60131
hsa-miR-3909	0,98172	-2,43467
hsa-miR-3615	0,980803	-1,80311
hsa-miR-3605-3p	0,979117	-2,41914
hsa-miR-1287-5p	0,973278	0,912274
hsa-miR-629-3p	0,972928	-1,67688
hsa-miR-191-5p	0,966031	-1,58239
hsa-miR-582-3p	0,961042	-4,14954
hsa-miR-3614-5p	0,958231	-2,97705
hsa-miR-663a	0,953849	2,67046
hsa-let-7b-3p	0,948477	-2,95843
hsa-miR-361-3p	0,947327	-2,15164
hsa-miR-135a-5p	0,94414	3,2574
hsa-miR-6724-5p	0,94401	1,83649
hsa-miR-940	0,94082	1,94991
hsa-miR-2110	0,937826	-1,59903
hsa-miR-486-3p	0,93044	4,38793
hsa-miR-10a-3p	0,929789	2,71702
hsa-miR-151a-5p	0,916771	2,88747
hsa-miR-197-3p	0,913718	-1,775
hsa-miR-1343-3p	0,912214	-1,6636
hsa-miR-20b-5p	0,911952	2,16427
hsa-miR-145-3p	0,911709	-1,86574
hsa-miR-516a-5p	0,911078	2,35014
hsa-miR-151a-3p	0,90933	2,61307
hsa-miR-1306-5p	0,906156	-1,69451
hsa-let-7e-5p	0,905923	2,08614
hsa-miR-4516	0,901787	2,13603

## ANEXO 8: miRNAs diferencialmente expressos em neutrófilos

feature_id	posterior_probability	eta1_2
hsa-miR-145-5p	1	4,17071
hsa-miR-4272	1	0,022371
hsa-miR-2115-3p	0,999999	3,16739
hsa-miR-4772-3p	0,999998	5,83885
hsa-miR-145-3p	0,999997	4,29442
hsa-miR-2115-5p	0,999997	2,77919
hsa-miR-339-5p	0,999996	-2,37352
hsa-miR-1271-5p	0,999994	-2,11669
hsa-miR-194-5p	0,999993	2,89082
hsa-miR-338-5p	0,999993	2,40402
hsa-miR-7706	0,999991	-1,97693
hsa-miR-144-3p	0,999988	4,43926
hsa-miR-4772-5p	0,999988	4,98156
hsa-miR-511-5p	0,999985	-3,4488
hsa-miR-1268b	0,999984	-2,79734
hsa-miR-500a-3p	0,999981	-1,72374
hsa-miR-501-3p	0,999981	-1,58019
hsa-miR-6503-3p	0,999979	-3,80243
hsa-miR-151a-3p	0,999965	-3,72049
hsa-miR-618	0,999956	2,29224
hsa-miR-143-5p	0,99994	5,00191
hsa-miR-92a-1-5p	0,999921	-1,648
hsa-miR-33b-3p	0,999908	-1,96127
hsa-miR-4750-5p	0,999901	-1,63109
hsa-miR-550a-5p	0,999885	1,71389
hsa-miR-378e	0,999858	-2,63538
hsa-miR-877-5p	0,999841	-2,02128
hsa-miR-4456	0,999815	-4,18335
hsa-miR-6503-5p	0,999804	-2,83011
hsa-miR-451a	0,999797	5,0245
hsa-miR-877-3p	0,999785	-1,80041
hsa-miR-146a-5p	0,999731	-2,75684
hsa-miR-34a-5p	0,999694	-2,11303
hsa-miR-873-5p	0,999668	2,8442
hsa-miR-149-5p	0,999634	-2,77866
hsa-miR-1296-5p	0,999612	-1,78799
hsa-miR-1249-3p	0,999595	-1,92315
hsa-miR-144-5p	0,999594	3,40651
hsa-miR-5010-3p	0,999575	-1,62838

hsa-miR-4676-3p	0,999483	1,49424
hsa-miR-146b-3p	0,999465	-1,27319
hsa-miR-143-3p	0,999449	5,42867
hsa-miR-326	0,999437	-1,64191
hsa-miR-194-3p	0,999365	2,12942
hsa-miR-7847-3p	0,999341	1,59885
hsa-miR-574-5p	0,99928	-1,57602
hsa-miR-378g	0,999265	-1,66855
hsa-miR-148a-5p	0,999204	1,88221
hsa-miR-532-3p	0,999204	-1,63459
hsa-miR-4787-3p	0,999054	-2,06746
hsa-miR-192-5p	0,998801	3,12446
hsa-miR-371a-3p	0,998728	1,40839
hsa-miR-6774-3p	0,998715	-1,82386
hsa-miR-222-5p	0,998546	-2,05722
hsa-miR-106b-3p	0,998501	1,07734
hsa-miR-551a	0,998474	2,19205
hsa-miR-450a-1-3p	0,998463	2,0284
hsa-miR-371b-5p	0,998377	1,32797
hsa-miR-6774-5p	0,998271	-2,01235
hsa-miR-155-5p	0,998066	-2,56136
hsa-miR-3150b-3p	0,99696	2,00365
hsa-miR-378f	0,996861	-1,79408
hsa-miR-181a-3p	0,996843	-1,50093
hsa-miR-556-5p	0,996789	1,99064
hsa-miR-188-5p	0,996757	-3,36824
hsa-miR-6718-5p	0,996712	-1,69105
hsa-miR-4664-3p	0,996681	-1,69501
hsa-miR-15b-5p	0,996671	1,31424
hsa-miR-342-5p	0,996621	-1,49736
hsa-miR-331-5p	0,996216	-2,0092
hsa-miR-940	0,995845	-1,94912
hsa-miR-1298-5p	0,995757	-2,3493
hsa-miR-582-5p	0,995544	2,03739
hsa-miR-3177-3p	0,995133	-0,966111
hsa-miR-378b	0,994873	-2,09822
hsa-miR-6753-3p	0,994781	-1,01932
hsa-miR-1908-5p	0,994629	-2,05675
hsa-miR-345-5p	0,994304	-1,46124
hsa-miR-378d	0,994299	-1,17034
hsa-miR-378c	0,994189	-2,01804
hsa-miR-627-3p	0,993637	1,16852
hsa-miR-505-3p	0,992884	1,42319
hsa-miR-1275	0,992825	-2,6287
hsa-miR-1-3p	0,9918	-3,2313

hsa-miR-486-5p	0,991407	2,0671
hsa-miR-195-5p	0,991238	1,83455
hsa-miR-95-3p	0,991123	-1,78627
hsa-miR-4802-3p	0,990348	-1,52654
hsa-miR-491-5p	0,990109	1,40651
hsa-miR-1180-3p	0,989903	-1,71203
hsa-miR-142-5p	0,989584	2,84423
hsa-miR-885-5p	0,9893	-2,27002
hsa-miR-342-3p	0,987779	-1,39256
hsa-miR-504-5p	0,987643	3,06153
hsa-miR-3143	0,987554	1,28694
hsa-miR-148a-3p	0,987543	3,00269
hsa-miR-4472	0,986464	-1,76139
hsa-miR-502-3p	0,986355	-1,03101
hsa-miR-196b-5p	0,986226	2,08504
hsa-miR-7-1-3p	0,985951	1,20773
hsa-miR-21-3p	0,985563	-1,96177
hsa-miR-223-3p	0,985094	2,78642
hsa-miR-625-5p	0,983282	1,35719
hsa-miR-3615	0,981801	1,29145
hsa-miR-6131	0,980795	1,86927
hsa-miR-628-3p	0,980712	1,11953
hsa-miR-365b-5p	0,976178	-1,37272
hsa-miR-1254	0,975227	-0,875759
hsa-miR-532-5p	0,975061	-1,59441
hsa-miR-1294	0,973745	1,62756
hsa-miR-548w	0,973687	1,0591
hsa-miR-26b-3p	0,973125	1,15708
hsa-miR-3688-3p	0,97181	0,867297
hsa-miR-1284	0,97071	1,06029
hsa-miR-1226-5p	0,969979	-1,57999
hsa-miR-4483	0,969326	-3,17087
hsa-miR-191-3p	0,968993	-1,0813
hsa-miR-6842-5p	0,965582	1,25015
hsa-miR-362-5p	0,963587	-1,47938
hsa-miR-335-5p	0,959347	1,78588
hsa-miR-199b-5p	0,958381	2,28427
hsa-miR-450b-5p	0,955791	1,85096
hsa-miR-6073	0,955122	2,05675
hsa-miR-152-3p	0,954704	-1,43284
hsa-miR-23b-3p	0,954088	1,02459
hsa-miR-378a-5p	0,953821	-2,18254
hsa-miR-6502-5p	0,953046	0,911459
hsa-miR-7850-5p	0,944822	1,65731
hsa-miR-3611	0,942937	1,3156

hsa-miR-378a-3p	0,942732	-2,13838
hsa-miR-3944-3p	0,936895	-1,22524
hsa-miR-454-3p	0,936229	1,48561
hsa-miR-365a-5p	0,935618	-1,70494
hsa-miR-146a-3p	0,928833	-1,90006
hsa-miR-320a	0,923137	-1,03064
hsa-miR-1908-3p	0,921291	-1,05014
hsa-miR-339-3p	0,920131	-1,22139
hsa-miR-23a-5p	0,918805	-0,978441
hsa-miR-450a-5p	0,915388	1,62096
hsa-miR-335-3p	0,914622	-1,43503
hsa-miR-378i	0,913847	-1,5023
hsa-miR-4455	0,909691	-2,08062
hsa-miR-628-5p	0,904739	1,54502
hsa-miR-671-3p	0,904591	-0,650251
hsa-miR-4773	0,903592	1,55526
hsa-miR-933	0,903516	1,25402
hsa-miR-556-3p	0,902242	1,51442



## ANEXO 9: miRNAs diferencialmente expressos em monócitos

feature_id	posterior_probability	eta1_3
hsa-miR-4787-3p	0,999926	-4,54034
hsa-miR-511-5p	0,999875	-4,94072
hsa-miR-34a-5p	0,999632	-3,81704
hsa-miR-1271-5p	0,999516	-1,95024
hsa-miR-221-3p	0,999339	-2,97576
hsa-miR-365b-5p	0,999242	-2,68945
hsa-miR-155-5p	0,999185	-2,59099
hsa-miR-342-3p	0,998695	-3,05044
hsa-miR-146a-5p	0,998258	-3,77122
hsa-miR-146b-3p	0,997453	-1,53207
hsa-miR-4455	0,996921	-5,60034
hsa-miR-6820-3p	0,996865	1,50396
hsa-miR-222-5p	0,995798	-2,81917
hsa-miR-99b-5p	0,993587	-3,15142
hsa-let-7e-5p	0,992233	-3,02594
hsa-miR-125a-5p	0,991604	-3,33966
hsa-miR-4483	0,991595	-3,26529
hsa-miR-132-3p	0,988763	-2,31522
hsa-miR-181a-2-3p	0,982834	1,3444
hsa-miR-1301-3p	0,979649	-1,45791
hsa-miR-7977	0,972141	1,49812
hsa-miR-486-5p	0,958976	1,7436
hsa-miR-203a-3p	0,949692	-1,58952
hsa-miR-4456	0,94457	-2,85471
hsa-miR-6503-3p	0,939764	2,2388
hsa-miR-618	0,939539	1,58361
hsa-miR-6515-5p	0,937077	-1,20733
hsa-let-7e-3p	0,932565	-2,04237
hsa-miR-6732-3p	0,924452	-1,43709
hsa-miR-4288	0,924369	-2,34375
hsa-miR-582-5p	0,923872	1,73575
hsa-miR-616-5p	0,923752	-1,42287
hsa-miR-193a-5p	0,911932	-1,48798
hsa-miR-3615	0,91179	1,14948
hsa-miR-6503-5p	0,908345	1,76488
hsa-miR-150-5p	0,904456	2,67338
hsa-miR-4742-3p	0,900071	-1,15883

## ANEXO 10: miRNAs diferencialmente expressos em mais de uma comparação

miRNA	my,ly	gmp,mep	mono,neut	macro,mono	m2m,m1m
hsa-let-7c-5p	1,47695	1,5509	NA	NA	NA
hsa-let-7e-5p	NA	2,4716	NA	-3,07068	NA
hsa-miR-106a-5p	2,43105	2,26216	NA	NA	NA
hsa-miR-1249-3p	NA	-3,21177	-2,23282	NA	NA
hsa-miR-125b-2-3p	2,34824	3,38882	NA	NA	NA
hsa-miR-126-3p	-4,14748	4,51256	NA	NA	NA
hsa-miR-126-5p	-4,38196	5,25256	NA	NA	NA
hsa-miR-1271-5p	2,47495	NA	-2,18868	-1,84285	NA
hsa-miR-1275	3,22063	3,38417	-2,87508	NA	NA
hsa-miR-1287-5p	1,13841	1,05317	NA	NA	NA
hsa-miR-1301-3p	NA	NA	-1,00478	-1,41116	NA
hsa-miR-142-5p	2,8743	NA	2,19098	NA	NA
hsa-miR-143-3p	-2,46101	-2,38723	5,45092	NA	NA
hsa-miR-145-3p	-2,84314	-1,88522	5,24085	NA	NA
hsa-miR-145-5p	-3,08081	NA	4,23374	NA	NA
hsa-miR-146a-5p	2,03438	NA	-2,72237	-3,76135	NA
hsa-miR-146b-3p	2,59804	1,77364	-1,83693	-2,5081	NA
hsa-miR-146b-5p	2,4979	2,16859	NA	-2,46632	NA
hsa-miR-150-5p	7,65476	NA	NA	2,9326	NA
hsa-miR-151a-3p	NA	3,13316	-3,82656	NA	NA
hsa-miR-155-5p	1,79913	NA	-2,93903	-2,85343	NA
hsa-miR-181a-2-3p	1,85456	NA	NA	1,44001	NA
hsa-miR-181b-5p	NA	1,70067	-0,656324	NA	1,17767
hsa-miR-181c-3p	NA	NA	1,15151	NA	1,83785
hsa-miR-181d-5p	NA	NA	1,24988	NA	1,67951
hsa-miR-186-5p	1,30809	NA	1,07054	NA	NA
hsa-miR-191-3p	-1,0542	NA	-1,06701	NA	NA
hsa-miR-194-5p	2,00976	NA	3,40213	NA	NA
hsa-miR-199b-5p	-2,83551	NA	2,0944	1,98412	NA
hsa-miR-20b-5p	NA	2,29531	NA	1,82041	NA
hsa-miR-2115-3p	NA	NA	3,79057	1,69095	NA
hsa-miR-2116-3p	NA	NA	-3,54729	-2,21312	NA
hsa-miR-221-3p	NA	NA	-1,17237	-2,91403	NA
hsa-miR-222-5p	NA	NA	-2,01268	-3,1424	NA
hsa-miR-223-3p	-3,95216	-1,82045	2,8353	1,60559	NA
hsa-miR-223-5p	-3,26106	NA	1,40639	NA	NA
hsa-miR-25-3p	NA	NA	1,60179	1,00558	NA
hsa-miR-26a-5p	1,41249	NA	1,51637	NA	NA
hsa-miR-26b-3p	NA	-1,33514	1,20286	1,40641	NA
hsa-miR-296-3p	1,59213	3,81961	NA	NA	NA
hsa-miR-3191-5p	NA	NA	-1,62124	-1,64988	NA

hsa-miR-3200-3p	NA	NA	-2,54652	-5,13263	NA
hsa-miR-331-5p	NA	1,64966	-1,59863	NA	NA
hsa-miR-335-3p	1,55392	5,31701	-1,52947	NA	NA
hsa-miR-335-5p	NA	NA	2,02232	1,42525	NA
hsa-miR-338-5p	-3,35555	-3,30959	2,42278	NA	NA
hsa-miR-342-3p	3,15822	NA	-1,87832	-3,23522	NA
hsa-miR-342-5p	4,77465	NA	-2,0035	-2,2253	NA
hsa-miR-345-5p	-3,33512	NA	-1,05252	NA	NA
hsa-miR-34a-5p	NA	NA	-1,89641	-4,96429	NA
hsa-miR-3605-3p	NA	-2,61615	1,16288	NA	NA
hsa-miR-361-3p	2,38895	-1,48327	NA	NA	NA
hsa-miR-3614-5p	NA	-2,71228	1,39833	NA	-2,32547
hsa-miR-3615	NA	-1,33297	1,33117	1,2439	NA
hsa-miR-363-3p	2,53812	1,78844	NA	1,48108	NA
hsa-miR-365b-5p	NA	NA	-2,20016	-2,53634	NA
hsa-miR-378f	NA	NA	-2,97017	-1,01935	NA
hsa-miR-378i	NA	NA	-2,57791	-1,27538	NA
hsa-miR-454-5p	-1,06442	NA	0,950225	0,84663	NA
hsa-miR-4665-5p	NA	NA	-1,73016	-1,41682	NA
hsa-miR-4709-5p	NA	NA	-2,38817	-2,24344	NA
hsa-miR-4787-3p	NA	NA	-2,91299	-4,4601	NA
hsa-miR-4796-3p	1,43515	NA	1,20091	NA	NA
hsa-miR-500a-3p	-1,41091	NA	-1,66476	NA	NA
hsa-miR-511-3p	NA	NA	NA	-5,19959	-2,31732
hsa-miR-511-5p	NA	NA	-5,13841	-5,26241	-2,7996
hsa-miR-550a-5p	-2,60591	NA	1,67114	NA	NA
hsa-miR-574-5p	-3,94024	NA	-2,10099	NA	NA
hsa-miR-582-3p	NA	-4,22534	1,91664	2,53624	NA
hsa-miR-582-5p	NA	-4,19762	2,29146	2,03125	NA
hsa-miR-618	-3218	-4,27338	2,29076	1,84496	NA
hsa-miR-628-5p	NA	-1,43053	1,98998	NA	NA
hsa-miR-6503-3p	NA	NA	-4,54022	3,12905	NA
hsa-miR-6503-5p	NA	NA	-4,14899	2,08408	NA
hsa-miR-6753-3p	NA	NA	-1,84955	-1,89598	NA
hsa-miR-6761-3p	-0,727341	NA	NA	NA	-0,219293
hsa-miR-6876-3p	NA	NA	0,420586	NA	-1,01248
hsa-miR-7155-3p	NA	NA	1,97304	NA	0,138139
hsa-miR-874-3p	3,80877	NA	-1,7744	-1,39925	NA
hsa-miR-95-3p	5,2255	NA	-2,02287	NA	NA
hsa-miR-99b-3p	NA	2,32244	NA	-2,52836	2,37089
hsa-miR-99b-5p	NA	1,89023	NA	-3,10018	NA

## ANEXO 11. Vias metabólicas alteradas em linfócitos

go_id	go_name	p_value
GO:0045766	positive regulation of angiogenesis	2,00E-05
GO:1903034	regulation of response to wounding	4,00E-05
GO:0070821	tertiary granule membrane	6,00E-05
GO:0031092	platelet alpha granule membrane	7,00E-05
GO:0061041	regulation of wound healing	7,00E-05
GO:0050731	positive regulation of peptidyl-tyrosine phosphorylation	8,00E-05
GO:0016722	oxidoreductase activity, oxidizing metal ions	9,00E-05
GO:0040001	establishment of mitotic spindle localization	1,80E-04
GO:0030246	carbohydrate binding	2,00E-04
GO:1904018	positive regulation of vasculature development	2,10E-04
GO:0051783	regulation of nuclear division	2,40E-04
GO:0014002	astrocyte development	2,60E-04
GO:0007599	hemostasis	3,40E-04
GO:1903036	positive regulation of response to wounding	3,90E-04
GO:1900046	regulation of hemostasis	4,20E-04
GO:0050730	regulation of peptidyl-tyrosine phosphorylation	4,50E-04
GO:0002237	response to molecule of bacterial origin	5,00E-04
GO:0032496	response to lipopolysaccharide	5,70E-04
GO:0007088	regulation of mitotic nuclear division	5,80E-04
GO:0030193	regulation of blood coagulation	5,80E-04
GO:0050727	regulation of inflammatory response	5,90E-04
GO:0050853	B cell receptor signaling pathway	7,50E-04
GO:0030823	regulation of cGMP metabolic process	7,70E-04
GO:0001819	positive regulation of cytokine production	8,20E-04
GO:0032675	regulation of interleukin-6 production	8,30E-04
GO:0031349	positive regulation of defense response	8,40E-04
GO:0001525	angiogenesis	8,80E-04
GO:0006954	inflammatory response	9,20E-04
GO:0090312	positive regulation of protein deacetylation	9,30E-04
GO:0050818	regulation of coagulation	9,50E-04
GO:0031065	positive regulation of histone deacetylation	9,90E-04
GO:0031406	carboxylic acid binding	1,00E-03
GO:0090303	positive regulation of wound healing	1,04E-03
GO:0050878	regulation of body fluid levels	1,06E-03
GO:0050817	coagulation	1,11E-03
GO:0048729	tissue morphogenesis	1,17E-03
GO:0007596	blood coagulation	1,19E-03
GO:0046683	response to organophosphorus	1,22E-03
GO:0045058	T cell selection	1,24E-03
GO:0046456	icosanoid biosynthetic process	1,24E-03
GO:0050729	positive regulation of inflammatory response	1,28E-03
GO:0043395	heparan sulfate proteoglycan binding	1,31E-03
GO:0019369	arachidonic acid metabolic process	1,32E-03
GO:0042108	positive regulation of cytokine biosynthetic process	1,35E-03
GO:0006636	unsaturated fatty acid biosynthetic process	1,39E-03
GO:1901342	regulation of vasculature development	1,45E-03
GO:0050840	extracellular matrix binding	1,47E-03
GO:0000079	regulation of cyclin-dependent protein serine/threonine kinase activity	1,56E-03
GO:0051294	establishment of spindle orientation	1,57E-03
GO:0000132	establishment of mitotic spindle orientation	1,58E-03
GO:0042303	molting cycle	1,59E-03
GO:0009408	response to heat	1,60E-03
GO:1904029	regulation of cyclin-dependent protein kinase activity	1,63E-03
GO:0019372	lipxygenase pathway	1,64E-03
GO:0042633	hair cycle	1,85E-03
GO:0051928	positive regulation of calcium ion transport	1,86E-03
GO:0045444	fat cell differentiation	2,09E-03
GO:0045987	positive regulation of smooth muscle contraction	2,33E-03
GO:0021782	glial cell development	2,39E-03
GO:0006968	cellular defense response	2,43E-03
GO:0034121	regulation of toll-like receptor signaling pathway	2,43E-03
GO:0034122	negative regulation of toll-like receptor signaling pathway	2,44E-03
GO:0043409	negative regulation of MAPK cascade	2,82E-03
GO:0034764	positive regulation of transmembrane transport	2,84E-03
GO:0043177	organic acid binding	2,90E-03
GO:0002705	positive regulation of leukocyte mediated immunity	2,91E-03
GO:0032103	positive regulation of response to external stimulus	3,05E-03
GO:0045840	positive regulation of mitotic nuclear division	3,05E-03
GO:0030218	erythrocyte differentiation	3,15E-03
GO:0032102	negative regulation of response to external stimulus	3,18E-03
GO:0009743	response to carbohydrate	3,20E-03
GO:0051785	positive regulation of nuclear division	3,20E-03
GO:0035587	purinergic receptor signaling pathway	3,28E-03
GO:0035588	G-protein coupled purinergic receptor signaling pathway	3,32E-03
GO:0070372	regulation of ERK1 and ERK2 cascade	3,33E-03
GO:0034612	response to tumor necrosis factor	3,35E-03
GO:0045765	regulation of angiogenesis	3,40E-03
GO:0010574	regulation of vascular endothelial growth factor production	3,42E-03
GO:0002009	morphogenesis of an epithelium	3,58E-03
GO:0035586	purinergic receptor activity	3,65E-03
GO:0038024	cargo receptor activity	3,67E-03
GO:0006910	phagocytosis, recognition	3,93E-03
GO:0002092	positive regulation of receptor internalization	3,97E-03
GO:0034605	cellular response to heat	4,02E-03
GO:0031100	animal organ regeneration	4,07E-03
GO:0051653	spindle localization	4,08E-03
GO:0009266	response to temperature stimulus	4,13E-03
GO:2000377	regulation of reactive oxygen species metabolic process	4,30E-03
GO:0000775	chromosome, centromeric region	4,35E-03

GO:0071900	regulation of protein serine/threonine kinase activity	4,42E-03
GO:0045124	regulation of bone resorption	4,61E-03
GO:0008329	signaling pattern recognition receptor activity	4,65E-03
GO:0021952	central nervous system projection neuron axonogenesis	4,67E-03
GO:0038187	pattern recognition receptor activity	4,76E-03
GO:0046850	regulation of bone remodeling	4,80E-03
GO:0032225	regulation of synaptic transmission, dopaminergic	4,95E-03
GO:0040036	regulation of fibroblast growth factor receptor signaling pathway	4,96E-03
GO:0048260	positive regulation of receptor-mediated endocytosis	5,27E-03
GO:0046427	positive regulation of JAK-STAT cascade	5,40E-03
GO:0019842	vitamin binding	5,51E-03
GO:1901137	carbohydrate derivative biosynthetic process	5,54E-03
GO:0045931	positive regulation of mitotic cell cycle	5,65E-03
GO:0007346	regulation of mitotic cell cycle	5,68E-03
GO:0002822	regulation of adaptive immune response based on somatic recombination of immune receptors built from immunoglobulin superfamily domains	5,82E-03
GO:0045933	positive regulation of muscle contraction	6,00E-03
GO:0070373	negative regulation of ERK1 and ERK2 cascade	6,00E-03
GO:0010575	positive regulation of vascular endothelial growth factor production	6,03E-03
GO:2000144	positive regulation of DNA-templated transcription, initiation	6,03E-03
GO:0019825	oxygen binding	6,05E-03
GO:0010042	response to manganese ion	6,29E-03
GO:0050873	brown fat cell differentiation	6,33E-03
GO:1904894	positive regulation of STAT cascade	6,33E-03
GO:0030318	melanocyte differentiation	6,36E-03
GO:0030947	regulation of vascular endothelial growth factor receptor signaling pathway	6,50E-03
GO:0051293	establishment of spindle localization	6,50E-03
GO:0050931	pigment cell differentiation	6,60E-03
GO:0048514	blood vessel morphogenesis	6,67E-03
GO:0009165	nucleotide biosynthetic process	6,75E-03
GO:0030949	positive regulation of vascular endothelial growth factor receptor signaling pathway	6,87E-03
GO:0050909	sensory perception of taste	6,89E-03
GO:0070665	positive regulation of leukocyte proliferation	7,00E-03
GO:0033044	regulation of chromosome organization	7,16E-03
GO:1901293	nucleoside phosphate biosynthetic process	7,22E-03
GO:0034097	response to cytokine	7,25E-03
GO:0046906	tetrapyrrole binding	7,38E-03
GO:0097305	response to alcohol	7,41E-03
GO:0006937	regulation of muscle contraction	7,47E-03
GO:0044772	mitotic cell cycle phase transition	7,76E-03
GO:1903524	positive regulation of blood circulation	7,76E-03
GO:0042035	regulation of cytokine biosynthetic process	7,88E-03
GO:0005539	glycosaminoglycan binding	8,14E-03
GO:0030851	granulocyte differentiation	8,20E-03
GO:0034284	response to monosaccharide	8,21E-03
GO:0044770	cell cycle phase transition	8,27E-03
GO:1901990	regulation of mitotic cell cycle phase transition	8,40E-03
GO:0046879	hormone secretion	8,48E-03
GO:0000188	inactivation of MAPK activity	8,82E-03
GO:0045117	azole transport	8,89E-03
GO:0043270	positive regulation of ion transport	8,99E-03
GO:0051861	glycolipid binding	9,00E-03
GO:0034502	protein localization to chromosome	9,15E-03
GO:0070613	regulation of protein processing	9,44E-03
GO:0050857	positive regulation of antigen receptor-mediated signaling pathway	9,47E-03
GO:0032645	regulation of granulocyte macrophage colony-stimulating factor production	9,52E-03
GO:0034330	cell junction organization	9,52E-03
GO:0032303	regulation of icosanoid secretion	9,53E-03
GO:0071901	negative regulation of protein serine/threonine kinase activity	9,58E-03
GO:1903317	regulation of protein maturation	9,58E-03
GO:0050671	positive regulation of lymphocyte proliferation	9,73E-03
GO:0002040	sprouting angiogenesis	9,84E-03
GO:0008209	androgen metabolic process	9,86E-03
GO:0071356	cellular response to tumor necrosis factor	9,97E-03

## ANEXO 12 Vias metabólicas alteradas em megacariócitos e eritroblatos

go_id	go_name	p_value
GO:0001101	response to acid chemical	2,00E-05
GO:0032675	regulation of interleukin-6 production	4,00E-05
GO:0035150	regulation of tube size	4,00E-05
GO:0097746	regulation of blood vessel diameter	4,00E-05
GO:0003018	vascular process in circulatory system	6,00E-05
GO:0050880	regulation of blood vessel size	6,00E-05
GO:0035296	regulation of tube diameter	7,00E-05
GO:0071396	cellular response to lipid	7,00E-05
GO:0001819	positive regulation of cytokine production	1,10E-04
GO:0003013	circulatory system process	1,10E-04
GO:0002237	response to molecule of bacterial origin	1,30E-04
GO:0006954	inflammatory response	1,50E-04
GO:0032496	response to lipopolysaccharide	1,70E-04
GO:1902532	negative regulation of intracellular signal transduction	1,70E-04
GO:0019229	regulation of vasoconstriction	2,70E-04
GO:0002683	negative regulation of immune system process	3,00E-04
GO:0098542	defense response to other organism	3,60E-04
GO:1903524	positive regulation of blood circulation	3,80E-04
GO:0006805	xenobiotic metabolic process	5,80E-04
GO:0019221	cytokine-mediated signaling pathway	6,10E-04
GO:0045671	negative regulation of osteoclast differentiation	7,00E-04
GO:0048660	regulation of smooth muscle cell proliferation	7,70E-04
GO:0043062	extracellular structure organization	7,90E-04
GO:1903556	negative regulation of tumor necrosis factor superfamily cytokine production	8,30E-04
GO:0030198	extracellular matrix organization	8,40E-04
GO:0032755	positive regulation of interleukin-6 production	8,40E-04
GO:0002819	regulation of adaptive immune response	8,80E-04
GO:0022602	ovulation cycle process	8,80E-04
GO:1903532	positive regulation of secretion by cell	8,90E-04
GO:0032720	negative regulation of tumor necrosis factor production	9,50E-04
GO:0022600	digestive system process	1,08E-03
GO:0031012	extracellular matrix	1,08E-03
GO:1903706	regulation of hemopoiesis	1,10E-03
GO:0050777	negative regulation of immune response	1,11E-03
GO:1903522	regulation of blood circulation	1,19E-03
GO:1901800	positive regulation of proteasomal protein catabolic process	1,31E-03
GO:0051262	protein tetramerization	1,36E-03
GO:0002695	negative regulation of leukocyte activation	1,37E-03
GO:0019369	arachidonic acid metabolic process	1,42E-03
GO:0032436	positive regulation of proteasomal ubiquitin-dependent protein catabolic process	1,42E-03
GO:0043114	regulation of vascular permeability	1,47E-03
GO:0051899	membrane depolarization	1,50E-03
GO:0051250	negative regulation of lymphocyte activation	1,51E-03
GO:1904407	positive regulation of nitric oxide metabolic process	1,51E-03
GO:0007200	phospholipase C-activating G-protein coupled receptor signaling pathway	1,55E-03
GO:0071219	cellular response to molecule of bacterial origin	1,56E-03
GO:0002822	regulation of adaptive immune response based on somatic recombination of immune receptors built from immunoglobulin superfamily domains	1,67E-03
GO:0045429	positive regulation of nitric oxide biosynthetic process	1,67E-03
GO:0032680	regulation of tumor necrosis factor production	1,68E-03
GO:0031099	regeneration	1,69E-03
GO:0071222	cellular response to lipopolysaccharide	1,69E-03
GO:1903555	regulation of tumor necrosis factor superfamily cytokine production	1,70E-03
GO:0045840	positive regulation of mitotic nuclear division	1,77E-03
GO:0051785	positive regulation of nuclear division	1,80E-03
GO:0002793	positive regulation of peptide secretion	1,86E-03
GO:0050714	positive regulation of protein secretion	1,87E-03
GO:1903428	positive regulation of reactive oxygen species biosynthetic process	1,92E-03
GO:0030595	leukocyte chemotaxis	2,02E-03
GO:1903364	positive regulation of cellular protein catabolic process	2,07E-03
GO:0032526	response to retinoic acid	2,37E-03
GO:0002791	regulation of peptide secretion	2,40E-03
GO:0007263	nitric oxide mediated signal transduction	2,41E-03
GO:0005179	hormone activity	2,46E-03
GO:0012507	ER to Golgi transport vesicle membrane	2,47E-03
GO:0008227	G-protein coupled amine receptor activity	2,53E-03
GO:0010817	regulation of hormone levels	2,69E-03
GO:0005791	rough endoplasmic reticulum	2,73E-03
GO:0042035	regulation of cytokine biosynthetic process	2,73E-03
GO:0050708	regulation of protein secretion	2,88E-03
GO:0051781	positive regulation of cell division	2,88E-03
GO:0002283	neutrophil activation involved in immune response	2,94E-03
GO:0005507	copper ion binding	2,96E-03
GO:0060326	cell chemotaxis	2,98E-03
GO:0043312	neutrophil degranulation	3,00E-03
GO:0009617	response to bacterium	3,02E-03
GO:0019842	vitamin binding	3,08E-03
GO:0032103	positive regulation of response to external stimulus	3,09E-03
GO:0043122	regulation of I-kappaB kinase/NF-kappaB signaling	3,13E-03
GO:0051101	regulation of DNA binding	3,15E-03
GO:0002691	regulation of cellular extravasation	3,17E-03
GO:0006935	chemotaxis	3,23E-03
GO:0043433	negative regulation of sequence-specific DNA binding transcription factor activity	3,24E-03

GO:1902105	regulation of leukocyte differentiation	3,28E-03
GO:0042330	taxis	3,31E-03
GO:0050864	regulation of B cell activation	3,32E-03
GO:0042277	peptide binding	3,52E-03
GO:0097529	myeloid leukocyte migration	3,55E-03
GO:0042108	positive regulation of cytokine biosynthetic process	3,56E-03
GO:0045823	positive regulation of heart contraction	3,63E-03
GO:0002687	positive regulation of leukocyte migration	3,74E-03
GO:0046649	lymphocyte activation	3,74E-03
GO:0071229	cellular response to acid chemical	3,79E-03
GO:0002286	T cell activation involved in immune response	3,81E-03
GO:0016829	lyase activity	3,89E-03
GO:0002521	leukocyte differentiation	4,03E-03
GO:0036230	granulocyte activation	4,05E-03
GO:0042119	neutrophil activation	4,07E-03
GO:0042534	regulation of tumor necrosis factor biosynthetic process	4,18E-03
GO:0043299	leukocyte degranulation	4,19E-03
GO:0032715	negative regulation of interleukin-6 production	4,46E-03
GO:0002027	regulation of heart rate	4,50E-03
GO:0031100	animal organ regeneration	4,50E-03
GO:0045428	regulation of nitric oxide biosynthetic process	4,63E-03
GO:0050866	negative regulation of cell activation	4,67E-03
GO:0042742	defense response to bacterium	5,60E-03
GO:0086010	membrane depolarization during action potential	5,78E-03
GO:1903038	negative regulation of leukocyte cell-cell adhesion	5,81E-03
GO:0030660	Golgi-associated vesicle membrane	5,88E-03
GO:0050868	negative regulation of T cell activation	5,91E-03
GO:0043271	negative regulation of ion transport	6,07E-03
GO:0031901	early endosome membrane	6,12E-03
GO:0051090	regulation of sequence-specific DNA binding transcription factor activity	6,14E-03
GO:1902475	L-alpha-amino acid transmembrane transport	6,17E-03
GO:0002685	regulation of leukocyte migration	6,30E-03
GO:0002250	adaptive immune response	6,58E-03
GO:0060348	bone development	6,68E-03
GO:0032727	positive regulation of interferon-alpha production	6,74E-03
GO:0031341	regulation of cell killing	6,77E-03
GO:0032649	regulation of interferon-gamma production	6,82E-03
GO:0043200	response to amino acid	6,82E-03
GO:0002703	regulation of leukocyte mediated immunity	7,01E-03
GO:0050707	regulation of cytokine secretion	7,06E-03
GO:0045637	regulation of myeloid cell differentiation	7,10E-03
GO:0055038	recycling endosome membrane	7,15E-03
GO:0032660	regulation of interleukin-17 production	7,18E-03
GO:0030277	maintenance of gastrointestinal epithelium	7,33E-03
GO:0030225	macrophage differentiation	7,43E-03
GO:0098581	detection of external biotic stimulus	7,46E-03
GO:0050729	positive regulation of inflammatory response	7,47E-03
GO:0055072	iron ion homeostasis	7,55E-03
GO:0009595	detection of biotic stimulus	7,56E-03
GO:0010669	epithelial structure maintenance	7,79E-03
GO:0032652	regulation of interleukin-1 production	7,85E-03
GO:0060333	interferon-gamma-mediated signaling pathway	7,92E-03
GO:0002790	peptide secretion	7,95E-03
GO:0022617	extracellular matrix disassembly	8,07E-03
GO:0002526	acute inflammatory response	8,35E-03
GO:0002285	lymphocyte activation involved in immune response	8,48E-03
GO:0002700	regulation of production of molecular mediator of immune response	8,74E-03
GO:0051289	protein homotetramerization	8,76E-03
GO:0034113	heterotypic cell-cell adhesion	8,86E-03
GO:1903707	negative regulation of hemopoiesis	8,88E-03
GO:0002577	regulation of antigen processing and presentation	8,89E-03
GO:0044390	ubiquitin-like protein conjugating enzyme binding	8,91E-03
GO:0071216	cellular response to biotic stimulus	8,95E-03
GO:0097060	synaptic membrane	9,23E-03
GO:0031624	ubiquitin conjugating enzyme binding	9,26E-03
GO:0043268	positive regulation of potassium ion transport	9,29E-03
GO:0010965	regulation of mitotic sister chromatid separation	9,30E-03
GO:0090075	relaxation of muscle	9,33E-03
GO:0005242	inward rectifier potassium channel activity	9,34E-03
GO:1903115	regulation of actin filament-based movement	9,35E-03
GO:0098900	regulation of action potential	9,40E-03
GO:0033137	negative regulation of peptidyl-serine phosphorylation	9,55E-03
GO:0086009	membrane repolarization	9,61E-03
GO:0030315	T-tubule	9,62E-03
GO:0086004	regulation of cardiac muscle cell contraction	9,66E-03
GO:0086091	regulation of heart rate by cardiac conduction	9,68E-03
GO:1901381	positive regulation of potassium ion transmembrane transport	9,71E-03
GO:0030111	regulation of Wnt signaling pathway	9,78E-03
GO:0005790	smooth endoplasmic reticulum	9,83E-03
GO:0042611	MHC protein complex	9,87E-03
GO:0086011	membrane repolarization during action potential	9,87E-03
GO:0061337	cardiac conduction	9,88E-03
GO:0071674	mononuclear cell migration	9,88E-03
GO:0086013	membrane repolarization during cardiac muscle cell action potential	9,90E-03
GO:1902099	regulation of metaphase/anaphase transition of cell cycle	9,90E-03
GO:0001508	action potential	9,94E-03
GO:1905818	regulation of chromosome separation	9,98E-03
GO:0048029	monosaccharide binding	9,99E-03
GO:0086001	cardiac muscle cell action potential	9,99E-03
GO:0099622	cardiac muscle cell membrane repolarization	9,99E-03
GO:0055075	potassium ion homeostasis	1,00E-02

## ANEXO 13. Vias metabólicas alteradas em Neutrófilos

go_id	go_name	p_value
GO:0010744	positive regulation of macrophage derived foam cell differentiation	1,00E-05
GO:0050678	regulation of epithelial cell proliferation	2,00E-05
GO:0051043	regulation of membrane protein ectodomain proteolysis	2,00E-05
GO:0070372	regulation of ERK1 and ERK2 cascade	3,00E-05
GO:0030546	receptor activator activity	6,00E-05
GO:0098794	postsynapse	6,00E-05
GO:0030545	receptor regulator activity	7,00E-05
GO:0045637	regulation of myeloid cell differentiation	9,00E-05
GO:0034341	response to interferon-gamma	1,30E-04
GO:0035587	purinergic receptor signaling pathway	1,40E-04
GO:0032103	positive regulation of response to external stimulus	1,50E-04
GO:0071526	semaphorin-plexin signaling pathway	1,60E-04
GO:0030032	lamellipodium assembly	1,70E-04
GO:0032506	cytokinetic process	1,70E-04
GO:0030198	extracellular matrix organization	1,80E-04
GO:0043062	extracellular structure organization	2,00E-04
GO:0001772	immunological synapse	2,20E-04
GO:0043268	positive regulation of potassium ion transport	2,40E-04
GO:0031406	carboxylic acid binding	2,60E-04
GO:0034097	response to cytokine	2,60E-04
GO:0034612	response to tumor necrosis factor	2,60E-04
GO:0001936	regulation of endothelial cell proliferation	3,00E-04
GO:0031343	positive regulation of cell killing	3,10E-04
GO:0043197	dendritic spine	3,20E-04
GO:0002761	regulation of myeloid leukocyte differentiation	3,30E-04
GO:1990138	neuron projection extension	3,30E-04
GO:0043177	organic acid binding	3,50E-04
GO:0001912	positive regulation of leukocyte mediated cytotoxicity	3,90E-04
GO:0008360	regulation of cell shape	4,20E-04
GO:0016597	amino acid binding	4,20E-04
GO:0050679	positive regulation of epithelial cell proliferation	5,00E-04
GO:0070664	negative regulation of leukocyte proliferation	5,10E-04
GO:0007405	neuroblast proliferation	5,20E-04
GO:0070374	positive regulation of ERK1 and ERK2 cascade	5,30E-04
GO:0048018	receptor agonist activity	5,60E-04
GO:0044309	neuron spine	5,80E-04
GO:1901381	positive regulation of potassium ion transmembrane transport	5,80E-04
GO:0001523	retinoid metabolic process	6,30E-04
GO:0033198	response to ATP	6,80E-04
GO:0043410	positive regulation of MAPK cascade	7,10E-04
GO:0035586	purinergic receptor activity	7,30E-04
GO:0035590	purinergic nucleotide receptor signaling pathway	7,30E-04
GO:0033293	monocarboxylic acid binding	7,70E-04
GO:1990777	lipoprotein particle	7,80E-04
GO:0048675	axon extension	7,90E-04
GO:0052689	carboxylic ester hydrolase activity	8,30E-04
GO:0034358	plasma lipoprotein particle	9,00E-04
GO:0072676	lymphocyte migration	9,00E-04
GO:0002831	regulation of response to biotic stimulus	9,20E-04
GO:0048247	lymphocyte chemotaxis	9,20E-04
GO:0071356	cellular response to tumor necrosis factor	9,50E-04
GO:0043266	regulation of potassium ion transport	1,03E-03
GO:1990266	neutrophil migration	1,05E-03
GO:0042130	negative regulation of T cell proliferation	1,10E-03
GO:0030593	neutrophil chemotaxis	1,16E-03
GO:1901379	regulation of potassium ion transmembrane transport	1,17E-03
GO:0002833	positive regulation of response to biotic stimulus	1,19E-03
GO:0015079	potassium ion transmembrane transporter activity	1,23E-03
GO:0048020	CCR chemokine receptor binding	1,24E-03
GO:0050672	negative regulation of lymphocyte proliferation	1,35E-03
GO:0032945	negative regulation of mononuclear cell proliferation	1,39E-03
GO:0014910	regulation of smooth muscle cell migration	1,48E-03
GO:0071805	potassium ion transmembrane transport	1,48E-03



GO:0006805	xenobiotic metabolic process	1,53E-03
GO:0071804	cellular potassium ion transport	1,56E-03
GO:0001910	regulation of leukocyte mediated cytotoxicity	1,58E-03
GO:0006813	potassium ion transport	1,58E-03
GO:0032720	negative regulation of tumor necrosis factor production	1,66E-03
GO:0071346	cellular response to interferon-gamma	1,66E-03
GO:0048588	developmental cell growth	1,75E-03
GO:0043270	positive regulation of ion transport	1,76E-03
GO:1903556	negative regulation of tumor necrosis factor superfamily cytokine production	1,80E-03
GO:0045765	regulation of angiogenesis	1,83E-03
GO:0016101	diterpenoid metabolic process	1,93E-03
GO:1901342	regulation of vasculature development	1,95E-03
GO:0051044	positive regulation of membrane protein ectodomain proteolysis	2,04E-03
GO:0030684	peribosome	2,06E-03
GO:0042136	neurotransmitter biosynthetic process	2,24E-03
GO:0033273	response to vitamin	2,30E-03
GO:0006721	terpenoid metabolic process	2,31E-03
GO:0032689	negative regulation of interferon-gamma production	2,34E-03
GO:0001614	purinergic nucleotide receptor activity	2,51E-03
GO:0098801	regulation of renal system process	2,54E-03
GO:0032870	cellular response to hormone stimulus	2,72E-03
GO:0006720	isoprenoid metabolic process	2,83E-03
GO:0005581	collagen trimer	2,87E-03
GO:0016502	nucleotide receptor activity	2,87E-03
GO:0099572	postsynaptic specialization	2,90E-03
GO:1901699	cellular response to nitrogen compound	2,98E-03
GO:0097581	lamellipodium organization	3,07E-03
GO:0005604	basement membrane	3,12E-03
GO:0014069	postsynaptic density of dendrite	3,17E-03
GO:0051249	regulation of lymphocyte activation	3,24E-03
GO:0009617	response to bacterium	3,35E-03
GO:1901616	organic hydroxy compound catabolic process	3,40E-03
GO:0006956	complement activation	3,63E-03
GO:0050920	regulation of chemotaxis	3,63E-03
GO:0004033	aldo-keto reductase (NADP) activity	3,65E-03
GO:0043011	myeloid dendritic cell differentiation	3,66E-03
GO:0005504	fatty acid binding	3,74E-03
GO:0097028	dendritic cell differentiation	3,77E-03
GO:0001773	myeloid dendritic cell activation	3,87E-03
GO:0071345	cellular response to cytokine stimulus	3,87E-03
GO:0010830	regulation of myotube differentiation	3,88E-03
GO:0017171	serine hydrolase activity	3,94E-03
GO:0042537	benzene-containing compound metabolic process	4,07E-03
GO:0050868	negative regulation of T cell activation	4,13E-03
GO:0098552	side of membrane	4,19E-03
GO:0010594	regulation of endothelial cell migration	4,22E-03
GO:0031341	regulation of cell killing	4,26E-03
GO:0018410	C-terminal protein amino acid modification	4,40E-03
GO:0002715	regulation of natural killer cell mediated immunity	4,42E-03
GO:0071219	cellular response to molecule of bacterial origin	4,45E-03
GO:1903038	negative regulation of leukocyte cell-cell adhesion	4,53E-03
GO:0071347	cellular response to interleukin-1	4,65E-03
GO:0044126	regulation of growth of symbiont in host	4,77E-03
GO:0015276	ligand-gated ion channel activity	4,81E-03
GO:0032941	secretion by tissue	4,83E-03
GO:0008144	drug binding	4,84E-03
GO:0050863	regulation of T cell activation	4,85E-03
GO:0071222	cellular response to lipopolysaccharide	4,86E-03
GO:0044144	modulation of growth of symbiont involved in interaction with host	4,92E-03
GO:0042269	regulation of natural killer cell mediated cytotoxicity	4,98E-03
GO:0044146	negative regulation of growth of symbiont involved in interaction with host	4,99E-03
GO:0070663	regulation of leukocyte proliferation	5,01E-03
GO:0022834	ligand-gated channel activity	5,02E-03
GO:0006898	receptor-mediated endocytosis	5,08E-03
GO:0007162	negative regulation of cell adhesion	5,20E-03
GO:0045670	regulation of osteoclast differentiation	5,24E-03
GO:0035584	calcium-mediated signaling using intracellular calcium source	5,37E-03
GO:0044243	multicellular organismal catabolic process	5,38E-03
GO:0031225	anchored component of membrane	5,47E-03
GO:0008076	voltage-gated potassium channel complex	5,52E-03
GO:0051209	release of sequestered calcium ion into cytosol	5,55E-03
GO:0072376	protein activation cascade	5,56E-03

GO:0043034	costamere	5,60E-03
GO:0030424	axon	5,67E-03
GO:0022604	regulation of cell morphogenesis	5,68E-03
GO:0034705	potassium channel complex	5,68E-03
GO:1903037	regulation of leukocyte cell-cell adhesion	5,69E-03
GO:0097553	calcium ion transmembrane import into cytosol	5,77E-03
GO:0030170	pyridoxal phosphate binding	5,98E-03
GO:0045471	response to ethanol	6,03E-03
GO:0051283	negative regulation of sequestering of calcium ion	6,06E-03
GO:0070279	vitamin B6 binding	6,08E-03
GO:0050839	cell adhesion molecule binding	6,12E-03
GO:0090075	relaxation of muscle	6,20E-03
GO:0051769	regulation of nitric-oxide synthase biosynthetic process	6,24E-03
GO:0009991	response to extracellular stimulus	6,25E-03
GO:0048593	camera-type eye morphogenesis	6,26E-03
GO:0044283	small molecule biosynthetic process	6,31E-03
GO:0031253	cell projection membrane	6,37E-03
GO:0001937	negative regulation of endothelial cell proliferation	6,43E-03
GO:0019233	sensory perception of pain	6,46E-03
GO:0034284	response to monosaccharide	6,51E-03
GO:0032963	collagen metabolic process	6,53E-03
GO:1901343	negative regulation of vasculature development	6,55E-03
GO:0008201	heparin binding	6,62E-03
GO:2000181	negative regulation of blood vessel morphogenesis	6,63E-03
GO:0016525	negative regulation of angiogenesis	6,66E-03
GO:0043535	regulation of blood vessel endothelial cell migration	6,66E-03
GO:0046464	acylglycerol catabolic process	6,66E-03
GO:0032374	regulation of cholesterol transport	6,76E-03
GO:0043409	negative regulation of MAPK cascade	6,77E-03
GO:0090025	regulation of monocyte chemotaxis	6,82E-03
GO:0008236	serine-type peptidase activity	6,97E-03
GO:0040013	negative regulation of locomotion	6,98E-03
GO:0071216	cellular response to biotic stimulus	6,99E-03
GO:0071333	cellular response to glucose stimulus	6,99E-03
GO:0008395	steroid hydroxylase activity	7,00E-03
GO:0003823	antigen binding	7,01E-03
GO:0032371	regulation of sterol transport	7,03E-03
GO:1903706	regulation of hemopoiesis	7,13E-03
GO:0070373	negative regulation of ERK1 and ERK2 cascade	7,16E-03
GO:0046461	neutral lipid catabolic process	7,17E-03
GO:0044449	contractile fiber part	7,18E-03
GO:0071677	positive regulation of mononuclear cell migration	7,21E-03
GO:0042605	peptide antigen binding	7,23E-03
GO:0071331	cellular response to hexose stimulus	7,30E-03
GO:0040019	positive regulation of embryonic development	7,33E-03
GO:0071322	cellular response to carbohydrate stimulus	7,35E-03
GO:0090026	positive regulation of monocyte chemotaxis	7,35E-03
GO:0006958	complement activation, classical pathway	7,41E-03
GO:0030666	endocytic vesicle membrane	7,43E-03
GO:0045639	positive regulation of myeloid cell differentiation	7,48E-03
GO:0050680	negative regulation of epithelial cell proliferation	7,69E-03
GO:0010959	regulation of metal ion transport	7,74E-03
GO:0007032	endosome organization	7,81E-03
GO:1901654	response to ketone	7,90E-03
GO:0043174	nucleoside salvage	7,97E-03
GO:0033013	tetrapyrrole metabolic process	7,99E-03
GO:0071326	cellular response to monosaccharide stimulus	8,03E-03
GO:1905952	regulation of lipid localization	8,19E-03
GO:0050922	negative regulation of chemotaxis	8,24E-03
GO:0060326	cell chemotaxis	8,24E-03
GO:0050770	regulation of axonogenesis	8,33E-03
GO:0071229	cellular response to acid chemical	8,36E-03
GO:0005231	excitatory extracellular ligand-gated ion channel activity	8,37E-03
GO:0005230	extracellular ligand-gated ion channel activity	8,58E-03
GO:0048873	homeostasis of number of cells within a tissue	8,60E-03
GO:0044420	extracellular matrix component	8,68E-03
GO:0042056	chemoattractant activity	8,85E-03
GO:0042742	defense response to bacterium	8,85E-03
GO:0014074	response to purine-containing compound	8,88E-03
GO:0045926	negative regulation of growth	8,94E-03
GO:0032104	regulation of response to extracellular stimulus	8,98E-03
GO:0032496	response to lipopolysaccharide	9,01E-03

GO:0050900	leukocyte migration	9,30E-03
GO:0010522	regulation of calcium ion transport into cytosol	9,31E-03
GO:0002548	monocyte chemotaxis	9,44E-03
GO:0032107	regulation of response to nutrient levels	9,47E-03
GO:0071396	cellular response to lipid	9,54E-03
GO:1903391	regulation of adherens junction organization	9,58E-03
GO:0098852	lytic vacuole membrane	9,62E-03
GO:0071674	mononuclear cell migration	9,71E-03
GO:0097035	regulation of membrane lipid distribution	9,72E-03
GO:0043325	phosphatidylinositol-3,4-bisphosphate binding	9,77E-03
GO:0097530	granulocyte migration	9,77E-03
GO:0071384	cellular response to corticosteroid stimulus	9,80E-03
GO:0050670	regulation of lymphocyte proliferation	9,81E-03
GO:0032944	regulation of mononuclear cell proliferation	9,85E-03
GO:0032994	protein-lipid complex	9,88E-03

## ANEXO 14. Vias metabólicas alteradas em Monócitos

go_id	go_name	p_value
GO:0051272	positive regulation of cellular component movement	7,00E-05
GO:0032368	regulation of lipid transport	1,20E-04
GO:0055102	lipase inhibitor activity	2,50E-04
GO:0023014	signal transduction by protein phosphorylation	3,10E-04
GO:0005912	adherens junction	4,50E-04
GO:0000165	MAPK cascade	4,60E-04
GO:0019838	growth factor binding	4,90E-04
GO:0005504	fatty acid binding	5,20E-04
GO:2000147	positive regulation of cell motility	5,50E-04
GO:0030335	positive regulation of cell migration	5,80E-04
GO:0005085	guanyl-nucleotide exchange factor activity	7,00E-04
GO:0043062	extracellular structure organization	7,60E-04
GO:0005924	cell-substrate adherens junction	8,00E-04
GO:0046464	acylglycerol catabolic process	8,00E-04
GO:0046461	neutral lipid catabolic process	8,10E-04
GO:0030055	cell-substrate junction	8,20E-04
GO:0019433	triglyceride catabolic process	8,30E-04
GO:0005925	focal adhesion	8,50E-04
GO:0030198	extracellular matrix organization	8,50E-04
GO:0032989	cellular component morphogenesis	9,10E-04
GO:0032374	regulation of cholesterol transport	1,02E-03
GO:0032371	regulation of sterol transport	1,04E-03
GO:0048010	vascular endothelial growth factor receptor signaling pathway	1,06E-03
GO:0040017	positive regulation of locomotion	1,25E-03
GO:0043410	positive regulation of MAPK cascade	1,25E-03
GO:0050873	brown fat cell differentiation	1,31E-03
GO:0010743	regulation of macrophage derived foam cell differentiation	1,33E-03
GO:0005539	glycosaminoglycan binding	1,44E-03
GO:0032103	positive regulation of response to external stimulus	1,44E-03
GO:0045598	regulation of fat cell differentiation	1,47E-03
GO:0034105	positive regulation of tissue remodeling	1,48E-03
GO:0008201	heparin binding	1,53E-03
GO:0052689	carboxylic ester hydrolase activity	1,72E-03
GO:0071356	cellular response to tumor necrosis factor	1,75E-03
GO:0005125	cytokine activity	1,77E-03
GO:0045639	positive regulation of myeloid cell differentiation	1,89E-03
GO:0070372	regulation of ERK1 and ERK2 cascade	2,47E-03
GO:0022617	extracellular matrix disassembly	2,48E-03
GO:0014068	positive regulation of phosphatidylinositol 3-kinase signaling	2,62E-03
GO:0006936	muscle contraction	2,76E-03
GO:0035924	cellular response to vascular endothelial growth factor stimulus	2,79E-03
GO:0031333	negative regulation of protein complex assembly	2,86E-03
GO:0038083	peptidyl-tyrosine autophosphorylation	2,90E-03
GO:1904018	positive regulation of vasculature development	2,96E-03
GO:0031406	carboxylic acid binding	3,08E-03
GO:0001816	cytokine production	3,13E-03
GO:0070570	regulation of neuron projection regeneration	3,39E-03
GO:0043177	organic acid binding	3,41E-03
GO:0048679	regulation of axon regeneration	3,46E-03
GO:0098552	side of membrane	3,46E-03
GO:0007566	embryo implantation	3,52E-03
GO:0051769	regulation of nitric-oxide synthase biosynthetic process	3,62E-03
GO:0001649	osteoblast differentiation	3,66E-03
GO:0004714	transmembrane receptor protein tyrosine kinase activity	3,81E-03
GO:0010811	positive regulation of cell-substrate adhesion	3,93E-03
GO:0033273	response to vitamin	4,25E-03
GO:0060192	negative regulation of lipase activity	4,28E-03
GO:0034377	plasma lipoprotein particle assembly	4,29E-03
GO:0032369	negative regulation of lipid transport	4,36E-03
GO:0045766	positive regulation of angiogenesis	4,39E-03
GO:0048675	axon extension	4,41E-03
GO:0001938	positive regulation of endothelial cell proliferation	4,44E-03
GO:0034364	high-density lipoprotein particle	4,44E-03

GO:0014074	response to purine-containing compound	4,47E-03
GO:0051004	regulation of lipoprotein lipase activity	4,49E-03
GO:0016209	antioxidant activity	4,51E-03
GO:0045940	positive regulation of steroid metabolic process	4,55E-03
GO:0048261	negative regulation of receptor-mediated endocytosis	4,60E-03
GO:0045806	negative regulation of endocytosis	4,61E-03
GO:0065005	protein-lipid complex assembly	4,70E-03
GO:0005088	Ras guanyl-nucleotide exchange factor activity	4,85E-03
GO:0035987	endodermal cell differentiation	4,87E-03
GO:0033293	monocarboxylic acid binding	5,02E-03
GO:0010927	cellular component assembly involved in morphogenesis	5,06E-03
GO:0050679	positive regulation of epithelial cell proliferation	5,12E-03
GO:0071363	cellular response to growth factor stimulus	5,13E-03
GO:0060142	regulation of syncytium formation by plasma membrane fusion	5,33E-03
GO:0010830	regulation of myotube differentiation	5,35E-03
GO:0004806	triglyceride lipase activity	5,41E-03
GO:0032370	positive regulation of lipid transport	5,44E-03
GO:1905952	regulation of lipid localization	5,52E-03
GO:0007218	neuropeptide signaling pathway	5,55E-03
GO:0050920	regulation of chemotaxis	5,55E-03
GO:0060143	positive regulation of syncytium formation by plasma membrane fusion	5,56E-03
GO:1905953	negative regulation of lipid localization	5,81E-03
GO:0006954	inflammatory response	6,21E-03
GO:0048514	blood vessel morphogenesis	6,40E-03
GO:0045211	postsynaptic membrane	6,69E-03
GO:0008083	growth factor activity	7,04E-03
GO:0005769	early endosome	7,07E-03
GO:1903708	positive regulation of hemopoiesis	7,11E-03
GO:0008211	glucocorticoid metabolic process	7,13E-03
GO:0046503	glycerolipid catabolic process	7,14E-03
GO:0001558	regulation of cell growth	7,16E-03
GO:0055072	iron ion homeostasis	7,45E-03
GO:0004713	protein tyrosine kinase activity	7,74E-03
GO:1903706	regulation of hemopoiesis	7,76E-03
GO:0006879	cellular iron ion homeostasis	7,82E-03
GO:0010984	regulation of lipoprotein particle clearance	7,89E-03
GO:0017127	cholesterol transporter activity	8,30E-03
GO:0030308	negative regulation of cell growth	8,31E-03
GO:0045926	negative regulation of growth	8,41E-03
GO:0001078	transcriptional repressor activity, RNA polymerase II core promoter proximal region sequence-specific binding	8,47E-03
GO:0000083	regulation of transcription involved in G1/S transition of mitotic cell cycle	8,50E-03
GO:0070848	response to growth factor	8,50E-03
GO:0045927	positive regulation of growth	8,75E-03
GO:0070374	positive regulation of ERK1 and ERK2 cascade	8,78E-03
GO:0001819	positive regulation of cytokine production	8,84E-03
GO:0005178	integrin binding	9,06E-03
GO:0043425	bHLH transcription factor binding	9,64E-03
GO:0071675	regulation of mononuclear cell migration	9,78E-03
GO:0002042	cell migration involved in sprouting angiogenesis	9,88E-03
GO:0010832	negative regulation of myotube differentiation	9,99E-03