



**Ministério da Saúde  
Instituto Nacional de Câncer  
Coordenação de Pós-graduação**

**INSTITUTO NACIONAL DE CÂNCER  
Pós-Graduação em Oncologia**

*PAULO THIAGO DE SOUZA SANTOS*

ESTUDO DE ALTERAÇÕES MOLECULARES EM CARCINOMA EPIDERMÓIDE DE  
ESÔFAGO ATRAVÉS DE SEQUENCIAMENTO DE RNA (RNA-seq)

**Orientador (es):** Prof. Dr. Luis Felipe Ribeiro Pinto  
Prof. Dr. Tatiana de Almeida Simão

**RIO DE JANEIRO  
2016**



**Ministério da Saúde  
Instituto Nacional de Câncer  
Coordenação de Pós-graduação**

**INSTITUTO NACIONAL DE CÂNCER  
Pós-Graduação em Oncologia**

*PAULO THIAGO DE SOUZA SANTOS*

ESTUDO DE ALTERAÇÕES MOLECULARES EM CARCINOMA EPIDERMÓIDE DE  
ESÔFAGO ATRAVÉS DE SEQUENCIAMENTO DE RNA (RNA-seq)

Tese apresentada ao Instituto Nacional de Câncer como  
parte dos requisitos para obtenção do título de Doutor  
em Oncologia

**Orientador (es):** Prof. Dr. Luis Felipe Ribeiro Pinto  
Prof. Dr. Tatiana de Almeida Simão

**RIO DE JANEIRO  
2016**

S237e Santos, Paulo Thiago de Souza.

Estudo de alterações moleculares em carcinoma epidermóide de esôfago através de sequenciamento de RNA (RNA-seq) / Paulo Thiago de Souza Santos. – Rio de Janeiro, 2016.

96 f.: il. color.

Tese (Doutorado em Oncologia) – Instituto Nacional de Câncer José Alencar Gomes da Silva, 2016.

Orientadores: Luis Felipe Ribeiro Pinto; Tatiana de Almeida Simão.

1. Neoplasias Esofágicas. 2. Carcinoma de Células Escamosas. 3. Análise de Sequência de RNA. 4. Transcriptoma. I. Pinto, Luis Felipe Ribeiro (Orient.). II. Simão, Tatiana de Almeida (Orient.). III. Instituto Nacional de Câncer José Alencar Gomes da Silva. IV. Título.

CDD 616.99432



**Ministério da Saúde  
Instituto Nacional de Câncer  
Coordenação de Pós-graduação**

**INSTITUTO NACIONAL DE CÂNCER  
Pós-Graduação em Oncologia**

***AUTOR: PAULO THIAGO DE SOUZA SANTOS***

**ESTUDO DE ALTERAÇÕES MOLECULARES EM CARCINOMA EPIDERMÓIDE  
DE ESÔFAGO ATRAVÉS DE SEQUENCIAMENTO DE RNA (RNA-seq)**

**ORIENTADOR (ES): Prof. Dr. Luis Felipe Ribeiro Pinto  
Prof. Dr. Tatiana de Almeida Simão**

**Aprovada em: \_\_\_\_/\_\_\_\_/\_\_\_\_**

**EXAMINADORES:**

**Prof. Dr.** Miguel Angelo Martins Moreira  
**Prof. Dra.** Eliana Saul Furquim Werneck Abdelhay  
**Prof. Dr.** Carlos Eduardo Guerra Schrago  
**Prof. Dr.** Michael Sammeth  
**Prof. Dr.** Martin Hernan Bonamino – Suplente I  
**Prof. Dr.** Milton Ozório Moraes – Suplente II

**RIO DE JANEIRO  
2016**

## **Dedicatória**

Dedico aos meus Pais:  
Paulo Cesar que me deu a vida  
e  
Adriano Caldeira de Araujo que me deu a vida acadêmica

## Agradecimentos

Ao meu orientador, Luis Felipe, pela confiança depositada e oportunidades oferecidas.

A minha orientadora, Tatiana Simão, por toda troca ao longo desses anos e por ter feito o impossível na correção da presente tese.

A Mariana Boroni pela facilitação na caminhada pedregosa do mundo dos comandos (Me manda seu script aí).

Ao Pedro Nicolau, que foi, é e sempre será um grande amigo e colaborador (isso soou mais como uma praga).

Ao Albert pelo companheirismo nos primeiros passos no Linux e NGS.

A Lilian que mesmo com o corpo pesado de disputar o triatlo do Aedes foi fundamental para as conclusões dos experimentos de bancada.

A Carolina Furtado pela ajuda na construção das bibliotecas e planejamento dos sequenciamentos.

Ao Renan e o Dr Miguel que gentilmente nos cederam os primers para o sequenciamento do TP53.

Ao Dr Hector que sempre nos recebeu bem na genética.

A Sheila pelas discussões e puxões de orelha logo após o anúncio de que estava indo embora (Pablo, desculpe fazer sua esposa perder a hora do jantar).

A Nathalia pelas discussões e apoio nas horas difíceis.

A Isabela pelo acolhimento, pelo ombro e incentivo nas horas de crise (e por conferir a lista de abreviaturas no fim do segundo tempo).

Aos amigos da PCM pela boa convivência nesses anos (Eu saio da senzala, mas a senzala não sairá de mim)

Aos amigos de um modo geral que torceram por mim nessa jornada

Enfim, a todos que deram alguma contribuição para a elaboração deste trabalho.

## RESUMO

O câncer de esôfago representa a oitava neoplasia mais incidente e a sexta mais letal em todo o mundo. Esta doença pode se apresentar de duas formas principais: o adenocarcinoma de esôfago e o carcinoma epidermóide de esôfago (CEE), sendo esta última, a mais frequente no Brasil e no mundo. O CEE é uma doença de etiologia bastante complexa, destacando-se o tabagismo e o etilismo. Além das questões acerca dos mecanismos de carcinogênese do esôfago, o CEE apresenta um grande desafio no que tange o seu diagnóstico e tratamento. No Brasil a sobrevida em 5 anos para essa neoplasia é menor que 10%. Isto é decorrente, principalmente, do estado avançado que os tumores são diagnosticados. O conhecimento das alterações moleculares presentes no tecido neoplásico pode responder perguntas sobre o processo de malignização. Além disso, este conhecimento pode apontar possíveis alvos para melhorar o diagnóstico e o tratamento. A utilização de métodos de análise em larga escala, como os microarranjos de DNA e mais recente o sequenciamento de nova geração, representam importantes ferramentas devido ao grande número de genes analisados sem a subjetividade do conhecimento pré-existente sobre a doença. Sendo assim, este trabalho objetivou a análise do perfil de expressão gênica, perfil mutacional bem como a presença de fusões gênicas por RNA-seq de amostras pareadas de CEE e mucosa não maligna adjacente ao tumor de 14 pacientes do INCA bem como amostras de esôfago de o pacientes sem câncer do Hospital Universitário Pedro Ernesto/UERJ. O sequenciamento do RNA foi realizado no sequenciador Illumina 2500, sendo a biblioteca construída com o kit *TruSeq* (Illumina) de acordo com as instruções do fabricante. Após o sequenciamento as leituras de baixa qualidade foram retiradas utilizando-se o software *PrinSeq*. Com as leituras de boa qualidade foram realizados os alinhamentos ao genoma humano de referência com o software TopHat2. Finalmente, com as sequências alinhadas foi possível determinar a expressão dos genes através do pacote para R/Bioconductor DESeq2. As mutações foram identificadas utilizando o protocolo do *Broad Institute* no programa *Genome Analysis Toolkit* (GATK). Para a análise de fusões gênicas foi usado o algoritmo *FusionMap*. Na comparação entre mucosa adjacente ao tumor e mucosa sadia foram classificados 522 genes como GDE (genes diferencialmente expressos) e os processos enriquecidos no GO (*Gene Ontology*) foram predominantemente relacionados a inflamação. Na comparação entre tumor e mucosa adjacente ao tumor 1.226 genes foram encontrados diferencialmente expressos. A análise de enriquecimento de vias do banco *Kyoto Encyclopedia of Genes and Genomes* (KEGG) revelou vias enriquecidas principalmente relacionadas a inflamação e outros ligados a câncer. Isso levou a validação do gene *IL8* em uma amostra independente por PCRq e o mesmo se comportou como o esperado sendo sua expressão maior no tumor do que na mucosa adjacente. A análise da via *Pathways in cancer* levou a observação do enriquecimento da via de sinalização do WNT onde 3 membros tiveram impacto na sobrevida global dos pacientes, quais sejam, *WNT16*, *WNT7B* e *PRKCQ*. Apenas o *WNT7B* teve seu valor prognóstico validado num grupo diferente de amostras. A análise de mutações foi restrita aquelas do tipo *missense*, *nonsense* e sinônima revelando uma média de 65 mutações por tumor. O gene mais mutado foi o *TP53* com uma frequência de 79% seguido pelo *LOC389831* com 43% dos tumores mutados. A frequência de mutações do gene *TP53* foi validada em uma amostra independente tendo sido de 70%. A avaliação das fusões gênicas revelou 21 eventos de fusão em sete tumores não tendo nenhum se repetido em mais de uma amostra. Nenhuma característica clínico-patológica testada apresentou associação com a presença de fusões, contudo nas amostras onde os eventos de fusão aconteceram a expressão do gene *ATR* encontrou-se aumentada. Frente a esses dados podemos concluir que a inflamação e a via de WNT são importantes para o desenvolvimento do tumor de esôfago assim como a mutação no gene *TP53*. Além disso, pode-se dizer que o fenômeno de fusão gênica é um evento consequente da carcinogênese esofágica e não causador da mesma.

Palavras-chave: RNA-seq; Carcinoma epidermóide de Esôfago; Transcriptoma

## Abstract

Esophageal cancer (EC) is the eighth most frequent and the sixth most lethal neoplasia worldwide. This disease has two main histological types: esophageal adenocarcinoma and esophageal squamous cell carcinoma (ESCC), the latter being the most common type of EC in Brazil and worldwide. ESCC has a complex aetiology, with tobacco smoking and alcohol consumption as the main risk factors in Western countries. In addition to a poorly understood molecular carcinogenesis process, ESCC represents a major challenge regarding its diagnosis and treatment. In Brazil, the 5-year survival for this cancer is less than 10%. This is mainly due to a late diagnosis. Therefore, a better comprehension of the molecular pathways that lead to ESCC development can help identifying potential targets to improve diagnosis and/or treatment. In this context, large-scale analysis methods such as DNA microarrays and next generation sequencing are important tools because of the number of genes simultaneously analyzed and the avoidance of selection bias. Thus, the present study aimed to analyze the gene expression profile, mutational profile and the presence of gene fusions in ESCC by using RNA-seq. A total of 14 patients with a confirmed diagnosis of ESCC at the Brazilian National Cancer Institute (INCA), who donated both tumor and non-malignant adjacent mucosa, and eight individuals without cancer from Hospital Universitário Pedro Ernesto were included in this analysis. RNA sequencing was performed on Illumina 2500 sequencer, and the library was prepared with TruSeq kit (Illumina), according to the manufacturer's instructions. After sequencing, the low quality reads were removed using the PrinSeq software. The remaining reads were aligned to the human genome reference using TopHat2 software. Finally, the aligned sequences were used to determine the gene expression levels (R/Bioconductor package DESeq2). Mutations were identified using the Broad Institute Protocol with Genome Analysis Toolkit (GATK), while for the gene fusion analysis we used the FusionMap algorithm. When comparing non-malignant adjacent mucosa from ESCC patients to healthy esophageal epithelium, we identified 522 differentially expressed genes (DEG). The pathway enrichment analysis showed an overrepresentation of genes related to inflammation (GO – Gene Ontology). The comparison between tumor and non-malignant adjacent mucosa revealed a total of 1,226 DEG. Once again, the enrichment analysis showed that inflammation-related genes were more represented than expected among DEG, as well as other pathways linked to cancer (KEGG - *Kyoto Encyclopedia of Genes and Genomes*). Based on this, we selected *IL8* for validation by qPCR in an independent set of samples, which showed the same profile observed in RNA-seq, a higher expression in tumors when compared with adjacent mucosa. Also, a more detailed evaluation of the “Pathways in cancer” enrichment revealed the WNT signaling pathway as altered. Three members of this pathway had an impact on overall survival, namely *WNT16*, *WNT7B* and *PRKCQ*. Only *WNT7B* had its prognostic value validated in a different set of samples. The mutational analysis, restricted to missense, nonsense, and synonymous alterations, revealed an average of 65 mutations per tumor. The most frequently mutated gene was *TP53*, with a frequency of 79%, followed by *LOC389831*, found to be mutated in 43% of the tumors. The validation in an independent group of samples confirmed the high frequency of *TP53* mutations (70%). Regarding gene fusions, a total of 21 fusion events were observed in seven tumors with no common alterations among cases. Finally, we found no associations between patients' clinic-pathological characteristics and the occurrence of fusion events, but tumor samples carrying fusions showed an increased *ATR* expression. Based on our findings, we may conclude that alterations on inflammatory response and WNT signaling pathway are important events for the development of esophageal cancer, as well as *TP53* mutations. Moreover, we may speculate gene fusion is a consequence of esophageal carcinogenesis and not its cause.

## LISTA DE ABREVIATURAS

°C	- Grau Celsius
µg	- Micrograma
µL	- Microlitro
µM	- Micromolar
ADE	- Adenocarcinoma de Esôfago
BCL	- <i>base calling</i>
BNT	- Banco Nacional de Tumores
BWA	- <i>Burrows-Wheeler Aligner</i>
cDNA	- DNA complementar
CE	- Câncer de Esôfago
CEE	- Epidermóide de Esôfago
Chr	- Cromossomo
CONICQ	- Comissão Nacional para Implementação da Convenção-Quadro
Cq	- Ciclos de Quantificação
DESeq2	- <i>Differential Gene Expression Analyses Based On The Negative Binomial Distribution</i>
DNA	- Ácido Desoxirribonucleico
dNTP	- Deoxinucleotídeo
FDA	- <i>United States Food and Drug Administration</i>
FDR	- <i>False Discovery Rate</i>
GATK	- <i>Genome Analysis Toolkit</i>
GDE	- Gene Diferencialmente Expressos
GO	- <i>Gene Ontology</i>
GSEA	- <i>Gene Set Enrichment Analysis</i>
HTSeq	- <i>Analysing High-Throughput Sequencing</i>
HUPE/UERJ	- Hospital Universitário Pedro Ernesto/ Universidade do Estado do Rio de Janeiro
ICGC	- <i>International Cancer Genome Consortium</i>
ID	- Identificação
INCA	- Instituto Nacional de Câncer
INDEL	- Inserções e/ou Deleções
KEGG	- <i>Kyoto Encyclopedia of Genes and Genomes</i>
logFC	- <i>Log Fold Change</i>
mAb	- Anticorpo Monoclonal
miR	- MicroRNA
mL	- Mililitro
ND	- Não Disponível
NGS	- <i>Next-Generation Sequencing</i>
nm	- Nanometro
nM	- Nanomolar
NxH	- Mucosa esofágica normal adjacente ao tumor <i>versus</i> mucosa esofágica de indivíduos sem câncer
OR	- <i>odds ratio</i>
pb	- Pares de Base
PCA	- Análise de Componente Principal
PCR	- Reação em Cadeia da Polimerase
PCRq	- Reação em Cadeia da Polimeras Quantitativa
pM	- Picomolar

pmol	- Picomol
QRT	- Quimiorradioterapia
RIN	- Número de Integridade do RNA
RJ	- Rio de Janeiro
RNA	- Ácido Ribonucleico
RNAm	- Ácido Ribonucleico mensageiro
RNA-Seq	- Transcriptoma
rpm	- Rotações por minute
rRNA	- Ácido Ribonucleico ribossômico
RS	- Rio Grande do Sul
SNP	- <i>Single Nucleotide Polymorphism</i>
SNV	- <i>Single Nucleotide Variant</i>
SP	- São Paulo
STAR	- <i>Spliced Transcripts Alignment to Reference</i>
TKIs	- Inibidores Tirosina Quinase
TxN	- Mucosa tumoral <i>versus</i> mucosa esofágica normal adjacente ao tumor
WES	- <i>Whole Exome Sequencing</i>

## LISTA DE FIGURAS

<b>Figura 1.1 -</b>	Representação geográfica mundial da incidência de câncer de esôfago, ajustado por idade, para ambos os sexos.....	16
<b>Figura 1.2 –</b>	História natural do CEE - Evolução temporal da carcinogênese de tumores epidermóides no esôfago e exemplificação dos momentos fundamentais para a prevenção e diagnóstico da doença.....	19
<b>Figura 1.3 –</b>	Gráfico mostrando a queda no consumo de cigarro de acordo com as implementações de políticas públicas anti-tabaco.....	20
<b>Figura 1.4 -</b>	Comparação do perfil mutacional do gene <i>TP53</i> em CEE de pacientes de diferentes regiões.....	22
<b>Figura 1.5 –</b>	Esquema ilustrando as etapas do NGS.....	26
<b>Figura 3.1 –</b>	Alinhamento em duas etapas com TOPHAT2.....	34
<b>Figura 4.1 -</b>	Avaliação de qualidade, na escala <i>Phred</i> , de uma biblioteca representativa antes (A) e depois do uso do filtro de qualidade ( <i>Prinseq</i> ).....	45
<b>Figura 4.2 –</b>	Análise de componente principal mostrando a distribuição das amostras de acordo com suas maiores diferenças. PC1 corresponde a 20% da variância e PC2 6%. Vermelho – amostras tumorais, Laranja – mucosa adjacente ao Tumor, Verde – mucosa de indivíduos sem doença no esôfago, Preto – <i>batch 1</i> , Roxo – <i>batch 2</i> e Azul – <i>batch 3</i> .	46
<b>Figura 4.3 –</b>	Seleção dos GDE baseada no percentual de pares de amostras que apresentaram $ \log_{2}FC $ maior que 1.....	49
<b>Figura 4.4 –</b>	Análise de interação proteína-proteína realizada no STRING. A seta indica <i>IL8</i> como central nas interações proteicas.....	52
<b>Figura 4.5 -</b>	Validação do gene <i>IL8</i> em 28 amostras independentes por PCRq. Comparação entre a expressão no tumor e mucosa normal adjacente mostrando maior expressão no tumor ( $*p < 0,00001$ ).....	52
<b>Figura 4.6 -</b>	Detalhamento da via enriquecida “ <i>Pathways in cancer</i> ” do KEGG a partir da comparação TxN (caixas em vermelho - genes mais expressos no tumor; caixas em branco - genes não classificados com GDE; caixas em azul - genes mais expressos na mucosa adjacente ao tumor.....	53
<b>Figura 4.7 –</b>	Gráfico de expressão hipotética representando os quatro perfis de expressão entre os diferentes tecidos analisados: saudável, adjacente ao tumor e CEE.....	54
<b>Figura 4.8 –</b>	Correlação inversa entre a expressão gênica de <i>WNK2</i> e <i>MAPK3</i> ( <i>ERK1</i> ).....	55
<b>Figura 4.9 -</b>	Validação do gene <i>MAPK3</i> em 28 amostras independentes por PCRq. (A): Diagrama de Tukey mostrando a diferença de expressão entre o tumor e normal adjacente. (B): Dot plot mostrando a mesma	55

	comparação par a par. ( $p = 0,002$ ).....	
<b>Figura 4.10 -</b>	Detalhamento dos GDE presentes na via de WNT (KEGG) (caixas em vermelho - genes mais expressos no tumor; caixas em branco - genes não classificados com GDE; caixas em azul - genes mais expressos na mucosa adjacente ao tumor).....	59
<b>Figura 4.11 -</b>	Curvas de Kaplan-Meier referentes a sobrevida global dos pacientes, em função da expressão de genes associados a via de WNT. HR: <i>Hazard Ratio</i> bruta.....	60
<b>Figura 4.12 -</b>	Correlação inversa entre a expressão gênica de <i>WNT16</i> e <i>WNT7B</i> .....	62
<b>Figura 4.13 -</b>	Validação do gene <i>WNT7B</i> em 28 amostras independentes por PCRq. (A) Comparação entre a expressão no tumor e mucosa normal adjacente mostrando maior expressão no tumor ( $p = 0,0002$ ). (B) Sobrevida global dos pacientes baseada na expressão de <i>WNT7B</i> pelo tumor com HR ajustado de 4,32 (1,08 - 17,24).....	63
<b>Figura 4.14 -</b>	Validação do gene <i>FZD6</i> em 28 amostras independentes por PCRq. (A) Comparação entre a expressão no tumor e mucosa normal adjacente mostrando maior expressão no tumor ( $*p < 0,0001$ ). (B) Sobrevida global dos pacientes baseada na expressão de <i>FZD6</i> pelo tumor com HR ajustado de 0,83 (0,28 - 2,48).....	63
<b>Figura 4.15 -</b>	Gráfico de setores mostrando o perfil de alterações encontradas nos transcritos expressos pelo tumor (A). Gráfico de setores mostrando a distribuição das alterações encontradas nos transcritos expressos pelo tumor de acordo com a região gênica, após a retirada das alterações do tipo AT:GC e InDels (B). Gráficos em <i>dot plot</i> mostrando a expressão diferencial dos genes que codificam enzimas que editam RNAm (C).....	64
<b>Figura 4.16 -</b>	Gráfico em barras mostrando a frequência dos genes mais mutados dentre as amostras de CEE, após a retirada das alterações do tipo AT:GC e InDels.....	66
<b>Figura 4.17 -</b>	Gráfico de setores ilustrando o perfil mutacional (A) bem como a proporção de cada consequência das mutações (B) no gene <i>TP53</i> analisadas por NGS em 42 tumores.....	66
<b>Figura 4.18 -</b>	Gráfico em barras mostrando a frequência de mutações encontradas por exon do gene <i>TP53</i> .....	67
<b>Figura 4.19 -</b>	Comparação entre a expressão genica do gene <i>ATR</i> mostrando maior expressão desse gene em amostras tumorais que apresentaram fusão comparadas as que não apresentaram fusão ( $*p = 0,03$ ).....	69

## LISTA DE QUADROS

<b>Quadro 3.1</b> –	Lista de oligonucleotídeos iniciadores para validação da expressão dos genes por PCRq.....	38
<b>Quadro 3.2</b> –	Lista de oligonucleotídeos iniciadores para amplificação do gene <i>TP53</i> ...	40

## LISTA DE TABELAS

<b>Tabela 4.1 -</b>	Dados clínico-patológicos dos pacientes com carcinoma epidermóide de esôfago.....	44
<b>Tabela 4.2 -</b>	Lista dos 5 genes mais sobre-expressos e dos 5 mais sub-expressos em NxH.....	47
<b>Tabela 4.3 -</b>	Processos biológicos enriquecidos com anotação do GO, resumidos pelo <i>REVIGO</i> , dentre os GDE de NxH.....	48
<b>Tabela 4.4 -</b>	Lista dos 5 genes mais sobre-expressos e dos 5 mais sub-expressos em TxN.....	50
<b>Tabela 4.5 -</b>	Processos biológicos enriquecidos com anotação do GO (com pós-análise pelo <i>REVIGO</i> ) dentre os GDE de TxN.....	50
<b>Tabela 4.6 -</b>	Vias enriquecidas com anotação do KEGG dentre os GDEs de TxN.....	51
<b>Tabela 4.7 -</b>	Genes correspondentes a cada um dos perfis de expressão encontrados na interseção das comparações NxH e TxN.....	54
<b>Tabela 4.8 -</b>	Genes diferencialmente expressos na análise TxN que fazem parte da via de WNT.....	56
<b>Tabela 4.9 –</b>	Regressão de Cox a partir das simulações de possíveis cenários para os casos com o estadiamento não informado.....	61
<b>Tabela 4.10 –</b>	Análise de fusão gênica. A tabela mostra os 21 transcritos quiméricos encontrados nos tumores usando o <i>software</i> FusionMap.....	68

## SUMÁRIO

<b>1. INTRODUÇÃO</b> .....	16
<b>1.1. INCIDÊNCIA DO CÂNCER DE ESÔFAGO</b> .....	16
<b>1.2 FATORES DE RISCO</b> .....	17
<b>1.3 HISTÓRIA NATURAL DO CEE</b> .....	18
<b>1.4 ALTERAÇÕES MOLECULARES EM CEE</b> .....	21
<b>1.5. RNA-seq e CEE</b> .....	25
<b>2. OBJETIVO</b> .....	28
<b>2.1. OBJETIVOS ESPECÍFICOS</b> .....	28
<b>3. METODOLOGIA</b> .....	29
<b>3.1. AMOSTRAS</b> .....	29
<b>3.2. EXTRAÇÃO DE RNA</b> .....	30
<b>3.3. CONFECÇÃO DA BIBLIOTECA DE cDNA</b> .....	30
<b>3.3.1. Purificação e fragmentação do RNA</b> .....	30
<b>3.3.2. Síntese da primeira fita de cDNA</b> .....	31
<b>3.3.3. Síntese da segunda fita de cDNA</b> .....	31
<b>3.3.4. Reparação de final de fita de cDNA e adenilação da porção 3'</b> .....	32
<b>3.3.5. Ligação dos Adaptadores</b> .....	32
<b>3.3.6. Enriquecimento dos Fragmentos de DNA</b> .....	32
<b>3.3.7. Validação das bibliotecas</b> .....	33
<b>3.4. SEQUENCIAMENTO</b> .....	33
<b>3.5. CONTROLE DE QUALIDADE</b> .....	33
<b>3.6. ALINHAMENTO A REFERÊNCIA</b> .....	34
<b>3.7. ANÁLISE DE EXPRESSÃO GÊNICA DIFERENCIAL</b> .....	34
<b>3.8. ANÁLISE DOS GENES DIFERENCIALMENTE EXPRESSOS</b> .....	35
<b>3.9. ANÁLISE DAS VARIANTES DE BASE ÚNICA</b> .....	35

3.10. ANÁLISE DE TRANSCRITOS QUIMÉRICOS.....	36
3.11. VALIDAÇÃO DA EXPRESSÃO GÊNICA POR PCR EM TEMPO REAL.....	36
3.11.1 Síntese de cDNA.....	36
3.11.2 PCR em tempo real (PCRq).....	37
3.12. AVALIAÇÃO DO PERFIL MUTACIONAL DO GENE TP53 EM CEE.....	38
3.12.1. Extração do DNA.....	39
3.12.2. Quantificação de DNA.....	39
3.12.3. PCR.....	39
3.12.4. Purificação do produto de PCR.....	40
3.12.5. Quantificação das bibliotecas.....	42
3.12.6. Análise in silico.....	42
3.13. ANÁLISE ESTATÍSTICA DOS DADOS HISTOPATOLÓGICOS E EPIDEMIOLÓGICOS.....	43
4. RESULTADOS .....	44
4.1. LEVANTAMENTO DE DADOS DOS PRONTUÁRIOS.....	44
4.2. ANÁLISE DA QUALIDADE DAS LEITURAS.....	45
4.3. MÉTRICAS DE LEITURA E DE MAPEAMENTO NO GENOMA.....	45
4.4. EXPRESSÃO GÊNICA .....	45
4.5. ANÁLISE DAS SNVs ( <i>single nucleotide variants</i> ).....	64
4.6. ANÁLISE DE TRANSCRITOS QUIMÉRICOS.....	67
5. DISCUSSÃO.....	70
6. CONCLUSÕES .....	82
7. REFERÊNCIAS BIBLIOGRÁFICAS .....	83

# 1. INTRODUÇÃO

## 1.1. INCIDÊNCIA DO CÂNCER DE ESÔFAGO

O câncer de esôfago (CE) é a sexta neoplasia mais incidente entre homens e a décima terceira entre as mulheres no mundo (Ferlay *et al*, 2015). As taxas de incidência do CE variam muito ao redor do Mundo, onde é possível notar áreas de alta incidência próximas a áreas de baixa incidência. As áreas de maior incidência são a região que vai do mar Cáspio ao centro da China (região conhecida como cinturão asiático do CE), Europa Ocidental (principalmente a França), o sudeste da África e o sul da América do sul (Figura 1) (Ferlay *et al*, 2013).

Estimativas recentes indicam que a incidência de CE não irá diminuir nos próximos anos e em 2025 são esperados mais de 640 mil novos casos em ambos os sexos, número 40% maior do que o observado em 2012 (Ferlay *et al*, 2013)

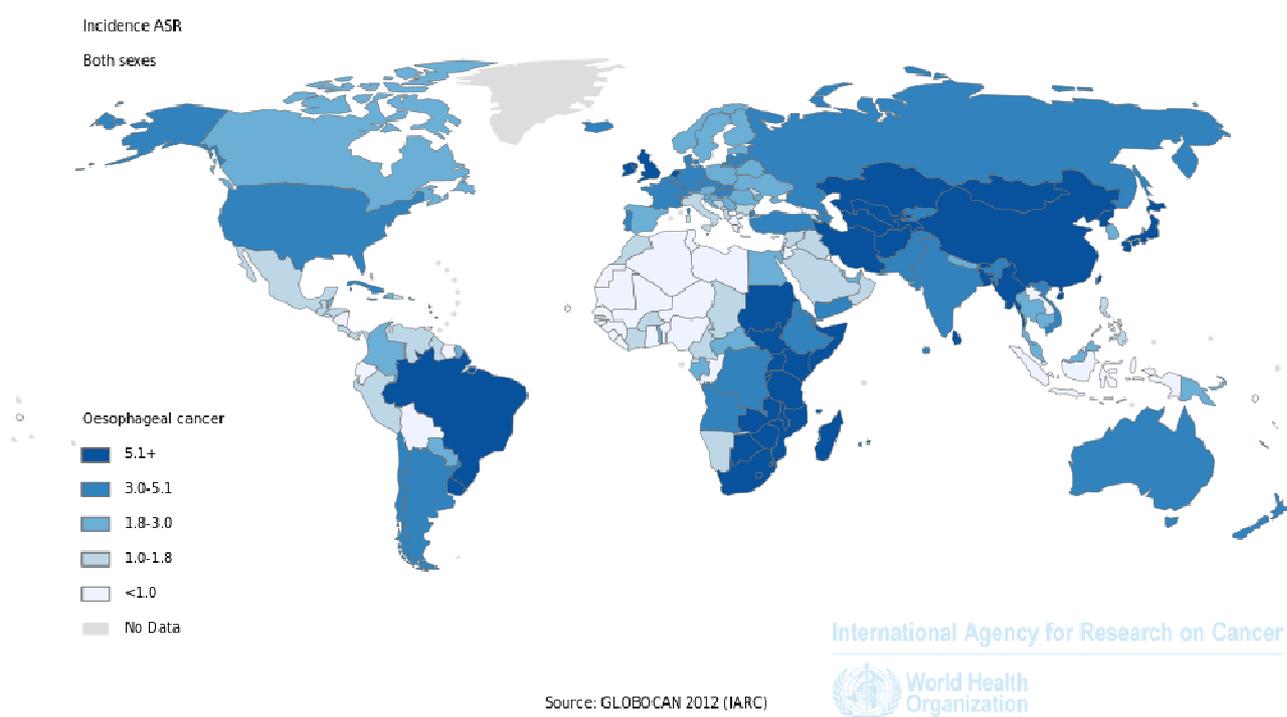


Figura 1.1 - Representação geográfica mundial da incidência de câncer de esôfago, ajustado por idade, para ambos os sexos (Ferlay *et al*, 2013).

No Brasil esse quadro é muito similar ao restante do mundo, ocupando a sexta posição entre os homens e a décima terceira entre as mulheres dentre as neoplasias mais incidentes. Existe uma grande variação entre as regiões do Brasil, por exemplo, na região sul o CE é o quinto tumor mais incidente entre os homens enquanto que nas regiões sudeste e nordeste figura na sétima posição (Estimativa

INCA, 2016). Essa discrepância se deve principalmente aos diferentes fatores de risco que serão discutidos abaixo.

## 1.2. FATORES DE RISCO

O CE é dividido principalmente em 2 tipos histológicos: Carcinoma Epidermóide de Esôfago (CEE) e Adenocarcinoma de Esôfago (ADE). O ADE responde a menos de 10% dos casos de CE no Brasil e é mais frequente em países desenvolvidos como Estados Unidos e Alemanha. Este tipo histológico tem como principais fatores de risco associados à doença, a obesidade, o refluxo gastroesofágico e a metaplasia intestinal do esôfago (Esôfago de Barrett) (Hongo, Nagasaki e Shoji, 2009).

O CEE, objeto de estudo deste trabalho, possui como principais fatores de risco o consumo moderado/alto de álcool e o hábito de fumar. Em estudo realizado por Castellsagué e colaboradores (1999), pacientes que não bebiam e fumavam apresentavam uma *odds ratio* (OR) entre 6,5 e 7,5 em relação aos controles ao passo que pacientes que não fumavam e bebiam apresentaram uma OR de 7,0 a 14 em relação aos controles. Quando foram avaliados pacientes que fumavam e bebiam a OR variou entre 4,0 e 51,0 mostrando que os efeitos carcinogênicos do cigarro e do álcool são potencializados um pelo outro. Além disso, há a sugestão de que o fumo tenha um papel importante tanto na iniciação quanto na promoção do tumor, enquanto o álcool teria um papel mais relevante na promoção, uma vez que após 10 anos de abstinência de álcool o risco retorna aos valores dos abstêmios, enquanto após 10 anos sem tabaco o risco diminui apenas 50,0% (Castellsagué *et al.*, 1999).

Outro fator de risco importante e presente em áreas de alta incidência como o sul da América do Sul e o “Cinturão Asiático” é o consumo de bebidas em altas temperaturas como chá e chimarrão (erva-mate consumida em altas temperaturas), sendo este último um hábito muito frequente no Rio Grande do Sul. Estima-se que, neste estado, 70% dos homens e 50% das mulheres sejam consumidores habituais (Victoria *et al.*, 1987). Alguns trabalhos mostraram que o consumo do chimarrão aumentou a chance de se desenvolver CEE em até 4 vezes dependendo da temperatura de consumo e quantidade consumida por dia (Castelleto *et al.*, 1994; Rolon *et al.*, 1995; Castellsagué *et al.*, 1999), sugerindo que a lesão térmica estaria gerando um processo inflamatório crônico no esôfago que contribuiria para o aparecimento do CEE.

Outros fatores de risco são conhecidos, mas possuem menor importância no Brasil como alimentos contaminados com nitrosaminas, compostos com ação carcinogênica potente e que apresentam grande organotropismo para o esôfago (Lijinsky, 1992).

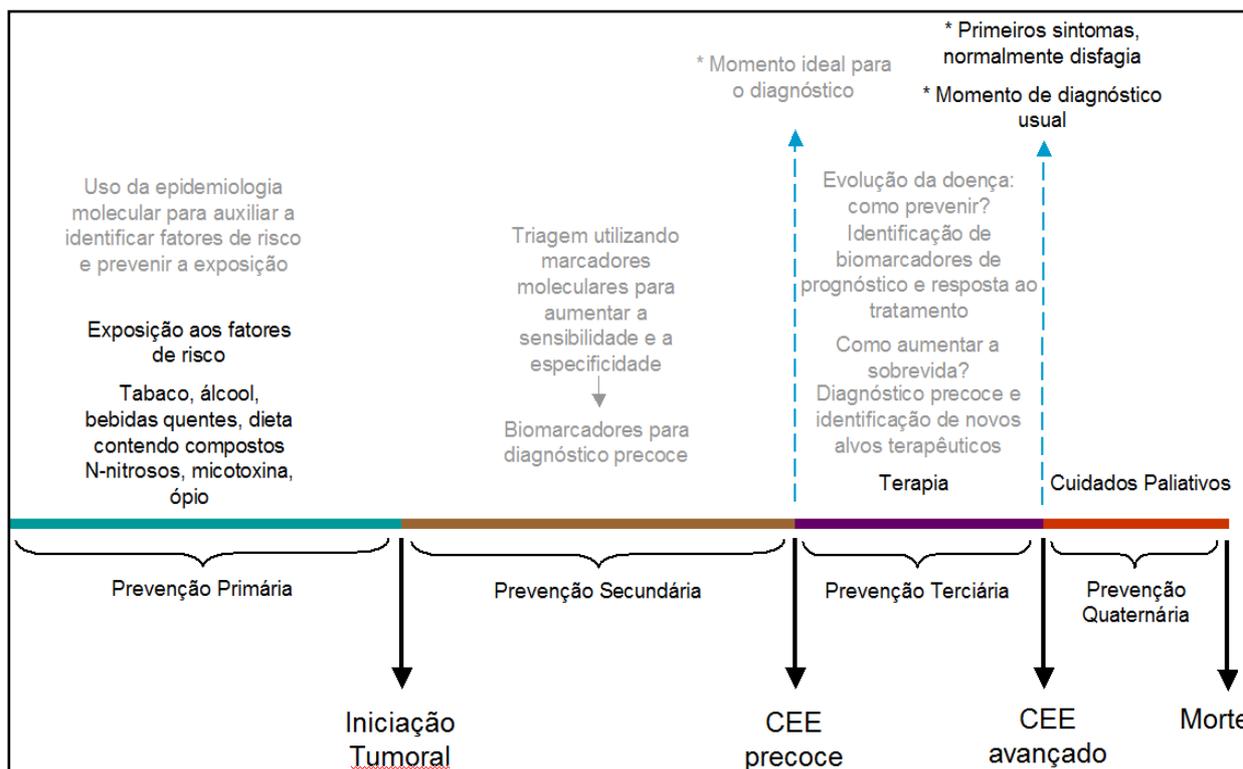
### **1.3. HISTÓRIA NATURAL DO CEE**

A história natural do CEE começa com uma exposição aos fatores de risco por um longo período que culminará na malignização da mucosa esofágica. Quando o tumor é detectado precocemente existe a possibilidade de excisão do tumor *in situ* com grande chance de cura (Chen *et al.*, 1999). Contudo, a natureza assintomática dessa doença leva a um diagnóstico tardio para a maioria dos pacientes.

A disfagia (dificuldade de deglutir) é o primeiro sintoma mais frequente do CEE e geralmente aparece quando mais de 70% da luz do esôfago está obstruída (Stein *et al.*, 2001), neste ponto 40% dos tumores já tem metástases linfonodais ou a distância (Jemal *et al.*, 2006). Além disso, características anatômicas do esôfago como a ausência de serosa, rica rede linfonodal e proximidade a estruturas nobres como a aorta e a coluna vertebral tornam a invasão mais rápida, a disseminação mais fácil e em alguns casos impossibilita a ressecção cirúrgica do tumor. Em alguns casos, quando a traqueia é acometida, há ainda a inelegibilidade do uso de radioterapia, pois nesse caso haveria a formação de uma fístula traqueo-esofágica impedindo o paciente de se alimentar por via oral (Reed, 1995).

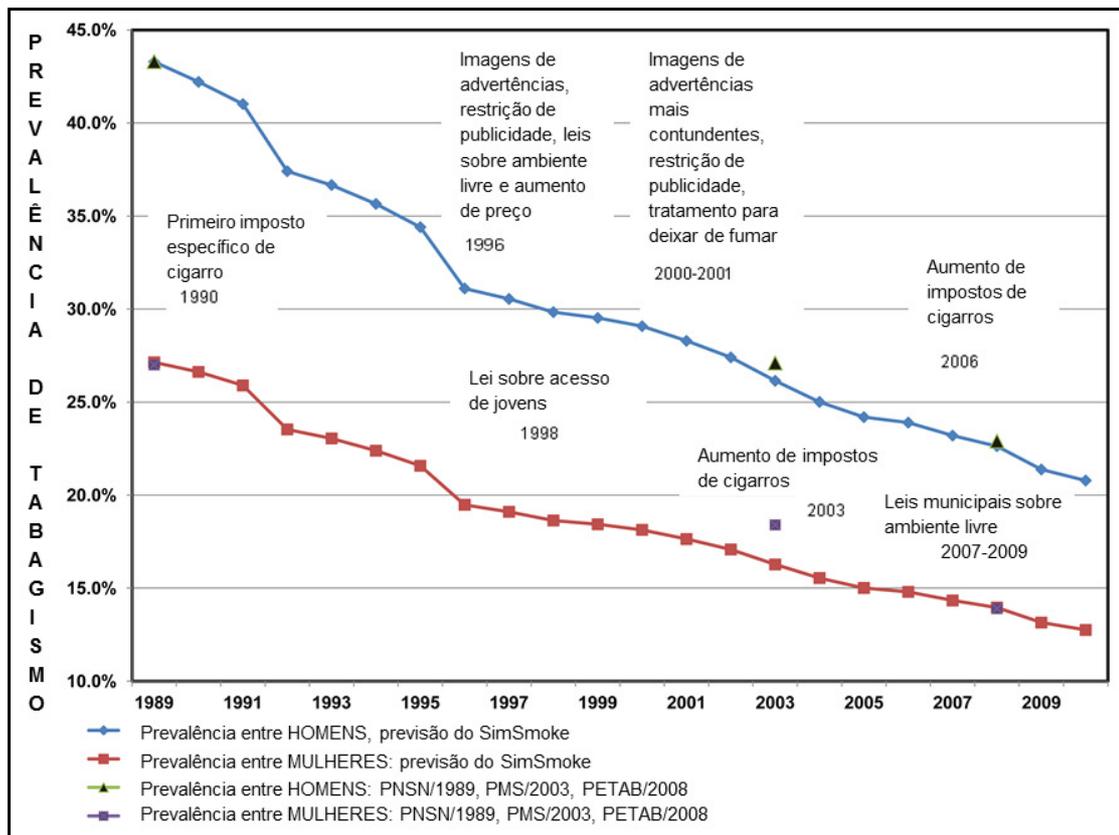
Nos casos em que o paciente não tem possibilidade de tratamento curativo alguns cuidados paliativos são tomados a fim de melhorar a qualidade de vida do paciente como a colocação de uma prótese no esôfago para aumentar a luz deste órgão permitindo que o paciente volte a se alimentar pela boca (Freeman, Ascioti e Mahidhara, 2012).

O cenário descrito acima leva a alta letalidade do CEE sendo a sobrevida em 5 anos inferior a 20% (Cohen e Ajani, 2011).



**Figura 1.2 – História natural do CEE - Evolução temporal da carcinogênese de tumores epidermóides no esôfago e exemplificação dos momentos fundamentais para a prevenção e diagnóstico da doença (adaptado de da Costa *et al.*, 2013).**

O curso dessa doença (figura 1.2) pode ser perturbado em vários momentos da sua evolução (da Costa *et al.*, 2013). Como medidas de prevenção primária podem-se adotar medidas que evitem a exposição aos fatores etiológicos. No Brasil, foram adotadas ações anti-tabaco que foram muito bem-sucedidas e estão ilustradas na figura 1.3. Em 1989 a prevalência de tabagismo em adultos homens era de 43,3% e 27,0% entre as mulheres, no ano de 2013 essa prevalência caiu para 18,9% e 11,0%, respectivamente (Levy, Almeida e Szklo, 2012). Infelizmente para o consumo de álcool a realidade é outra. Em 2006 a prevalência de abuso de bebidas alcoólicas no estado do Rio de Janeiro entre homens adultos era de 17,9% e 8,6% entre as mulheres. Em 2013 essa prevalência cresceu para 22,6% e 10,2%, respectivamente (VIGITEL, DATASUS). Esses dados sugerem a necessidade da adoção de políticas mais efetivas contra o etilismo.



**Figura 1.3 – Gráfico mostrando a queda no consumo de cigarro de acordo com as implementações de políticas públicas anti-tabaco (Levy, Almeida e Szklo, 2012, adaptado por Secretaria-Executiva da CONICQ)**

Uma outra possibilidade de intervenção com grande impacto no desfecho desfavorável do CEE, seria o diagnóstico precoce oriundo de rastreamento. Uma estratégia já utilizada para isso foi o uso do balão citológico que consistia em inflar um balão no estômago e trazê-lo inflado pela boca raspando as células da mucosa esofágica, essas células seriam então examinadas na busca por células morfolologicamente malignas. Este procedimento foi utilizado na China em uma província de alta incidência de CEE e a sensibilidade ficou abaixo de 40% (Dawsey *et al.*, 1997).

A técnica mais sensível para detecção do CEE é a endoscopia. No Japão devido à alta incidência de câncer no estômago é realizado o rastreamento endoscópico deste tumor na população. Eventualmente são detectados tumores no esôfago durante este rastreamento o que tem um efeito positivo por fazer um diagnóstico do CEE mais cedo. Este efeito benéfico se reflete na letalidade do CEE no Japão (62%) ser menor que a da China (80%) que não faz esse rastreamento endoscópico.

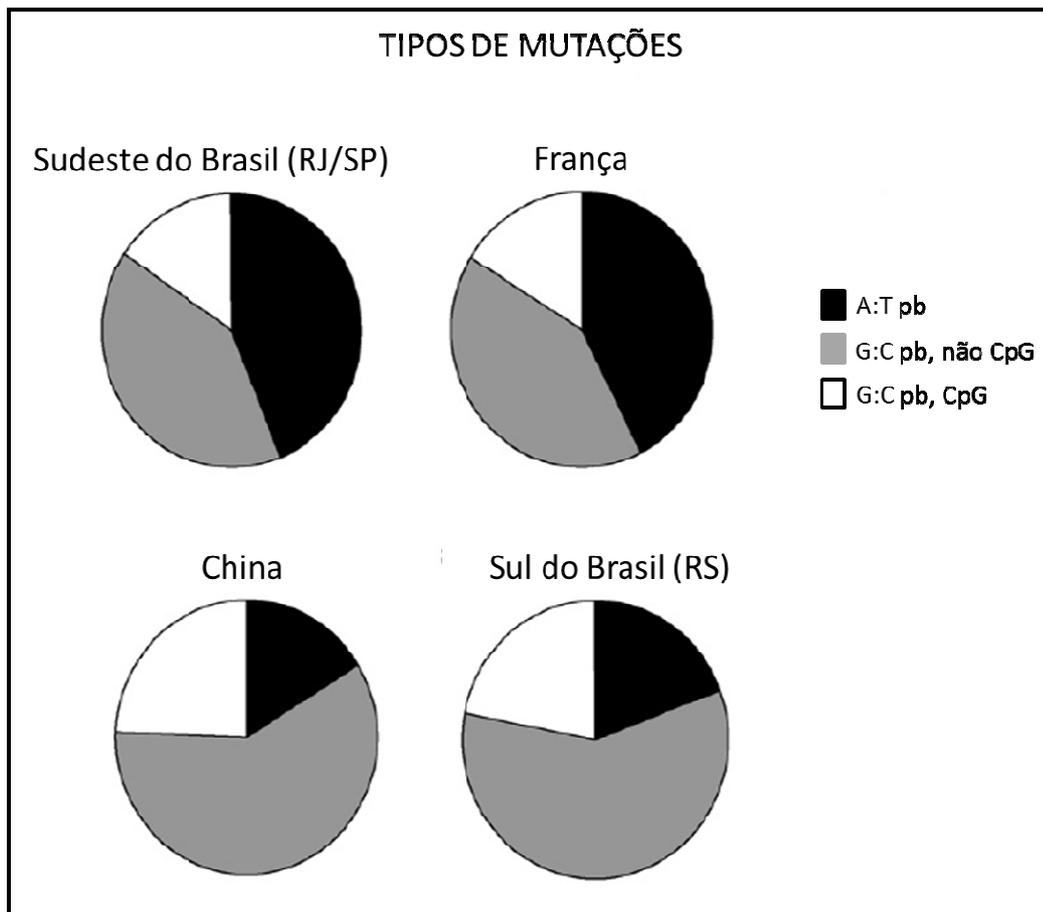
#### 1.4. ALTERAÇÕES MOLECULARES EM CEE

A identificação e caracterização das principais alterações moleculares em tumores podem trazer vários esclarecimentos sobre a gênese de um tumor bem como apontar novos caminhos terapêuticos e estratégias de diagnóstico.

O gene *TP53*, que codifica a proteína p53, é um dos genes mais alterados em CEE e a frequência de mutações nesse gene pode variar de 25 a 89% dependendo do país de origem do paciente (Olivier *et al*, 2002). A proteína p53 desempenha um papel central na homeostasia da célula por participar ativamente de processos como parada do ciclo celular na presença de danos no DNA, ativar vias de reparo de DNA e induzir apoptose quando o reparo dos danos não é possível. Por essa razão essa proteína tem um papel fundamental no processo carcinogênico.

A presença de mutações no gene *TP53* foi associada com menor ou nenhuma resposta a quimiorradioterapia (QRT) em pacientes com CEE em dois estudos asiáticos (Makino *et al*, 2010; Yamasaki *et al*, 2010) contudo o mesmo não foi visto por Fernandes (2013) que não encontrou valor preditivo a QRT nas mutações no *TP53* numa amostra de pacientes do Rio de Janeiro.

No que tange as mutações no gene *TP53* o perfil mutacional de pacientes do Rio de Janeiro foi semelhante ao perfil de mutações encontrado em pacientes franceses enquanto que o perfil de mutação de indivíduos do Rio Grande do Sul foi semelhante a indivíduos chineses (Figura 1.4). A diferença regional no padrão de mutações encontrado pode estar relacionada com a exposição a diferentes fatores de risco como o consumo de bebidas em alta temperatura no Sul do Brasil e na China (Rossini *et al*, 2010).



**Figura 1.4 - Comparação do perfil mutacional do gene *TP53* em CEE de pacientes de diferentes regiões (Rossini *et al*, 2010). pb = pares de bases**

A via dos receptores HER está alterada em diferentes tipos de câncer (Sergina e Moasser, 2007). Alguns membros dessa via como EGFR, HER2, Ras e Raf possuem fármacos específicos que bloqueiam suas funções e, portanto, auxiliam o combate de tumores que são dependentes desta via para seu crescimento. Atualmente, diversos fármacos são desenvolvidos para bloquearem proteínas específicas e, por essa razão, são chamadas de terapias-alvo. Duas classes de fármacos se destacam neste sentido: os inibidores tirosina quinase (TKIs) e os anticorpos monoclonais (mAbs) (Baselga, 2002).

Os TKIs apresentam capacidade inibitória através da sua competição e maior interação pelos sítios de ligação ao ATP de suas proteínas-alvo, bloqueando sua atividade. Fármacos como Gefitinib e Erlotinib, ambos inibidores de EGFR, foram aprovados pelo FDA e são utilizados na terapia de pacientes acometidos com câncer de pulmão não pequenas células (Harari, 2004; FDA, 2014).

Os mAbs foram desenvolvidos para atuar na porção extracelular de proteínas competindo pelo sitio de interação ao ligante de receptores transmembrana. Além

disso, esta classe de fármaco é capaz de levar esses receptores-alvos à endocitose, via clatrina, encaminhando-os para degradação via proteossomo (Capdevilla, 2009).

A terapia do CEE é bastante ineficaz, como já mencionado anteriormente, portanto é desejável a busca por novas estratégias terapêuticas. O sucesso do uso das terapias-alvo em tumores como os de pulmão, colorretal e mama motivou a busca por marcadores que justificassem o uso dessas terapias também no CEE.

Liu *et al* (2011) em um estudo conduzido na China encontraram em 14% das amostras de CEE mutações no gene *EGFR* e em 12% mutações no gene *KRAS*. Por outro lado, Gonzaga *et al* (2012) estudando a população brasileira não encontraram nenhuma mutação em *EGFR*, *KRAS* e *BRAF* utilizando mais que o dobro de pacientes do estudo chinês, embora tenham encontrado superexpressão proteica de HER2 em 4% dos casos. Desta forma, pacientes chineses e brasileiros apesar de terem o mesmo tipo histológico de tumor poderiam se beneficiar com estratégias terapêuticas diferentes.

O advento do sequenciamento de próxima geração (que será discutido mais detalhadamente abaixo) possibilitou a descoberta de diversas alterações importantes para a carcinogênese esofágica. Contudo, o uso desta ferramenta na investigação dos mecanismos envolvidos no desenvolvimento do CEE permanece exíguo, uma vez que utilizando como palavras-chave “*esophageal squamous cell carcinoma*” e “*exome*” em uma busca no *PUBMED* (<https://www.ncbi.nlm.nih.gov/pubmed/>) são encontrados menos de dez artigos. Dos trabalhos disponíveis, cinco avaliaram apenas tumores de pacientes orientais, principalmente chineses, criando um lapso no conhecimento da genômica de tumores de pacientes com CEE no ocidente.

Até o momento há apenas um trabalho que avaliou o sequenciamento de exoma total (WES, do inglês, *Whole Exome Sequencing*) em uma população ocidental (Agrawal *et al.*, 2012). Neste estudo, foram sequenciados 12 amostras de tecido tumoral e mucosa adjacente ao tumor de pacientes americanos. Foram encontrados 884 genes mutados com uma média de 83 genes mutados por tumor sendo *TP53*, *NOTCH1*, *NOTCH3*, *FBXW7*, *KIF16B*, *KIF21B* e *MYCBP2*, os genes frequentes em pelo menos três amostras.

Song *et al* (2014) fizeram o WES de 71 amostras de CEE e mucosas adjacentes ao tumor. A partir de um painel de 96 mutações foi possível estabelecer um agrupamento dos pacientes que nunca consumiram bebida alcoólica e tinham um prognóstico melhor. Além disso, encontraram 8 genes dos quais 7 são genes

classicamente envolvidos no processo de carcinogênese (*TP53*, *RB1*, *CDKN2A*, *PIK3CA*, *NOTCH1*, *NFE2L2* e *ADAM29*) e um ainda não tinha sido associado ao câncer, o gene *FAM135B* (família com similaridade de sequência 135, membro B). O *FAM135B* foi usado em ensaios funcionais com linhagens celulares provenientes de CEE e sua inibição foi capaz de reduzir o crescimento celular, formação de colônias, migração e invasão, mostrando a importância desse gene na carcinogênese do esôfago.

Lin *et al* (2014) também fizeram WES de 20 pares de amostras (tumor e mucosa normal adjacente ao tumor) e encontraram 1.042 genes mutados. Classificaram como principais 13 genes: *TP53*, *KMT2D*, *FAT1*, *FAT2*, *RB1*, *NOTCH1*, *EP300*, *ZNF750*, *PIK3CA*, *KDM6A*, *NFE2L2*, *PTEN* e *CDKN2A*. Além disso, a proteína APOBEC3B, uma DNA citidina deaminase, foi apontada como possível responsável pelas mutações em citidina que estavam presentes em sítios de reconhecimento desta enzima, uma vez que a expressão de *APOBEC3B* foi encontrada aumentada nos tumores.

Ainda em 2014, foi publicado um trabalho que analisou 113 pares de amostras tumorais e mucosa normal adjacente e encontraram mais de 9.000 genes mutados. Dentre os principais genes mutados estão o *TP53*, *CDKN2A*, *AJUBA*, *RB1*, *NOTCH1* e *NFE2L2*. Os autores apontaram como significativo o número de mutações em genes relacionados à modificação em histonas, quais sejam, *EP300*, *CREBBP*, *KMT2D*, *KMT2C*, *KDM6A* e *NSD1*. Vale ressaltar ainda que o status mutado para o gene *EP300* foi associado a um pior prognóstico tanto nas amostras sequenciadas por WES quanto em uma coorte independente de 246 pacientes (Gao *et al.*, 2014).

Após análise de 192 amostras tumorais por WES, Zhang *et al* (2015) identificaram 3 tipos de assinatura no perfil mutacional uma correspondente a mutações induzidas por APOBEC, uma assinatura em sítios CpG e uma assinatura de procedência desconhecida. Além disso eles foram capazes de associar o perfil de APOBEC com mutações no gene *PIK3CA*. Os genes mais frequentemente mutados foram *TP53*, *NOTCH1*, *PIK3CA* e *FAT1*.

Recentemente, foram analisadas por WES 144 amostras de CEE e concluíram que o tipo de mutação mais frequente foi a substituição de C por T e que essa transição acontecia em sítios alvo de APOBEC. Eles ainda foram capazes de associar a assinatura de APOBEC com mutações no gene *ZNF750* que é um regulador da diferenciação epidérmica. Os genes mais frequentemente mutados

foram *TP53*, *NOTCH1*, *MLL2*, *NFE2L2*, *ZNF750*, *FAT1* e *PIK3CA*. Além disso, mutações no gene *TET2* e *EP300*, ambos relacionados à regulação epigenética, foram associados a um pior prognóstico dos pacientes (Sawada *et al.*, 2016).

A ativação de um proto-oncogene ou inativação de um gene supressor de tumor não acontece exclusivamente por meio de mutações. Alterações epigenéticas são igualmente importantes no processo carcinogênico de qualquer tecido. Lima *et al* (2011) utilizando microarranjo para avaliar o perfil de metilação global em pacientes brasileiros com CEE encontraram mudanças no padrão de metilação da região promotora de genes associados a proteção tecidual como *TFF1*, a diferenciação celular como *DSG1* e indução de proliferação como *BCL3*.

Além de metilação do DNA, outras alterações epigenéticas foram encontradas em CEE como é o caso dos microRNAs. MicroRNAs são capazes de regular a expressão de genes de forma pós-transcricional já que participam da degradação do RNAm ou progressão do ribossomo, desta forma impedindo que a proteína seja traduzida (Grosshans e Filipowicz, 2008). Yang *et al* (2013) observaram que os microRNAs miR-338-3p, miR-218 e miR-139-5p apresentavam expressão suprimida no tumor em relação ao tecido adjacente não tumoral enquanto que miR-183, miR-574-5p, miR-21 e miR-601 apresentavam superexpressão no tumor.

### **1.5. RNA-seq e CEE**

O sequenciamento de próxima geração tem possibilitado a aquisição de sequências genômicas de diversos espécimes em um tempo curto e com custo relativamente baixo. Quando aplicado com o objetivo de obter o transcriptoma (RNA-seq) ele supera em eficiência outras metodologias em larga escala como o microarranjo, pois a metodologia de RNA-seq é mais eficiente que a de microarranjo na determinação das diferenças na expressão de genes. Isso ocorre porque o RNA-seq não se baseia em hibridização, ou seja, não gera falsos positivos devido à hibridização cruzada entre sondas similares e, além disso, aumenta a chance de encontrar um RNA raro já que é capaz de fazer a contagem absoluta dos RNAm (Marioni *et al*, 2008).

A tecnologia mais usada atualmente e que foi utilizada neste trabalho é a de sequenciamento por síntese da Illumina. Trata-se de um sequenciamento por síntese similar, em conceito, ao método de Sanger. A diferença principal está na aplicação de nanotecnologia e marcações na construção da biblioteca que permitem sequenciar várias amostras de uma vez em uma superfície sólida.

A figura 1.5 esquematiza as etapas de sequenciamento que serão mais detalhadas a frente. Resumidamente, o material a ser sequenciado (DNA ou cDNA) é fragmentado por sonicação ou enzimaticamente, a esses fragmentos gerados são adicionados adaptadores que irão identificar a amostra e se ligar na matriz da *flowcell*. Uma vez ligados à matriz, os fragmentos são amplificados por um processo chamado amplificação por ponte. Depois da amplificação é feito o sequenciamento por síntese onde cada base inserida emite uma luz diferente (Bras, Guerreiro e Hardy, 2012).

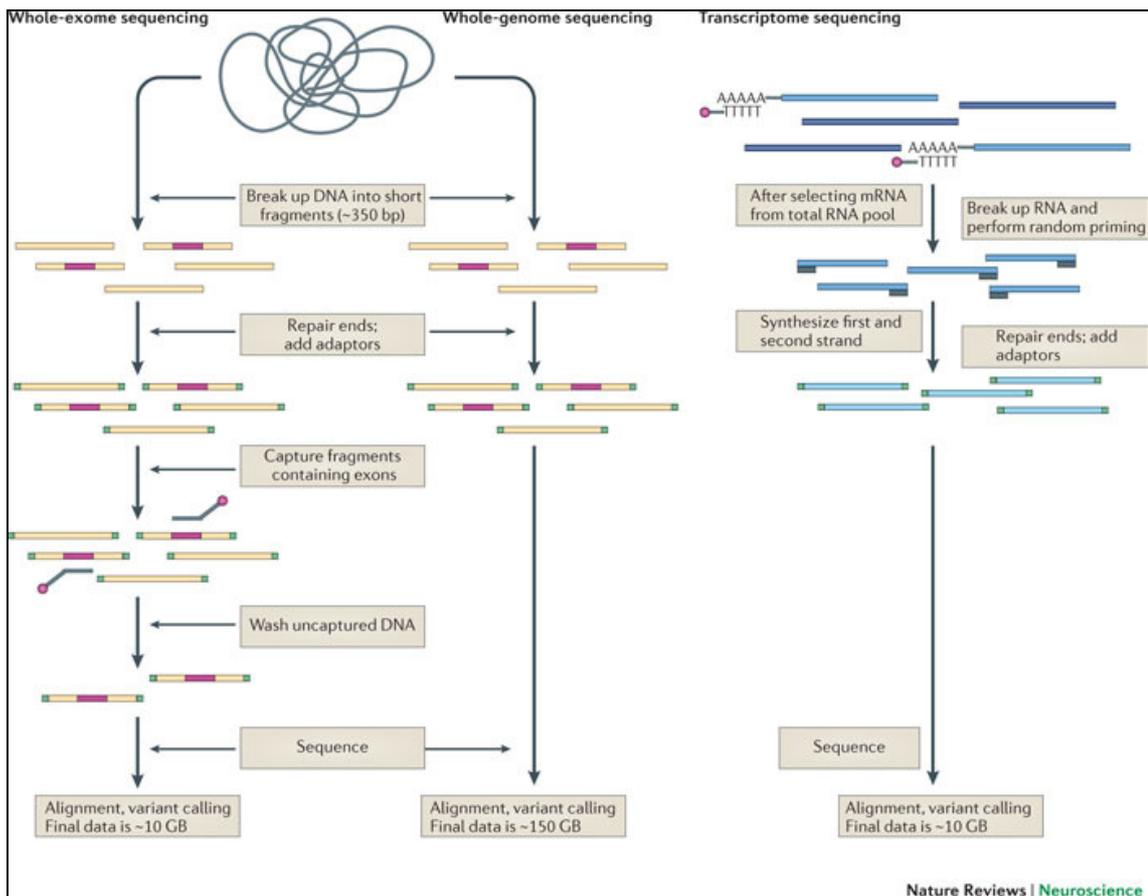


Figura 1.5 – Esquema ilustrando as etapas do NGS (Bras, Guerreiro e Hardy, 2012).

Assim como já foi dito anteriormente para exomas, o RNA-seq de CEE também é muito raro na literatura e os três trabalhos encontrados até o momento serão mencionados a seguir.

O primeiro trabalho publicado em CEE, realizou um RNA-seq em 3 pares de amostras de CEE e mucosa adjacente ao tumor tendo encontrado 1.425 genes diferencialmente expressos. Após análises de enriquecimento desse conjunto de genes observaram que grande parte dos processos enriquecidos estavam relacionados à migração celular, apoptose, reparo e replicação de DNA, regulação do ciclo celular e resposta ao dano no DNA. O gene *PTK6* foi escolhido para

validação funcional porque estava subexpresso nos três tumores. Os autores concluíram que o *PTK6* é um importante gene supressor de tumor no esôfago, pois sua superexpressão impediu o crescimento de tumores xenotransplantados em camundongos *nude* e sua inibição fez os tumores crescerem mais que o controle. Além disso, sugeriram que a ação supressora de tumor de *PTK6* esteja relacionada com a regulação negativa que essa proteína exerce sobre *AKT* (Ma *et al.*, 2012).

No mesmo ano, Tong *et al* (2012) analisaram a expressão diferencial de sete amostras de pacientes com CEE e cinco amostras não tumorais de esôfago por RNA-seq. Eles encontraram 1.730 genes diferencialmente expressos dos quais encontrava-se enriquecida a via de sinalização de integrina. Analisando os genes presentes nessa via apenas um gene se encontrava mais expresso no tecido saudável do que no tumor, o gene *RAB25*. A partir dessa análise foi avaliada a influência de *RAB25* em diversos parâmetros de malignidade e *RAB25* mostrou uma atividade supressora de tumor reduzindo a migração, a tumorigênese e a angiogênese de linhagens neoplásicas de esôfago.

Finalmente Jiang *et al* (2014) sequenciaram o RNA de 63 pares de amostras (tumor e normal adjacente), e encontraram 2.419 genes diferencialmente expressos. Nesse conjunto de genes, as vias de aderência e motilidade celular, resposta imune, metabolismo multi-celular e desenvolvimento estavam sobre-representadas. Além disso, no mesmo trabalho foi relatado uma fusão gênica detectada por RNA-seq, entre os genes *HLA-E* e *HLA-B*, ambos no cromossomo 6.

Assim, torna-se claro a partir da revisão da literatura que existem poucos trabalhos publicados na avaliação de alterações genômicas em larga escala em CEE, particularmente em pacientes ocidentais que possuem fatores de risco diferentes dos orientais.

## **2. OBJETIVO**

Avaliar o perfil global de alterações moleculares em carcinoma epidermóide de esôfago (CEE).

### **2.1. OBJETIVOS ESPECÍFICOS**

- Analisar o perfil de expressão gênica global das amostras de CEE a fim de identificar os genes diferencialmente expressos comparando tecido esofágico saudável, mucosa adjacente ao tumor e tecido tumoral;

- Selecionar e enriquecer os genes diferencialmente expressos utilizando as plataformas *Gene Ontology* e *KEGG* com a finalidade de obter vias e processos biológicos que estejam alterados em CEE;

- Analisar o perfil mutacional global das amostras de CEE comparando tecido esofágico saudável, mucosa adjacente ao tumor e tecido tumoral;

- Avaliar a presença de transcritos quiméricos;

- Selecionar genes para posterior validação dos resultados obtidos nas análises de expressão gênica em um número maior de amostras pela técnica de PCR em tempo real;

- Validar o perfil mutacional dos genes selecionados em um número maior de amostras;

- Determinar como as alterações moleculares se relacionam com os dados clínico-patológicos.

### **3. METODOLOGIA**

#### **3.1. AMOSTRAS**

Foram utilizadas para esse trabalho amostras armazenadas no Banco Nacional de Tumores (BNT) do Instituto Nacional de Câncer (INCA), de pacientes com diagnóstico confirmado de CEE que foram admitidos na Instituição no período de 2006 a 2011. Somente foram incluídos pacientes que não foram submetidos a qualquer tratamento prévio de radio e/ou quimioterapia. Foram coletados tecido tumoral e tecido adjacente ao tumor, caracterizado histologicamente como normal de todos os pacientes. Os dados clínico-patológicos dos pacientes foram coletados retrospectivamente, através da análise dos prontuários eletrônico e físico. Os pacientes foram divididos em três grupos sendo, 14 pacientes usados para o RNA-seq, 28 usados para PCRq e 42 pacientes para sequenciamento de *TP53*.

Além das amostras de pacientes com CEE foram coletadas biópsias de esôfago de oito pacientes sem qualquer alteração nesse tecido e sem histórico de câncer. Estes pacientes foram recrutados no Ambulatório de Gastroenterologia do Hospital Universitário Pedro Ernesto - HUPE/UERJ, durante procedimento endoscópico para avaliação de possíveis patologias do trato aerodigestivo superior, e após exame, foi confirmada a ausência de problema esofágico. Os indivíduos sem doença no esôfago, assim como os pacientes com CEE, assinaram um termo de consentimento informado (Apêndice A) e responderam um questionário sobre os seus hábitos alimentares, sociais, aspectos de saúde e histórico familiar de câncer (Tabela 1; Apêndice B).

O projeto foi aprovado pelos comitês de ética do HUPE/UERJ (416-CEP/HUPE; Apêndice C) e INCA (116/11; Apêndice D).

#### **3.2. EXTRAÇÃO DE RNA**

Para a extração do RNA total, foi utilizado o RNeasy Mini kit (Qiagen), seguindo as orientações do fabricante:

- Um fragmento da biópsia de até 30 mg foi colocado em um microtubo de 1,5 mL junto com 200  $\mu$ L do tampão RLT com beta-mercaptoetanol. Após macerar o tecido, com auxílio de um pilão, foram adicionados mais 400  $\mu$ L do tampão RLT e posteriormente foram centrifugados em microcentrífuga por 3 min à velocidade de 12.000 x g para sedimentar o material não lisado. Em seguida, o sobrenadante foi transferido para um novo tubo e 1 volume de etanol 70% foi adicionado. Após

homogeneização, 700 µL da solução foram transferidas para uma coluna de purificação alocada em um tubo coletor de 2 mL e centrifugados por 15 s a 8.000 x g. Ao término, o líquido do tubo coletor foi descartado e o procedimento foi repetido até passar toda a solução pela coluna. Em seguida, 700 µL do tampão RW1 foram adicionados na coluna de purificação e centrifugados por 15 s a 8.000 x g. Ao término, o líquido do tubo coletor foi descartado e 500 µL do tampão RPE foram adicionados a coluna de purificação e, posteriormente, centrifugados por 2 min a 8.000 x g. Em seguida, a coluna foi transferida para um novo tubo coletor de 2 mL e centrifugados por 1 min a 12.000 x g. Após essa centrifugação, a coluna foi transferida para um microtubo de 1,5 mL e 30 µL de água foram adicionados no centro da membrana sem encostar a ponteira na mesma a fim de eluir o RNA total. Após 1 minuto a temperatura ambiente, os tubos foram centrifugados por mais 1 min a 8.000 x g. A coluna foi descartada e posteriormente os tubos contendo o RNA total extraído foram armazenados à -80°C.

A concentração de RNA nas amostras foi determinada a partir de 1,0 µL de RNA extraído das amostras, utilizando o espectrofotômetro NanoDrop (Uniscience) no qual a absorvância da amostra foi determinada no comprimento de onda de 260 nm. O cálculo da concentração do RNA tomou por base a informação de que uma unidade de absorvância corresponde a uma concentração de 40 µg / mL de RNA. Ainda, a relação  $A_{260}/A_{280}$  foi utilizada como um parâmetro de avaliação da qualidade das preparações do RNA, uma vez que as proteínas absorvem luz no comprimento de onda de 280 nm. . Valores inferiores a 1,7 resultam de contaminação com proteína.

Para determinarmos a integridade do RNA extraído, utilizamos o equipamento Bioanalyzer (Agilent), com o chip RNA 6000 Nano, seguindo o protocolo do fabricante. Como resultado, o Bioanalyzer fornece um número de integridade do RNA (RIN) e amostras com RIN inferior a 8 foram descartadas. A integridade do RNA foi determinada pela razão entre os rRNAs 18S e 28S e a presença de material degradado.

### **3.3. CONFECÇÃO DA BIBLIOTECA DE cDNA**

As bibliotecas de cDNA das amostras de tumor de esôfago, tecido adjacente e mucosa sadia foram realizadas seguindo o protocolo TruSeq RNA (Illumina):

#### **3.3.1. Purificação e fragmentação do RNA**

Inicialmente, 1,0 µg de RNA total de cada amostra foi diluído para volume final de 50,0 µL e transferido para placa com poços de 0,3 mL. Ao RNA foram adicionados 50,0 µL de *beads* de purificação e homogeneizado. A placa foi selada e colocada no termociclador para desnaturação inicial por 4 minutos a 65°C. Após esse período, a placa permaneceu por 5 minutos a temperatura ambiente. A placa foi colocada sobre uma estante magnética para captura das *beads* ligadas ao RNA e o sobrenadante foi descartado. A placa foi retirada da estante e foi adicionado a cada poço 200,0 µL de *Bead Washing Buffer*. A placa foi recolocada sobre a estante magnética para captura das *beads* ligadas ao RNA e o sobrenadante foi descartado. A placa foi retirada da estante e foi adicionado a cada poço 50,0 µL de *Elution Buffer*. Após selada, a placa foi colocada no termociclador para incubação por 2 minutos a 80°C. Após esse período, a placa permaneceu por 5 minutos à temperatura ambiente. A placa foi colocada sobre a estante magnética para captura das *beads* ligadas ao RNA poly-A e o sobrenadante foi descartado. Retirou-se a placa da estante magnética e foram adicionados 200,0 µL de *Bead Washing Buffer* que permaneceu por 5 minutos a temperatura ambiente. A placa foi colocada novamente na estante magnética e o sobrenadante contendo o RNA ribossomal e outros RNAs não mensageiros foi descartado. A placa foi retirada da estante e foram adicionados 19,5 µL de *Elute Prime Fragmentation Mix* a cada poço. A placa foi selada e colocada no termociclador para incubação por 5 minutos a 94°C.

### **3.3.2. Síntese da primeira fita de cDNA**

Após a reação de fragmentação a placa foi retirada do termociclador e incubada a temperatura ambiente por 5 minutos na estante magnética. O adesivo foi retirado e 17,0 µL do sobrenadante foram utilizados para a síntese do cDNA, sendo adicionados em uma nova placa. Foram adicionados 8 µL em cada poço de mix contendo a proporção *First Strand Master Mix* e *SuperScript II* (Invitrogen) na proporção de 9:1. A placa foi selada e colocada no termociclador para incubação seguindo 25°C por 10 minutos, 42°C por 50 minutos e 70°C por 15 minutos.

### **3.3.3. Síntese da segunda fita de cDNA**

A placa foi retirada do termociclador e o selo foi removido. Foram adicionados 25,0 µL de *Second Strand Master Mix* em cada poço. A placa foi novamente selada e incubada por 1 hora a 16°C. Após a incubação a placa foi removida e incubada a temperatura ambiente. O cDNA dupla-fita foi purificado utilizando as *beads*

magnéticas AMPure XP, seguindo protocolo do fabricante, sendo eluído em 52,5  $\mu$ L de *Ressuspension Buffer*.

#### **3.3.4. Reparação de final de fita de cDNA e adenilação da porção 3'**

Após a síntese da dupla-fita de cDNA foi realizada a reparação das "saliências" resultantes do processo de fragmentação do RNA. Foram adicionados 10,0  $\mu$ L de *Ressuspension Buffer* em cada poço, seguido de 40,0  $\mu$ L do *End Repair Mix*. A placa foi selada e incubada por 30 minutos a 30 °C. Após essa incubação foi realizada a purificação com *beads* magnéticas AMPure XP, seguindo protocolo do fabricante, sendo eluído em 17,5  $\mu$ L de *Ressuspension Buffer*. Após a eluição, 15,0  $\mu$ L de cada reação foram transferidos para nova placa para a realização da adenilação da porção 3' das fitas de cDNA, com adição de 2,5  $\mu$ L de *Ressuspension Buffer* e 12,5  $\mu$ L de *A-Tailing Mix* em cada poço. Após a adição dos reagentes, as reações foram incubadas por 30 minutos a 37 °C.

#### **3.3.5 Ligação dos Adaptadores**

Este processo consiste na adição na porção terminal dos cDNAs de adaptadores e índices que servirão como ancoradores a matriz da *flowcell* e marcadores de identificação nas reações de sequenciamento, respectivamente. Após a adenilação, foram adicionados a cada reação 2,5  $\mu$ L do controle de ligação diluído 100 vezes, 2,5  $\mu$ L de *Ligation Mix* e 2,5  $\mu$ L de cada um dos adaptadores selecionados. As reações foram incubadas por 10 minutos a 30 °C no termociclador. Ao fim da incubação foram adicionados 5,0  $\mu$ L de *Stop ligation buffer* em cada reação, com subsequente purificação com *beads* magnéticas AMPure XP, seguindo protocolo do fabricante, sendo eluído em 22,5  $\mu$ L de *Ressuspension Buffer*.

#### **3.3.6. Enriquecimento dos Fragmentos de DNA**

Ao término da ligação dos adaptadores 20  $\mu$ L de cada amostra foi transferido para nova placa e foram adicionados a cada poço 5,0  $\mu$ L de *PCR Primer Cocktaile* 25,0  $\mu$ L de *PCR Master Mix*. As de enriquecimento seguiram a seguinte ciclagem:

- 1 ciclo: 98°C por 30 segundos
- 15 ciclos: 98°C por 10 segundos  
60°C por 30 segundos  
72°C por 30 segundos
- 1 ciclo: 72°C por 5 minutos
- Manter 10°C até retirar as reações do termociclador

Após a reação de enriquecimento foi realizada purificação do material amplificado com *beads* magnéticas AMPure XP, seguindo protocolo do fabricante, sendo eluído em 32,5 µL de *Ressuspension Buffer*.

### 3.3.7. Validação das bibliotecas

Como recomendado pela Illumina, após a preparação das bibliotecas de cDNA foi realizada a análise de controle de qualidade de cada amostra através de eletroforese capilar, utilizando a plataforma Bioanalyzer, chip Agilent DNA-1000, observando que as bibliotecas devem apresentar fragmentos de aproximadamente 260-300 bp. As bibliotecas foram normalizadas a 10 nM em nova placa.

### 3.4. SEQUENCIAMENTO

Após o preparo da biblioteca, 3 pmol de cada amostra foi sequenciada no equipamento Illumina HiSeq 2000, seguindo as recomendações do fabricante. Em cada canaleta do sequenciador foram colocadas quatro amostras. As reações de sequenciamento foram feitas com 100 ciclos utilizando o primer senso e 100 ciclos utilizando o primer antisenso formando para cada fragmento da biblioteca produtos de 100 pb em cada ponta do fragmento. Estes produtos de 2x 100pb serão chamados doravante de *reads*.

Durante a corrida, são geradas e analisadas as imagens que fizeram a captura da fluorescência durante a síntese da *read*, a análise dessas imagens é salva nos chamados arquivos *base calling* (BCL). Os arquivos BCL, juntamente com os arquivos contendo as localizações dos clusters, arquivos LOCS, foram convertidos para arquivos FastQ utilizando o programa CASAVA (Illumina), onde ocorre simultaneamente a separação de bibliotecas que utilizaram diferentes índices. Os arquivos FastQ contém as sequências das *reads* bem como a qualidade do sequenciamento de cada base na escala *Phred*.

### 3.5. CONTROLE DE QUALIDADE

A qualidade de todas as *reads* foi avaliada utilizando-se o programa FastQC ([www.bioinformatics.babraham.ac.uk/projects/fastqc](http://www.bioinformatics.babraham.ac.uk/projects/fastqc)), que possui diferentes módulos de análises para o controle de qualidade das sequências brutas produzidas pelos equipamentos de NGS. Os gráficos gerados pelo programa permitem avaliar a qualidade geral das bibliotecas e tomar decisões importantes que afetam as análises posteriores.

Para a filtragem das *reads* ou de partes das *reads* que tiveram qualidade média abaixo de 20 na escala Phred foi utilizado o programa PrinSeq (Schmieder e Edwards 2011). Foram consideradas *reads* de boa qualidade aquelas na escala Phred igual ou maior a 20.

### 3.6. ALINHAMENTO A REFERÊNCIA

Após criterioso controle de qualidade, as *reads* foram alinhadas ao genoma de referência de *Homo sapiens*, GRCh37, disponível no site do ENSEMBL ([www.ensembl.org](http://www.ensembl.org)) com a finalidade de identificar a origem genômica de cada uma delas. O genoma de referência foi devidamente indexado como recomenda o manual do TOPHAT2, programa usado para o alinhamento das leituras ao genoma.

O TOPHAT2 é um alinhador que funciona em duas etapas. Na primeira as *reads* são alinhadas a uma referência de transcriptoma gerada a partir das anotações de genes para a versão GRCh37, pelo próprio TOPHAT2. Depois as *reads* não mapeadas são alinhadas ao genoma como mostra o esquema na figura 5 (Trapnel, Pachter e Salzberg, 2009).

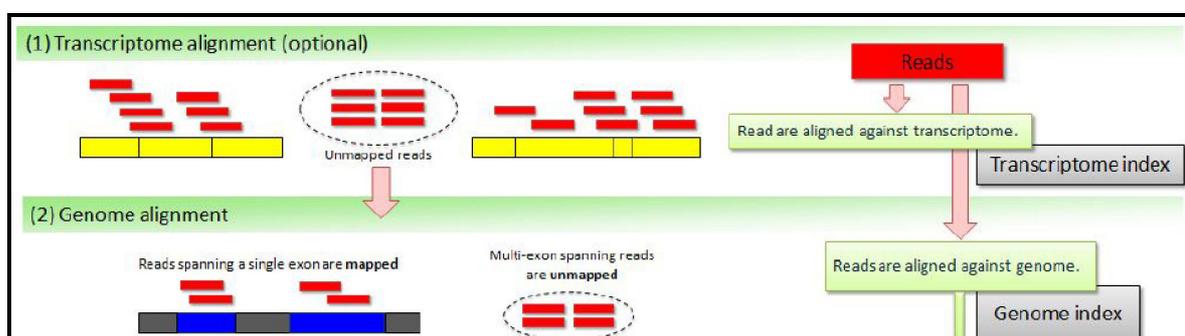


Figura 3.1 – Alinhamento em duas etapas com TOPHAT2 (Kim *et al*, 2013).

### 3.7. ANÁLISE DE EXPRESSÃO GÊNICA DIFERENCIAL

Logo após os alinhamentos serem feitos, os arquivos no formato *Sequence Alignment/Map* (SAM), contendo as coordenadas de alinhamento de cada *read*

foram filtrados para a retirada das *reads* que não tiveram alinhamento único. Após essa etapa, as *reads* foram ordenadas de forma que a primeira leitura no arquivo seja a leitura que alinhou na porção 5' do cromossomo 1 seguindo essa ordem até a região 3' do cromossomo Y. Para ambos os procedimentos foi usado o *software* SAMTools (Li *et al.*, 2009).

Em seguida as contagens das *reads* por gene foram feitas com o *software* HTSeq seguindo as instruções do manual (Anders, Pyl e Huber, 2014). Os arquivos com as contagens foram então carregados no *software* R e fundidos em uma matriz cujas colunas correspondiam as amostras e as linhas aos genes.

No ambiente R/Bioconductor, a matriz com as contagens das *reads* serviu de base para a análise de expressão diferencial entre 2 comparações: mucosa normal adjacente ao tumor contra mucosa de indivíduos sem câncer (NxH) e tecido tumoral contra mucosa adjacente ao tumor (TxN).

O pacote usado para análise dessas comparações foi o DESeq2 (Love, Huber e Anders, 2014) e os critérios para determinação de um gene como diferencialmente expresso (GDE) foram *logFold-change* absoluto maior que 1 e valor de p-ajustado (FDR) menor que 0,05.

### **3.8. ANÁLISE DOS GENES DIFERENCIALMENTE EXPRESSOS**

Uma vez que a lista de genes diferencialmente expressos (GDEs) foi determinada, as mesmas foram submetidas à análise de enriquecimento de processos biológicos, utilizando como base de dados o *Gene Ontology* e enriquecimento de vias utilizando o banco do *KEGG*. Depois do enriquecimento pelo *Gene Ontology* os processos biológicos foram submetidos à análise com o *software online* REVIGO.

Ambas as análises de enriquecimento e agrupamento foram executadas no R/Bioconductor.

### **3.9. ANÁLISE DAS VARIANTES DE BASE ÚNICA**

Para a análise das variantes de base única (SNV, do inglês, *Single Nucleotide Variant*) foi utilizado o protocolo desenvolvido pelo *Broad institute* baseado no uso do *Genome Analyses Toolkit* (GATK) e disponível no site <https://www.broadinstitute.org/gatk/guide/best-practices>. A partir dos arquivos FASTQ filtrados pela qualidade foi feito o mapeamento das *reads* usando o alinhador STAR (Dobin *et al.*, 2012).

Após o alinhamento foram retiradas as *reads* advindas provavelmente de duplicação por PCR e as *reads* cobrindo junções de exons tiveram a qualidade de mapeamento corrigida. Esse passo é estritamente técnico para manter a compatibilidade entre o STAR e o GATK.

A partir deste ponto foi feito o realinhamento de INDELs (inserções e deleções) e a recalibração das variações encontradas utilizando INDELs e SNPs conhecidos presentes no site <https://genome.ucsc.edu> sempre utilizando como referência o genoma de *H. sapiens*, versão hg19.

Finalmente foi realizado o *variant calling* seguido da aplicação de filtros para selecionar as variantes de maior confiabilidade. Os filtros aplicados foram: *Fisher Strand* < 30, *Quality by Depth* > 2 e *Depth* > 10.

A anotação das variações encontradas foi feita utilizando o pacote ANNOVAR (Wang, Li, e Hakonarson, 2010).

Uma vez que foram determinados os SNVs do CEE, mucosa normal adjacente e mucosa esofágica sadia aqueles SNVs presentes na mucosa adjacente e mucosa sadia foram excluídos da análise com a intenção de estudar apenas SNVs somáticas. Com a mesma finalidade também foram excluídas SNVs presentes em bancos públicos de SNP: 1000 genomes, Watson, Venter e YH.

### **3.10. ANÁLISE DE TRANSCRITOS QUIMÉRICOS**

Para análise de transcritos quiméricos ou fusionados foi utilizado o programa FusionMap (Ge *et al.*, 2011). A partir dos arquivos SAM contendo as coordenadas de mapeamento das *reads* alinhadas ao genoma de referência foram aplicados os seguintes parâmetros: *Sample.SeedCount* >3, *SplicePatternClass* = *CanonicalPattern[Major] or CanonicalPattern[Minor]*, *Filter* = vazio, *FrameShiftClass* = *InFrame* e *OnExonBoundary* = *Both*.

### **3.11. VALIDAÇÃO DA EXPRESSÃO GÊNICA POR PCR EM TEMPO REAL (PCRq)**

Para o ensaio de PCR em tempo real (PCRq) o RNA das amostras foi extraído como descrito anteriormente no item 3.2

### 3.11.1 Síntese de cDNA

O RNA total extraído das amostras foi reversamente transcrito em cDNA para análises de PCR em tempo real utilizando a transcriptase reversa *SuperScript II@* (Invitrogen), conforme descrito abaixo:

– Em um microtubo foram adicionados 250 ng de *primer* randômico (Promega), 500 ng de RNA alvo, 1  $\mu$ L do set de dNTPs a 10mM e água RNase *FREE* suficiente para completar o volume final de 20  $\mu$ L de reação, aquecer por 5 minutos à 65°C. Em seguida, foram adicionados 4  $\mu$ L do tampão da enzima transcriptase reversa (5X) fornecido pelo fabricante (Invitrogen) e 2  $\mu$ L de DTT 0,1M, incubar por 2 min à 25°C. Em seguida, foi adicionado 1  $\mu$ L de *SuperScript II@* contendo 200 U e as reações foram incubadas no termociclador por 10 min à 25°C, 50 min à 42°C e inativar a enzima a 70°C por 15 min. Após a síntese, o cDNA foi armazenado à -20°C

### 3.11.2 PCR em tempo real (PCRq).

As reações de PCR em tempo real (PCRq) foram realizadas na plataforma Rotor-gene (Qiagen). Em cada reação foi adicionado 7,5  $\mu$ L de tampão *SYBR green* FAST 2x (Qiagen), 0,3  $\mu$ L de cada um dos oligonucleotídeos específicos da reação à 10  $\mu$ M, 1,0  $\mu$ L de cDNA diluídos 1:20 e 5,5  $\mu$ L de água deionizada autoclavada. As reações ocorreram com a seguinte ciclagem:

40 ciclos  $\left\{ \begin{array}{l} 95^{\circ}\text{C por 5 min} \\ 95^{\circ}\text{C por 5 seg;} \\ 60^{\circ}\text{C por 10 seg} \end{array} \right.$

Para a quantificação dos resultados e determinação da eficiência dos oligonucleotídeos iniciadores específicos para cada gene foi estabelecido uma curva padrão de expressão utilizando uma diluição seriada de cDNA de um controle positivo da reação.

**Quadro 3.1 – Lista de oligonucleotídeos iniciadores para validação da expressão dos genes por PCRq.**

<b>Genes</b>	<b>Oligonucleotídeos</b>
<b><i>IL8</i></b>	<b>F: 5' CAG AGA CAG CAG AGC ACA CA 3'</b>
	<b>R: 5' GGC AAA ACT GC ACCT TCA CA 3'</b>
<b><i>WNK2</i></b>	<b>F: 5' GTG GGG CTG TGT TCC TAT CC 3'</b>
	<b>R: 5' ATA GGG CGC TGA ATC TGG AG 3'</b>
<b><i>MAPK3</i></b>	<b>F: 5' ATG GTC AGC TCG GCC TAT 3'</b>
	<b>R: 5' GCT GGC AGT AGG TCT GAT GTT 3'</b>
<b><i>WNT7B</i></b>	<b>F: 5' GGC TAC TAC AAC CAA GCC GA 3'</b>
	<b>R: 5' CCT CTA GAA CCT TCC TGC CG 3'</b>
<b><i>WNT16</i></b>	<b>F: 5' ACA CCA CGG GCA AAG AAA AC 3'</b>
	<b>R: 5' ACT TGG CGA CAG CCT GC 3'</b>
<b><i>PRKCQ</i></b>	<b>F: 5' GGC CGA AAG TGA AAT CAC CA 3'</b>
	<b>R: 5' CAG TGC TCT GTC GGC AAA TG 3'</b>
<b><i>FZD6</i></b>	<b>F: 5' CAG CGG CTT GTA TCT TGT GC 3'</b>
	<b>R: 5' CGA GCT TTT GCT TTT GCC TGA 3'</b>

F – Forward, R - Reverse

Após a amplificação dos produtos foi feita a análise da curva de dissociação (curva de *Melt*). Essa etapa é importante para verificar se os produtos observados na PCRq são específicos e para verificar se há formação de estruturas secundárias entre os oligonucleotídeos e contaminação na reação

A quantificação relativa de cada gene contra o controle interno, GAPDH, foi possível pelo método de  $\Delta Cq$ . Um gráfico de amplificação, cujo eixo Y era o sinal de fluorescência e o eixo X era o número de ciclos, foi desenhado. No primeiro ciclo da reação de PCR houve pouca mudança no sinal de fluorescência. Isso definiu a linha de base para o gráfico de amplificação. Um nível de fluorescência acima da linha de base indicou o acúmulo de produto de PCR. Foi determinado, então, um limiar de detecção de fluorescência fixo acima da linha de base durante a fase exponencial da PCR. O parâmetro Cq (limiar) foi definido como o número de ciclos, inteiro ou não, no qual a fluorescência ultrapassou os limiares previamente fixados. A diferença entre as médias (de três experimentos) do gene de interesse e do gene de referência foi calculada pelo programa Microsoft Excel e o valor de quantificação relativa foi expresso como  $2^{-\Delta Cq}$ . (Livak e Schmittgen, 2001).

### **3.12. AVALIAÇÃO DO PERFIL MUTACIONAL DO GENE TP53 EM CEE**

A reação de sequenciamento procedeu como descrito no item 3.4, mas antes algumas etapas foram necessárias e estão descritas abaixo.

### **3.12.1. Extração do DNA**

A extração foi realizada como descrito pelo fabricante do DNEasy Blood and Tissue kit (QIAGEN®). As biópsias de CEE foram maceradas com pilão dentro de um microtubo de 1,5 mL contendo 150 µL de tampão ATL. Após esse processo, foram adicionados mais 150 µL de tampão ATL e a proteinase K, com posterior incubação a 56°C em agitação constante por 2 horas no equipamento *Thermomixer Comfort* (EPPENDORF®).

A inativação da proteinase K foi realizada pela manutenção das amostras a 90°C em agitação constante durante 10 minutos. As amostras foram então centrifugadas a 14000 rpm por 30 segundos e foi adicionado 200 µL de tampão de ligação e 200 µL de etanol absoluto. Após a homogeneização das soluções, este conteúdo foi então transferido para a uma coluna com tubo coletor (contidos no kit) e centrifugado a 14000 rpm por 1 minuto. A solução coletada no tubo foi descartada. Adicionou-se em seguida 500 µL de tampão de lavagem a coluna e esta foi centrifugada a 14000 rpm por 1 minuto. A solução coletada no tubo foi descartada. Foi adicionado em seguida 500 µL de um segundo tampão de lavagem a coluna e esta foi centrifugada a 14000 rpm por 1 minuto. A solução coletada no tubo foi descartada. Foi realizada uma centrifugação a 14000 rpm por 3 minutos para a retirada de qualquer resíduo do tampão. O tubo coletor foi então descartado e substituído por um microtubo de 1,5 mL. Em seguida, foi acrescentado ao centro da coluna 30 µL de tampão de eluição (contido no ki) que foi posteriormente centrifugado a 14000 rpm por 1 minuto.

### **3.12.2. Quantificação de DNA**

Uma alíquota do material extraído (1 µL) foi quantificada através da leitura de absorvância em espectrofotômetro Nanodrop (Thermo) utilizando comprimento de onda igual a 260 nm. A pureza do DNA foi verificada a partir da leitura a 280 nm, para a detecção de uma eventual contaminação da amostra por proteínas. Quando a relação das densidades óticas, A260/A280, foi igual ou maior que 1,7 o material foi considerado puro.

### **3.12.3. PCR**

O gene *TP53* foi amplificado através da reação da polimerase em cadeia (PCR) utilizando 3 pares de oligonucleotídeos iniciadores, compreendendo os exons 2 ao 11 do gene. Estes oligonucleotídeos estão listados no quadro 3.1. Em cada

reação foram adicionados 2,5 µL do tampão 10X, 0,8 µL de MgCl<sub>2</sub>, 2,5 µL do mix dos três pares de oligos, 0,25 µL de dNTP, 1 unidade de Taq Platinum (Invitrogen), 50 ng de DNA e água suficiente para completar volume final de 25,0 µL.

**Quadro 3.2 – Lista de oligonucleotídeos iniciadores para amplificação do gene *TP53*.**

Região	Oligonucleotídeos	Tamanho
Exons 2-4	F - 5' CCAGGTGACCCAGGGTTGGA 3'	1781 pb
	R - 5' AGCATCAAATCATCCATTGC 3'	
Exons 5-9	F - 5' TGTTCACCTTGTGCCCTGACT 3'	1215 pb
	R - 5' ACTTGATAAGAGGTCCCAAG 3'	
Exons 10-11	F - 5' CAATTGTAACCTGAACCATC 3'	778 pb
	R - 5' AGGCTGTCAGTGGGGAACAA 3'	

F: *forward*; R: *reverse*; pb: pares de bases

### 3.12.3. Purificação do produto de PCR

Os produtos de PCR foram purificados com o kit PureLink™ Genomic DNA Purification de acordo com o protocolo do fabricante (INVITROGEN, USA). A cada micro tubo de capacidade de 1,5ml foram adicionados 500 µL de tampão de captura a 100 µL de produto de PCR. Esta solução foi então transferida a uma coluna Microspin com tubo coletor e centrifugada a 14000 rpm por 30 segundos. A solução coletada no tubo foi descartada. Foi adicionado em seguida 500 µL de tampão de lavagem a coluna e esta foi centrifugada a 14000 rpm por 30 segundos. A solução coletada no tubo foi descartada. Foi realizada uma centrifugação a 14000 rpm por 3 minutos para a retirada de quaisquer resíduos do tampão. O tubo coletor foi então descartado e substituído por um micro tubo de capacidade de 1,5 mL. Em seguida, foram adicionados ao centro da coluna 25 µL de tampão de eluição e esta etapa apresentou duração de 5 minutos a temperatura ambiente. Este material foi então centrifugado a 14000 rpm por 2 minuto e 2 µL do DNA purificado foram submetidos a uma corrida em gel de agarose 2 % para avaliação da amostra purificada. Por fim, este material foi armazenado a -20°C para posterior reação de sequenciamento.

As bibliotecas foram preparadas com o kit *Nextera XT DNA sample Preparation* (Illumina), seguindo o protocolo do fabricante. Na primeira etapa ocorre simultaneamente a fragmentação do DNA em regiões aleatórias e sua marcação com etiquetas (*tags*) específicas, que são sequencias terminais de 14 pb, através da ação de um *transposome* (complexo *transposon-transposase*). Para isso, foram adicionados 10 µL do tampão TD (*Tagment DNA Buffer*), 5 µL do produto de PCR

purificado a 0,4 mg / $\mu$ L e 5  $\mu$ L de ATM (*Amplicon Tagment Mix*) em cada reação e em seguida foi incubada a 55°C por 5 minutos. Em seguida, foi adicionado 5  $\mu$ L do tampão NT (*Neutralize Tagment Buffer*) e a reação foi incubada a temperatura ambiente por 5 minutos, para neutralizar a ação do *transposome*.

Na segunda etapa, o DNA fragmentado é enriquecido através de uma reação de PCR, que utiliza como iniciadores sequências contendo uma região 3' complementar as etiquetas (*tags*), além de uma sequência identificadora (*indices/barcodes*; i5 e i7) e ainda na região 5' encontra-se uma sequência chamada de adaptador (P5 e P7), que são complementares aos adaptadores da etapa de clusterização.

Para a reação de amplificação, foram adicionados 5  $\mu$ L de cada um dos índices (i5 e i7) e 15  $\mu$ L de NPM (*Nextera PCR Master Mix*) ao produto da neutralização. Em seguida a reação foi submetida a uma ciclagem de 72°C por 3 minutos, 95°C por 30 segundos e 12 ciclos de 95°C por 10 segundos, 55°C por 30 segundos e 72°C por 30 segundos, além de uma extensão final de 72°C por 5 minutos no termociclador *Veriti 96 well Thermal Cycler* (Applied Biosystems).

A purificação das bibliotecas foi realizada utilizando microesferas magnéticas, que além de purificar a biblioteca, também promovem uma seleção de tamanho dos fragmentos, retirando fragmentos de tamanhos inadequados. Para isto, em uma placa de 96 poços foram adicionados 30  $\mu$ L de *AMPure XP* (*Agencourt Bioscience Corporation*) em cada reação, misturando-se até gerar uma solução homogênea. Após 5 minutos de incubação na temperatura ambiente, a placa foi colocada sobre uma placa magnética por 2 minutos. Neste passo, as microesferas ligadas ao DNA foram precipitadas na parede de cada poço da placa, permitindo a retirada do sobrenadante. Em seguida, foram feitos dois processos de lavagem, adicionando 200  $\mu$ L de etanol 80% (Merck) e retirando o sobrenadante, com posterior incubação de 15 minutos ainda sobre a placa magnética a temperatura ambiente para evaporação de resquícios de etanol. Após a este processo as microesferas ligadas ao DNA foram ressuspensas em 32,5  $\mu$ L de tampão RSB (*Resuspension Buffer*). Em seguida, com a placa sobre a placa magnética, as microesferas foram precipitadas na parede do poço liberando o DNA no tampão RSB. O produto final, a biblioteca, foi analisado quanto a distribuição de tamanho dos fragmentos através de eletroforese de 3  $\mu$ L do produto final em gel de agarose 0,8%. Ao final, foi observado o tamanho médio dos fragmentos da biblioteca, comparando com o marcador de peso molecular e a biblioteca foi armazenada a -20°C.

### 3.12.4. Quantificação das bibliotecas

As bibliotecas foram quantificadas utilizando a técnica de quantificação absoluta por PCR em tempo real no sistema ECO (Illumina) com o kit *KAPA Library Quantification* (KAPA Biosystems), seguindo o protocolo do fabricante. Inicialmente as bibliotecas foram diluídas 10.000 vezes em solução 0,1% Tween20 + 10 mM Tris-HCl a pH 8,5. Foram utilizados seis padrões fornecidos com quantificações conhecidas (20 pM, 2 pM, 0,2 pM, 0,02 pM, 0,002 pM e 0,0002 pM), além de controle negativo. Em cada reação foram adicionados 6 µL de Master Mix e 2 µL de água ultrapura e, ao final, foram adicionados 2 µL da biblioteca, ou padrão de concentração ou solução de diluição para controle negativo. Todas as reações foram realizadas em triplicatas e foram submetidas a seguinte ciclagem:

- 1ª etapa: 1 X 95°C por 5 min
- 2ª etapa: 35X 95°C por 30 seg  
60°C por 45 seg
- 3ª etapa: 1 X 95°C por 15 seg  
55°C por 15 seg  
95°C por 15 seg

A concentração das bibliotecas foi estimada através da equação da reta  $f(x) = -b \ln(x) + a$ , gerada a partir dos ciclos de quantificação (Cq) dos padrões, sendo o  $x$  correspondente à média do Cq de cada biblioteca. A concentração final foi normalizada pelo tamanho médio das bibliotecas, que foi observado na eletroforese, e pela taxa de diluição de 10.000 vezes.

### 3.12.5. Análise *in silico*

Uma vez sequenciado as reads foram filtradas por qualidade como descrito no item 3.5. As sequencias contidas nos arquivos FASTQ foram alinhadas ao genoma humano utilizando o alinhador BWA (Li, 2013). O *Variant calling* e a anotação foram feitas como descrito acima.

### **3.13. ANÁLISE ESTATÍSTICA DOS DADOS HISTOPATOLÓGICOS E EPIDEMIOLÓGICOS**

As análises relacionando as alterações moleculares com os dados clínicos dos pacientes foram feitas utilizando o R. Para a construção das curvas de Kaplan-Meier bem como os testes de log rank e regressão de Cox foi usado o pacote survival. As correlações lineares, testes de Fisher, Kolmogorov-Smirnov, t, Wilcoxon e Mann-Whitney foram feitos utilizando o pacote básico do R.

Para todas as análises utilizou-se  $p < 0,05$  como critério para determinar valores estatisticamente significativos.

## 4. RESULTADOS

### 4.1. LEVANTAMENTO DE DADOS DOS PRONTUÁRIOS

Os dados clínico-patológicos e demográficos dos pacientes que fizeram parte desse estudo foram obtidos no arquivo de prontuários do INCA. A tabela 4.1 mostra o resultado deste levantamento de dados contendo as variáveis consideradas relevantes para o presente trabalho. Infelizmente muitos dados foram perdidos e classificados como “Não Disponível” (ND).

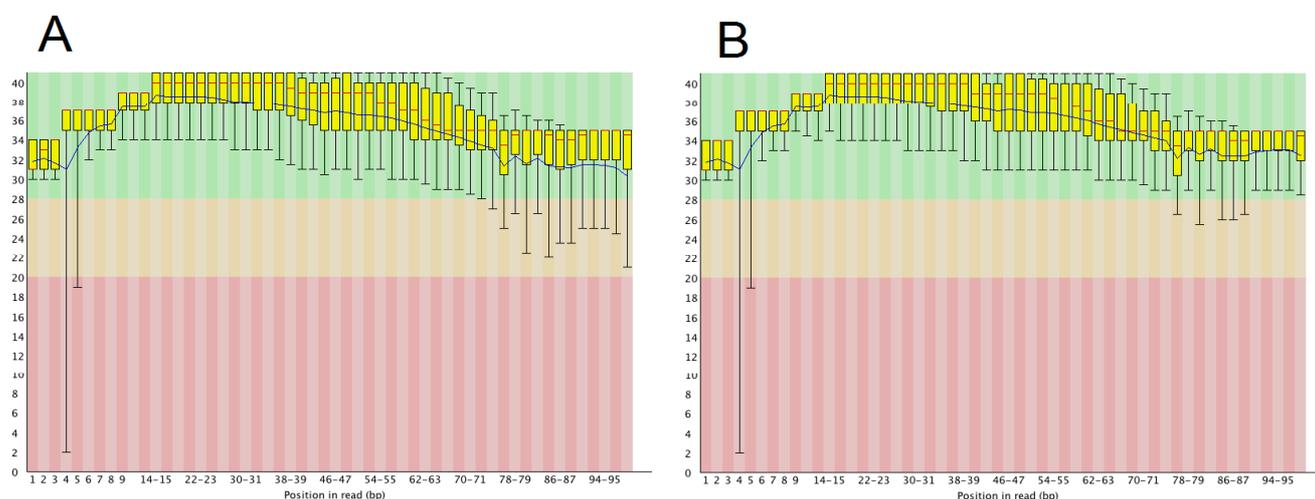
**Tabela 4.1 - Dados clínico-patológicos dos pacientes com carcinoma epidermóide de esôfago.**

	Pacientes RNA-seq	Pacientes Validações
<b>Gênero</b>		
Homem	10 (71,4%)	41 (74,5%)
Mulher	4 (28,6%)	14 (25,5%)
<b>Idade</b>		
	60 (45 – 78)	59 (39 – 79)
<b>Tabagismo</b>		
Sim	7 (50,0%)	16 (29,1%)
Não	1 (7,1%)	1 (1,8%)
Ex	1 (7,1%)	7 (12,7%)
ND	5 (35,7%)	31 (56,4%)
<b>Etilismo</b>		
Sim	5 (35,7%)	14 (25,5%)
Não	1 (7,1%)	2 (3,6%)
Ex	2 (14,3%)	8 (14,5%)
ND	6 (42,9%)	31 (56,4%)
<b>Grau de Diferenciação</b>		
Bem	1 (7,1%)	1 (1,8%)
Moderadamente	8 (57,1%)	36 (65,5%)
Pouco	5 (35,7%)	13 (35,7%)
<b>Estadiamento</b>		
1 e 2	5 (35,7%)	6 (10,9%)
3 e 4	5 (35,7%)	24 (43,6%)
ND	4 (28,6%)	25 (45,5%)
<b>Mediana da sobrevida global em meses (Mín – Máx)</b>		
	11,8 (1,9 – 41,7)	11,1 (0,2 – 41,7)

## 4.2. ANÁLISE DA QUALIDADE DAS LEITURAS

A etapa inicial da geração de resultados a partir do sequenciamento de RNA consistiu de uma análise da qualidade por base sequenciada nas 36 bibliotecas analisadas. A figura 4.1 mostra a comparação das qualidades das leituras das bases, antes e depois da aplicação do filtro de qualidade (*Prinseq*), em uma biblioteca representativa. A distribuição das qualidades está ilustrada em um gráfico em escala *Phred* que estima o erro no sequenciamento.

De um modo geral foi possível avaliar que a maioria das *reads* teve boa qualidade de sequenciamento já que as distribuições ficaram acima de 20 na escala *Phred*, que representa 1 erro a cada 100 bases sequenciadas, isto é, uma quantidade de erro aceitável para não interferir com as análises futuras (Figura 4.1A).



**Figura 4.1 - Avaliação de qualidade, na escala *Phred*, de uma biblioteca representativa antes (A) e depois do uso do filtro de qualidade (*Prinseq*).**

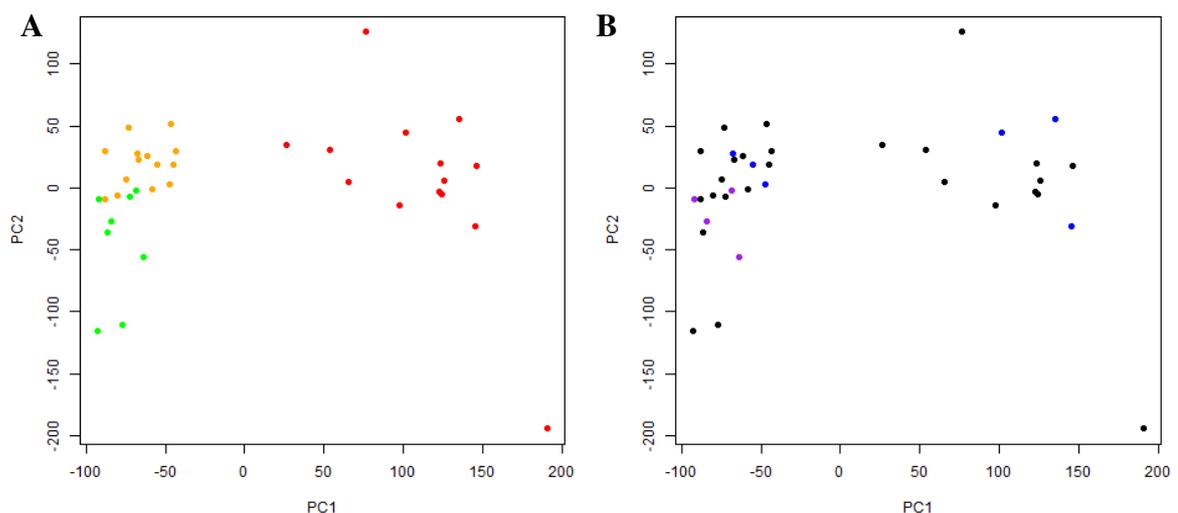
## 4.3. MÉTRICAS DE LEITURA E DE MAPEAMENTO NO GENOMA

Após a filtragem de qualidade, as *reads* foram alinhadas no genoma de referência utilizando o programa TOPHAT2, que é um algoritmo capaz de reconhecer junções exon-exon sendo, portanto ideal para remontar os transcritos presentes nas bibliotecas. Em média cada biblioteca teve 87,57% das *reads* mapeadas com um desvio-padrão de 5,06%.

#### 4.4. EXPRESSÃO GÊNICA

A fim de determinar diferenças nas abundâncias dos RNAs mensageiros, nas condições estudadas, foi utilizado o pacote DESeq2 no ambiente R/Bioconductor. Das três condições estudadas, duas comparações foram feitas, quais sejam, tumor contra mucosa adjacente ao tumor (TxN) e mucosa adjacente ao tumor contra mucosa de indivíduos sem doença no esôfago (NxH).

A fim de ter um panorama de como as amostras se relacionavam entre si e se estariam sujeitas ao *batch effect* foi feita uma análise de componente principal (PCA) visto na figura 4.2. Nela podemos observar que as amostras se agrupam em função do laudo histopatológico (figura 4.2A) e não em função do compartilhamento da corrida de sequenciamento (figura 4.2B). Além disso, podemos observar que a mucosa adjacente ao tumor e mucosa sadia são muito mais parecidas entre si e bem diferentes dos tumores. O espalhamento dos tumores no gráfico denota uma pronunciada heterogeneidade tumoral.



**Figura 4.2 – Análise de componente principal mostrando a distribuição das amostras de acordo com suas maiores diferenças. PC1 corresponde a 20% da variância e PC2 6%. Vermelho – amostras tumorais, Laranja – mucosa adjacente ao Tumor, Verde – mucosa de indivíduos sem doença no esôfago, Preto – *batch* 1, Roxo – *batch* 2 e Azul – *batch* 3.**

Para a comparação NxH foram classificados como diferencialmente expressos 522 RNAm. A tabela 4.2 mostra os cinco RNAm anotados mais expressos na mucosa adjacente ao tumor e os cinco mais expressos na mucosa de indivíduos saudáveis.

**Tabela 4.2 - Lista dos 5 genes mais sobre-expressos e dos 5 mais sub-expressos em NxH**

Ensembl_ID	$\log_2\text{FoldChange}$	p ajustad o	Símbolo
Sobre-expressos			
ENSG00000175084	5,99	$1,2 \times 10^{-9}$	<i>DES</i>
ENSG00000125999	4,48	$1,9 \times 10^{-2}$	<i>BPIFB1</i>
ENSG00000163017	4,16	$2,3 \times 10^{-3}$	<i>ACTG2</i>
ENSG00000188089	4,12	$2,4 \times 10^{-6}$	<i>PLA2G4E</i>
ENSG00000215182	3,96	$3,8 \times 10^{-2}$	<i>MUC5AC</i>
Sub-expressos			
ENSG00000238741	-3,18	$5,0 \times 10^{-5}$	<i>SCARNA7</i>
ENSG00000125740	-3,06	$2,0 \times 10^{-3}$	<i>FOSB</i>
ENSG00000173662	-2,96	$4,5 \times 10^{-2}$	<i>TASIR1</i>
ENSG00000160200	-2,94	$2,3 \times 10^{-3}$	<i>CBS</i>
ENSG00000242265	-2,53	$1,2 \times 10^{-2}$	<i>PEG10</i>

A interpretação de um número extenso de genes alterados representa uma grande barreira para as etapas seguintes a experimentos de larga escala como o RNA-seq. Com o intuito de facilitar a compreensão das listas e atribuir um sentido biológico aos dados obtidos foi realizada uma análise de enriquecimento de grupos gênicos (GSEA, do inglês, *Gene set enrichment analysis*).

A GSEA pode ser entendida da seguinte forma: dos pouco mais de 30.000 genes anotados no genoma humano, 10% desses estão envolvidos em uma mesma via hipotética. Dito isto, em uma lista de 2.000 genes proveniente de um experimento em larga escala como o RNA-seq, é esperado ao acaso que 200 genes dessa lista estejam envolvidos nessa via. Contudo, se ao invés de 200 genes forem observados 600 genes pertencentes a essa via diz-se que esta via está enriquecida ou sobre-representada.

Como já mencionado na seção 3.9, para o presente trabalho foram utilizados dois bancos de anotação de relação gênica, o Gene Ontology (GO) e o *Kyoto Encyclopedia of Genes and Genomes* (KEGG). Na lista de GDE da comparação NxH nenhuma via do *KEGG* foi identificada como enriquecida utilizando como critério um

valor de p menor que 0,001. Porém, utilizando o banco de anotação GO, foram encontrados 29 processos biológicos enriquecidos e após o resumo feito pelo *REVIGO* restaram 13 processos que estão descritos na tabela 4.3. *REVIGO* é uma sigla para a ferramenta online chamada *Reduce plus Visualize Gene Ontology*. Esta ferramenta permite o resumo dos processos enriquecidos no GO através da retirada de processos redundantes como, por exemplo, se o enriquecimento mostrar *cell adhesion* e *negative regulation of cell adhesion* apenas a primeira será mostrada pelo *REVIGO* já que a segunda está incluída na primeira.

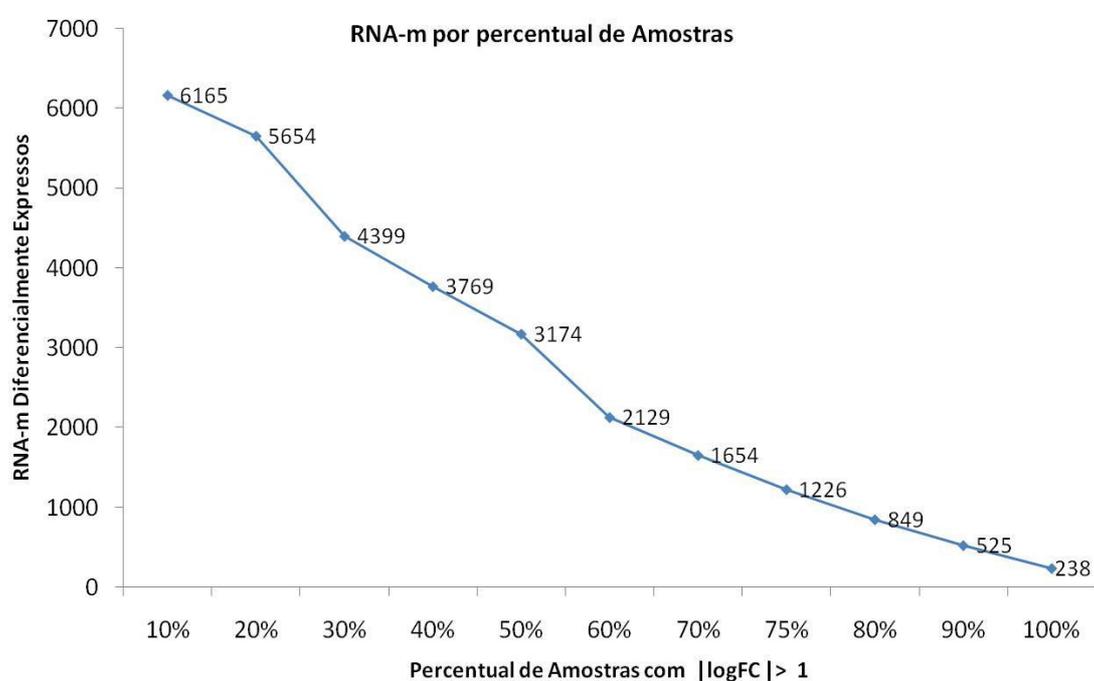
**Tabela 4.3 - Processos biológicos enriquecidos com anotação do GO, resumidos pelo *REVIGO*, dentre os GDE de NxH.**

Termo no GO	p-ajustado	N° de genes observados
single-multicellular organism process	2,99x10 <sup>-6</sup>	103
immune response	1,79x10 <sup>-5</sup>	35
regulation of leukocyte chemotaxis	2,12x10 <sup>-5</sup>	7
ear development	3,47x10 <sup>-5</sup>	11
leukocyte proliferation	3,74x10 <sup>-5</sup>	12
cell adhesion	8,36x10 <sup>-5</sup>	28
positive regulation of defense response	2,09x10 <sup>-4</sup>	12
embryonic organ morphogenesis	3,91x10 <sup>-4</sup>	9
cellular component movement	4,54x10 <sup>-4</sup>	35
L-cysteine catabolic process	4,88x10 <sup>-4</sup>	2
negative regulation of interleukin-6 secretion	4,88x10 <sup>-4</sup>	2
regulation of dendritic cell dendrite assembly	4,88x10 <sup>-4</sup>	2
negative regulation of monocyte differentiation	9,68x10 <sup>-4</sup>	2

Os processos biológicos que se destacam na comparação NxH estão em sua maioria ligados a resposta inflamatória, demonstrando mais uma vez a importância deste processo para a carcinogênese esofágica. Assim, podemos observar que das 13 vias identificadas pelo *REVIGO*, oito (61%) estavam diretamente ou indiretamente ligadas a processos inflamatórios (*immune response*, *regulation of leukocyte chemotaxis*, *leukocyte proliferation*, *positive regulation of defense response*, *cellular component movement*, *negative regulation of interleukin-6*

*secretion, regulation of dendritic cell dendrite assembly e negative regulation of monocyte differentiation*), conforme demonstrado na tabela 4.3.

Na comparação TxN foram classificados como diferencialmente expressos 9.886 RNAm. A estratégia empregada pelo *DESeq2* para testar a igualdade entre os dois grupos, normal adjacente e tumor, é sensível a *outliers* e, portanto, esses devem ser eliminados conforme descrito recentemente por Love, Huber e Anders (2014). Desta forma, para ser incluído na nova análise, um transcrito foi considerado diferencialmente expresso quando sua abundância no tumor foi maior que o dobro ou menor que a metade em relação a sua abundância na mucosa adjacente em pelo menos 75% das amostras. Sendo assim, após o uso desse critério de seleção, foram classificados 1.226 RNAm como mostra a figura 4.3.



**Figura 4.3 – Seleção dos GDE baseada no percentual de pares de amostras que apresentaram  $|\log FC|$  maior que 1.**

A tabela 4.4 mostra os 5 RNAm mais sobre-expressos e os 5 mais sub-expressos encontrados nessa análise.

**Tabela 4.4 - Lista dos 5 genes mais sobre-expressos e dos 5 mais sub-expressos em TxN.**

Ensembl_ID	log <sub>2</sub> FoldChange	P ajustado	Símbolo
<b>Sobre-expressos</b>			
ENSG00000139800	8,49	2,3x10 <sup>-17</sup>	ZIC5
ENSG00000128714	8,47	3,0x10 <sup>-21</sup>	HOXD13
ENSG00000173157	7,89	4,7x10 <sup>-13</sup>	ADAMTS20
ENSG00000149968	7,79	1,4x10 <sup>-43</sup>	MMP3
ENSG00000185686	7,71	3,7x10 <sup>-33</sup>	PRAME
<b>Sub-expressos</b>			
ENSG00000188257	-5,46	3,8x10 <sup>-16</sup>	PLA2G2A
ENSG00000261127	-4,95	2,2x10 <sup>-12</sup>	LOC390705
ENSG00000096006	-4,92	1,7x10 <sup>-15</sup>	CRISP3
ENSG00000168631	-4,52	1,7x10 <sup>-14</sup>	DPCR1
ENSG00000170477	-4,46	1,3x10 <sup>-13</sup>	KRT4

GSEA desta lista revelou um grande número de vias enriquecidas em ambos os bancos. As tabelas 4.5 e 4.6 mostram as 10 processos com menor valor de p ajustado no GO (com pós-análise pelo *REVIGO*) e 30 vias no KEGG, respectivamente.

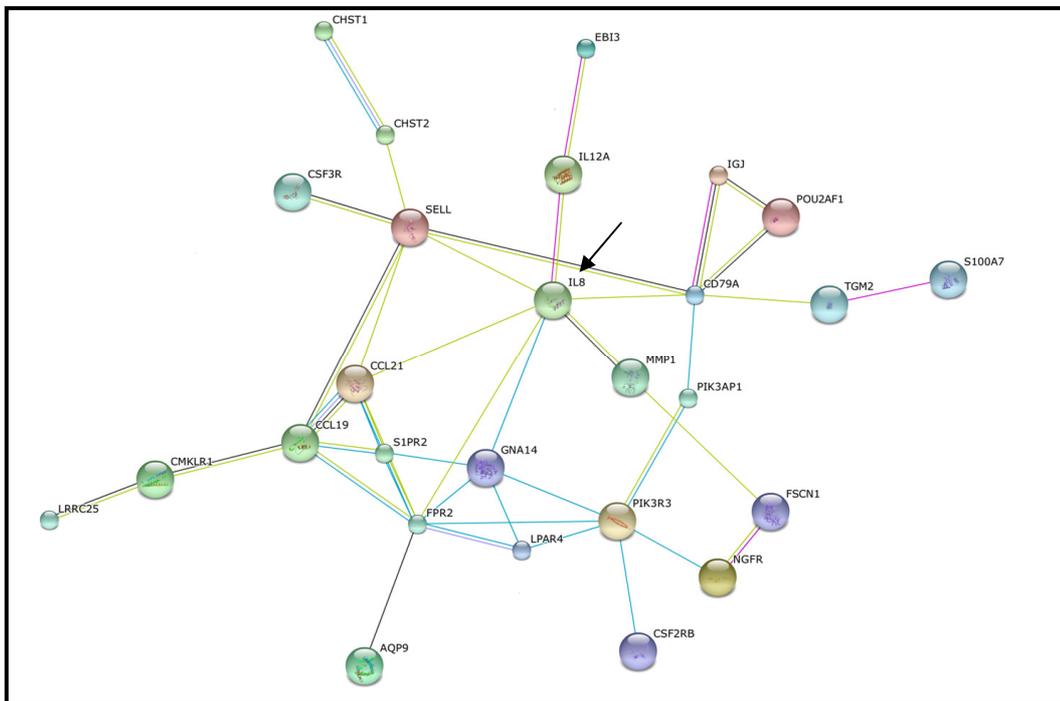
**Tabela 4.5 - Processos biológicos enriquecidos com anotação do GO (com pós-análise pelo *REVIGO*) dentre os GDE de TxN.**

Termo no GO	p-ajustado	Nº de genes observados
<i>collagen catabolic process</i>	1,1 x10 <sup>-16</sup>	30
<i>cell adhesion</i>	1,7 x10 <sup>-16</sup>	64
<i>response to external stimulus</i>	1,2 x10 <sup>-15</sup>	191
<i>extracellular matrix organization</i>	4,3 x10 <sup>-15</sup>	49
<i>Ossification</i>	5,2 x10 <sup>-11</sup>	53
<i>Chemotaxis</i>	7,9 x10 <sup>-11</sup>	63
<i>cell-cell signaling</i>	4,0 x10 <sup>-9</sup>	117
<i>Locomotion</i>	1,3 x10 <sup>-8</sup>	46
<i>regulation of biological quality</i>	3,8 x10 <sup>-8</sup>	258
<i>positive regulation of cell proliferation</i>	1,8 x10 <sup>-7</sup>	86

**Tabela 4.6 - Vias enriquecidas com anotação do KEGG dentre os GDEs de TxN.**

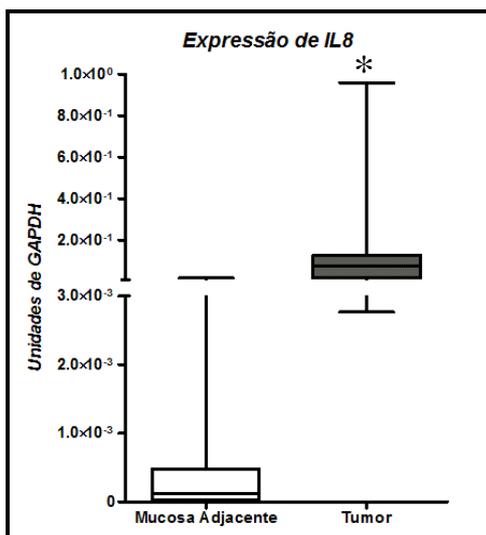
<b>Nome da Via</b>	<b>p-ajustado</b>	<b>N° de genes observados</b>
<i>Cytokine-cytokine receptor interaction</i>	1,6x10 <sup>-14</sup>	142
<i>ECM-receptor interaction</i>	9,4x10 <sup>-14</sup>	60
<i>Cell adhesion molecules (CAMs)</i>	3,8x10 <sup>-13</sup>	82
<i>Amoebiasis</i>	1,1x10 <sup>-11</sup>	67
<i>Focal adhesion</i>	3,8x10 <sup>-11</sup>	107
<i>Neuroactive ligand-receptor interaction</i>	5,7x10 <sup>-9</sup>	130
<i>Leukocyte transendothelial migration</i>	1,4x10 <sup>-6</sup>	61
<i>Chemokine signaling pathway</i>	6,9x10 <sup>-6</sup>	88
<i>Glycosaminoglycan biosynthesis - chondroitin sulfate</i>	1,2x10 <sup>-5</sup>	17
<i>Malaria</i>	1,4x10 <sup>-5</sup>	31
<i>Protein digestion and absorption</i>	1,4x10 <sup>-5</sup>	44
<i>Calcium signaling pathway</i>	1,8x10 <sup>-5</sup>	82
<i>Dilated cardiomyopathy</i>	2,9x10 <sup>-5</sup>	47
<i>Complement and coagulation cascades</i>	3,6x10 <sup>-5</sup>	38
<i>Arrhythmogenic right ventricular cardiomyopathy</i>	4,1x10 <sup>-5</sup>	40
<i>Rheumatoid arthritis</i>	4,2x10 <sup>-5</sup>	47
<i>Axon guidance</i>	4,9x10 <sup>-5</sup>	62
<i>Staphylococcus aureus infection</i>	1,1x10 <sup>-4</sup>	31
<i>Hypertrophic cardiomyopathy (HCM)</i>	1,9x10 <sup>-4</sup>	42
<i>Vascular smooth muscle contraction</i>	4,1x10 <sup>-4</sup>	54
<i>Pathways in cancer</i>	5,5x10 <sup>-4</sup>	130
<i>Salivary secretion</i>	5,9x10 <sup>-4</sup>	43
<i>Osteoclast differentiation</i>	6,0x10 <sup>-4</sup>	58
<i>TGF-beta signaling pathway</i>	6,0x10 <sup>-4</sup>	41
<i>Tight junction</i>	8,2x10 <sup>-4</sup>	59
<i>Regulation of actin cytoskeleton</i>	1,2x10 <sup>-3</sup>	88
<i>Leishmaniasis</i>	1,6x10 <sup>-3</sup>	35
<i>Hematopoietic cell lineage</i>	1,9x10 <sup>-3</sup>	41
<i>Gap junction</i>	6,1x10 <sup>-3</sup>	40

Novamente grande parte das vias enriquecidas estava relacionada a processos inflamatórios. Além disso, uma análise feita no STRING aponta o gene *IL8* como um gene central nas interações proteicas na lista de genes TxN (Figura 4.4).



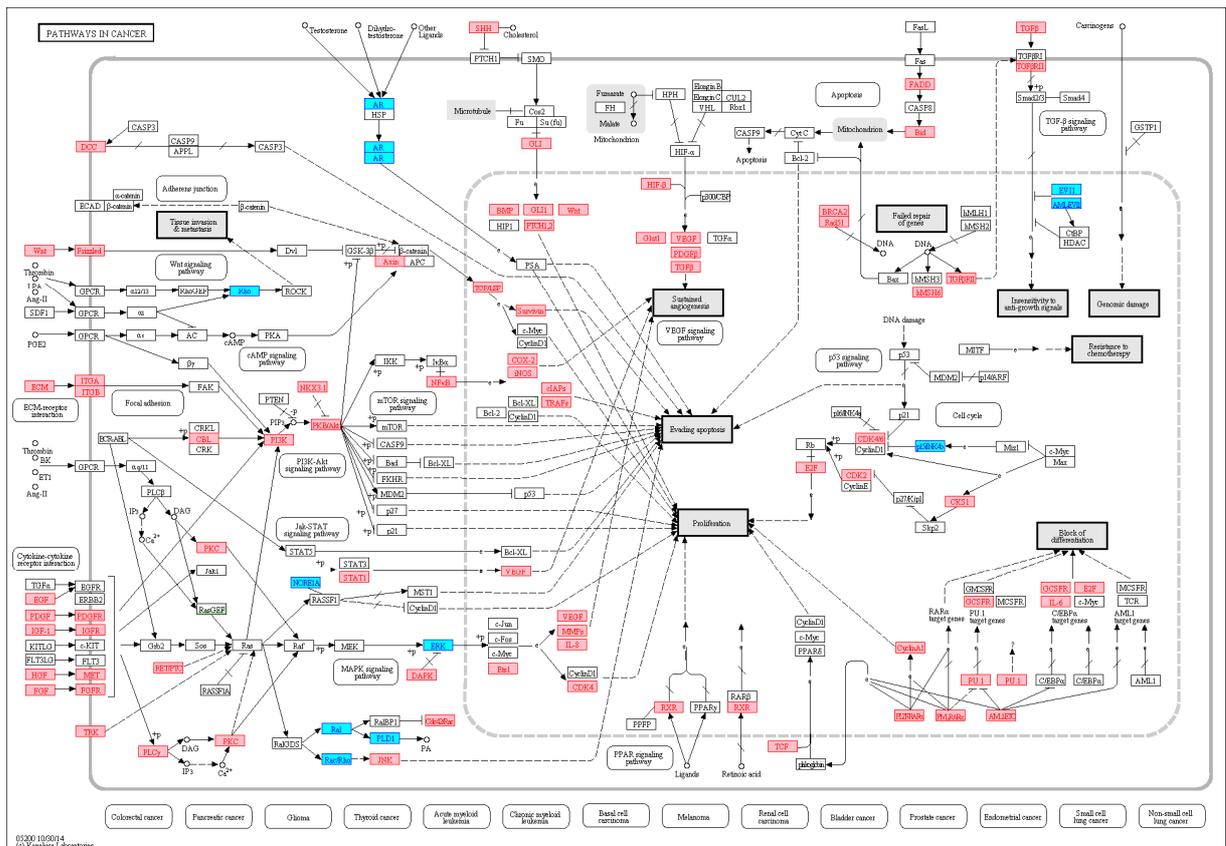
**Figura 4.4 – Análise de interação proteína-proteína realizada no STRING.** A seta indica *IL8* como central nas interações proteicas.

À luz desses dados o gene *IL8* foi validado por PCRq em um grupo independente de amostras e sua expressão relativa é mostrada na figura 4.5.



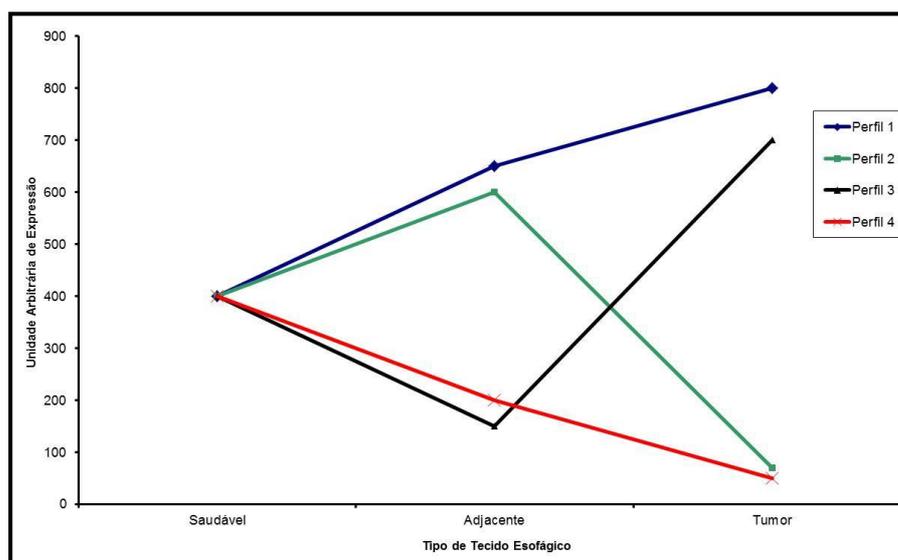
**Figura 4.5 - Validação do gene *IL8* em 28 amostras independentes por PCRq.** Comparação entre a expressão no tumor e mucosa normal adjacente mostrando maior expressão no tumor (\* $p < 0,00001$ ).

Por se tratar da comparação do tumor com o tecido normal adjacente a via “Pathways in cancer” foi explorada mais a fundo e é detalhada na figura 4.6.



**Figura 4.6 - Detalhamento da via enriquecida “Pathways in cancer” do KEGG a partir da comparação TxN (caixas em vermelho - genes mais expressos no tumor; caixas em branco - genes não classificados com GDE; caixas em azul - genes mais expressos na mucosa adjacente ao tumor.**

Uma análise com os genes que foram encontrados diferencialmente expressos tanto na comparação NxH quanto na comparação TxN foi feita. A interseção entre as listas foi composta de 125 genes, compondo 4 possíveis perfis, ilustrados na Figura 4.7, quais sejam: perfil 1, aqueles genes que apresentaram expressão aumentada no tecido adjacente em relação à mucosa sadia, assim como no tecido tumoral em relação ao tecido adjacente e à mucosa sadia; perfil 2, aqueles genes que apresentaram expressão aumentada no tecido adjacente em relação à mucosa sadia, entretanto menor expressão no tecido tumoral em relação ao tecido adjacente; perfil 3, aqueles genes que apresentam expressão diminuída no tecido adjacente em relação à mucosa sadia, por outro lado apresentam aumento de expressão no tecido tumoral em relação à mucosa adjacente e perfil 4, aqueles genes que apresentam expressão diminuída no tecido adjacente em relação à mucosa sadia, assim como no tecido tumoral em relação ao tecido adjacente.



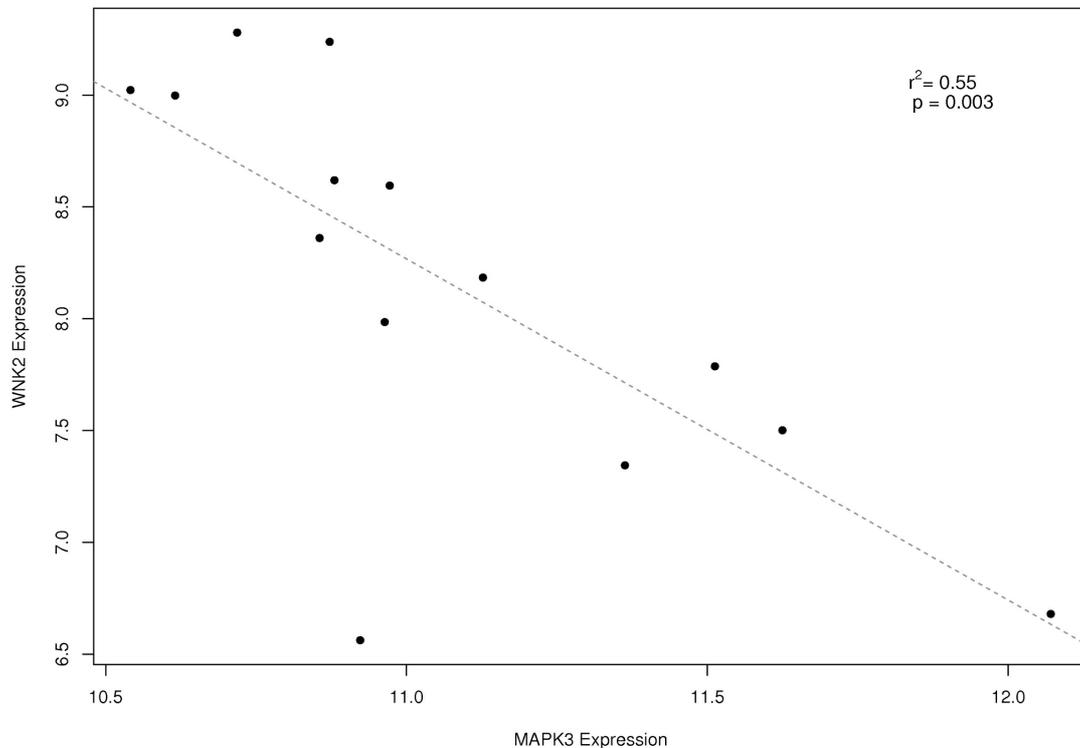
**Figura 4.7 – Gráfico de expressão hipotética representando os quatro perfis de expressão entre os diferentes tecidos analisados: saudável, adjacente ao tumor e CEE.**

Dos 125 genes encontrados na interseção entre as listas de GDE na comparação NxH e TxN, 94 se enquadravam no perfil 1, apenas um no perfil 2, três genes no perfil 3 e 27 pertenciam ao perfil 4. Alguns genes em cada perfil foram destacados na tabela 4.7.

**Tabela 4.7 - Genes correspondentes a cada um dos perfis de expressão encontrados na interseção das comparações NxH e TxN.**

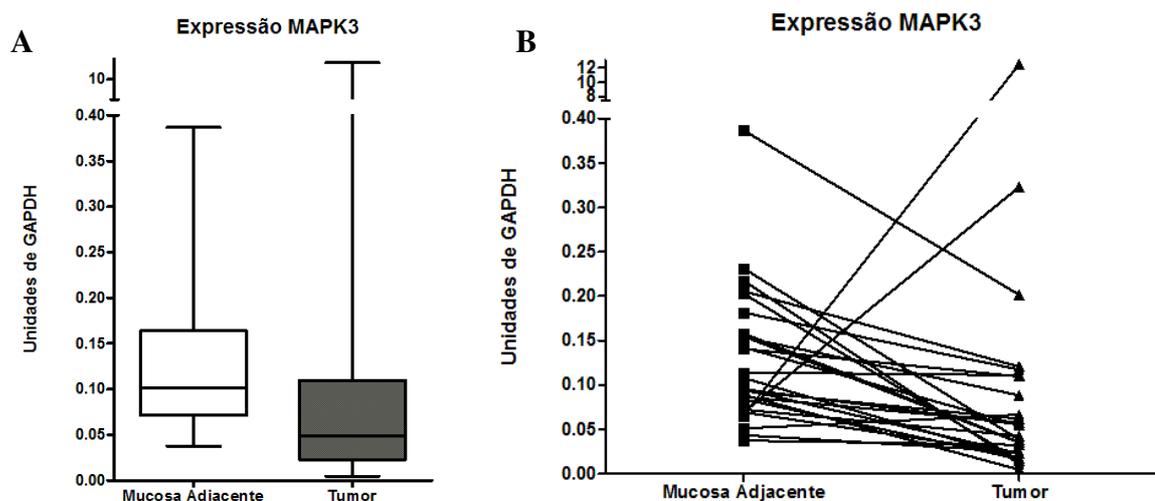
Gene	Perfil	logFC no NxH	logFC no TxN
<i>FBP4</i>	1	3,0	4,6
<i>MCAM</i>	1	2,3	2,7
<i>SLC2A6</i>	1	2,0	2,7
<i>NEK6</i>	1	1,1	2,3
<i>PLA2G2A</i>	2	2,9	-5,5
<i>WNK2</i>	3	-1,3	3,1
<i>BST1</i>	3	-1,4	2,5
<i>CBS</i>	3	-2,9	6,8
<i>IL12A</i>	4	-1,4	-2,4
<i>NOG</i>	4	-1,8	-1,8
<i>PARD6B</i>	4	-1,2	-1,2
<i>LPAR4</i>	4	-1,4	-1,2

O aumento na expressão de ligantes da via de WNT e genes *FZD*, bem como a diminuição na expressão de *MAPK3* (*ERK1*) (figura 4.6) e o aumento de *WNK2* (tabela 4.7) além da constatação de que *WNK2* e *MAPK3* se correlacionavam negativamente (figura 4.8) nos levou a um aprofundamento da via de WNT.



**Figura 4.8 – Correlação inversa entre a expressão gênica de *WNK2* e *MAPK3* (*ERK1*) .**

Os genes *WNK2* e *MAPK3* foram selecionados para validação por PCRq. A figura 4.9 traz a expressão do *MAPK3* no grupo de amostras de validação. Infelizmente o gene *WNK2* não foi possível avaliar por esta técnica.



**Figura 4.9 - Validação do gene *MAPK3* em 28 amostras independentes por PCRq. (A): Diagrama de Tukey mostrando a diferença de expressão entre o tumor e normal adjacente. (B): Dot plot mostrando a mesma comparação par a par. ( $p = 0,002$ )**

Dos 224 genes anotados para via de WNT, 84 se mostraram diferencialmente expressos em TxN e estão descritos na tabela 4.8.

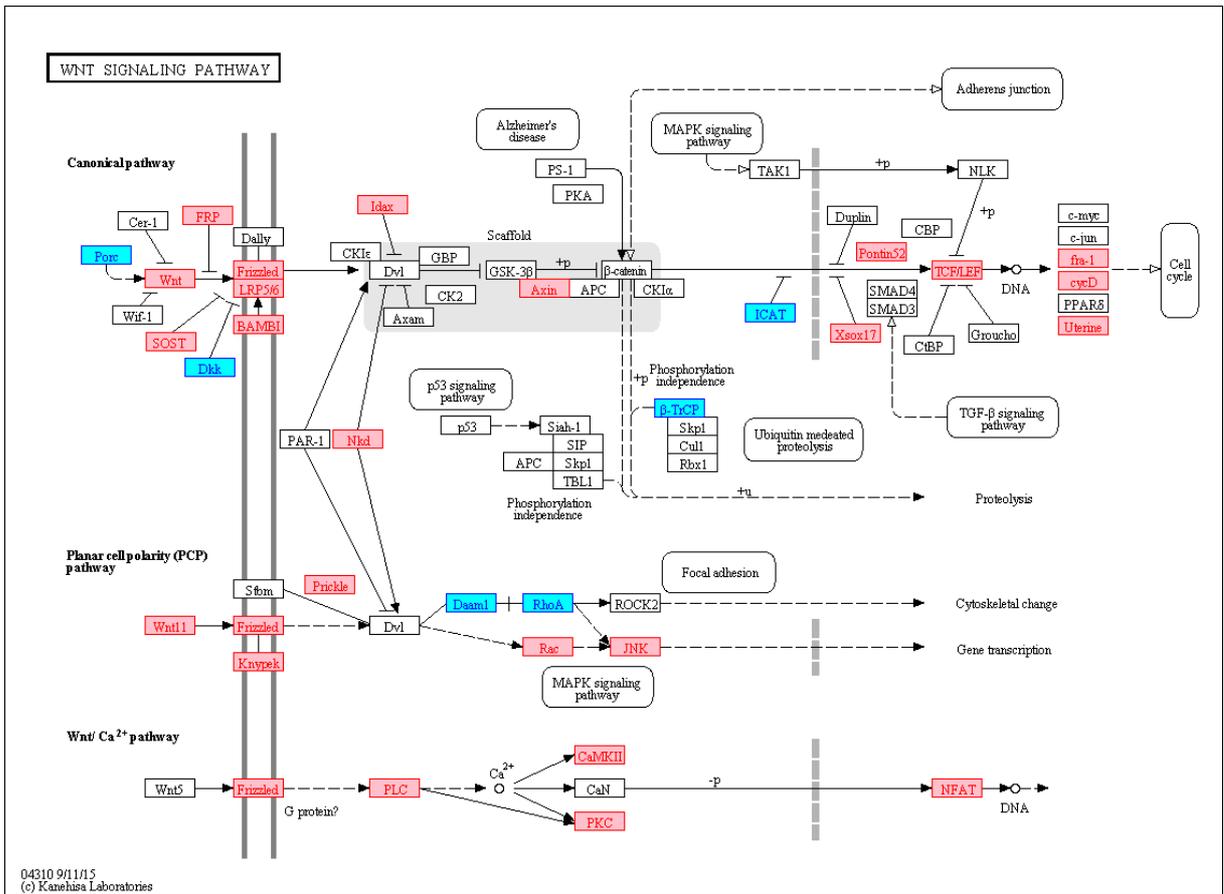
**Tabela 4.8 - Genes diferencialmente expressos na análise TxN que fazem parte da via de WNT.**

<i>Gene da via de WNT</i>	<i>Valor ajustado de p</i>	<i>logFC</i>
<i>LGR5</i>	5,5x10 <sup>-19</sup>	7,5
<i>WNT2</i>	1,2x10 <sup>-9</sup>	6,3
<i>SOX1</i>	8,7x10 <sup>-9</sup>	5,3
<i>SALL1</i>	1,3x10 <sup>-16</sup>	5,0
<i>SOST</i>	1,3x10 <sup>-7</sup>	4,9
<i>GPC6</i>	1,8x10 <sup>-16</sup>	4,0
<i>ROR2</i>	7,4x10 <sup>-22</sup>	4,0
<i>SFRP4</i>	2,0x10 <sup>-6</sup>	3,9
<i>NKD2</i>	2,6x10 <sup>-19</sup>	3,8
<i>MAP1B</i>	3,5x10 <sup>-20</sup>	3,7
<i>PRKCG</i>	2,2x10 <sup>-3</sup>	3,7
<i>MMP7</i>	1,4x10 <sup>-4</sup>	3,7
<i>PAX2</i>	2,9x10 <sup>-2</sup>	3,7
<i>CXXC4</i>	1,0x10 <sup>-7</sup>	3,5
<i>SDC2</i>	3,2x10 <sup>-17</sup>	3,5
<i>RSPO3</i>	1,6x10 <sup>-6</sup>	3,4
<i>PLCB4</i>	4,4x10 <sup>-13</sup>	3,2
<i>FZD2</i>	8,3x10 <sup>-25</sup>	3,2
<i>BAMBI</i>	1,5x10 <sup>-15</sup>	3,2
<i>NFATC1</i>	1,5x10 <sup>-25</sup>	3,1
<i>DLG2</i>	1,9x10 <sup>-10</sup>	-2,9
<i>SFRP5</i>	8,5x10 <sup>-3</sup>	2,8
<i>PRKD1</i>	7,4x10 <sup>-12</sup>	2,8
<i>PLCB1</i>	1,9x10 <sup>-10</sup>	2,7
<i>DAB2</i>	2,1x10 <sup>-33</sup>	2,6
<i>PRICKLE1</i>	6,0x10 <sup>-13</sup>	2,5
<i>RUNX2</i>	7,6x10 <sup>-13</sup>	2,5

<i>RSPO2</i>	1,4x10 <sup>-2</sup>	2,4
<i>GPC4</i>	3,2x10 <sup>-11</sup>	2,3
<i>WNT8B</i>	5,0x10 <sup>-3</sup>	2,1
<i>ROR1</i>	1,7x10 <sup>-8</sup>	-2,1
<i>WNT16</i>	1,1x10 <sup>-3</sup>	2,0
<i>TCF7L1</i>	6,8x10 <sup>-14</sup>	2,0
<i>GPC3</i>	1,9x10 <sup>-10</sup>	2,0
<i>NFATC4</i>	4,6x10 <sup>-10</sup>	2,0
<i>FZD8</i>	6,1x10 <sup>-10</sup>	1,9
<i>RAC2</i>	6,2x10 <sup>-14</sup>	1,9
<i>GPC2</i>	6,5x10 <sup>-7</sup>	1,8
<i>PRKCA</i>	7,2x10 <sup>-15</sup>	1,7
<i>MAPK3</i>	4,7x10 <sup>-14</sup>	-1,7
<i>SOX17</i>	9,4x10 <sup>-7</sup>	1,7
<i>CRYBB2</i>	1,2x10 <sup>-2</sup>	1,7
<i>DAAM2</i>	3,2x10 <sup>-5</sup>	1,7
<i>TCF4</i>	2,6x10 <sup>-10</sup>	1,7
<i>WNT3A</i>	4,7x10 <sup>-6</sup>	1,7
<i>WNT11</i>	3,0x10 <sup>-3</sup>	1,6
<i>PRKCQ</i>	1,9x10 <sup>-5</sup>	1,6
<i>ARRB2</i>	4,0x10 <sup>-28</sup>	1,6
<i>MAGI3</i>	1,2x10 <sup>-14</sup>	-1,6
<i>WNT6</i>	9,1x10 <sup>-4</sup>	1,6
<i>RSPO4</i>	2,9x10 <sup>-2</sup>	1,6
<i>PORCN</i>	1,8x10 <sup>-6</sup>	-1,6
<i>CDC25C</i>	1,3x10 <sup>-17</sup>	1,6
<i>CTNNBIP1</i>	2,1x10 <sup>-12</sup>	-1,6
<i>DKK1</i>	1,5x10 <sup>-2</sup>	-1,6
<i>MAPK10</i>	2,5x10 <sup>-5</sup>	1,5
<i>CDK1</i>	5,6x10 <sup>-17</sup>	1,5
<i>AXIN2</i>	7,5x10 <sup>-5</sup>	1,5
<i>SFRP2</i>	6,4x10 <sup>-3</sup>	1,4

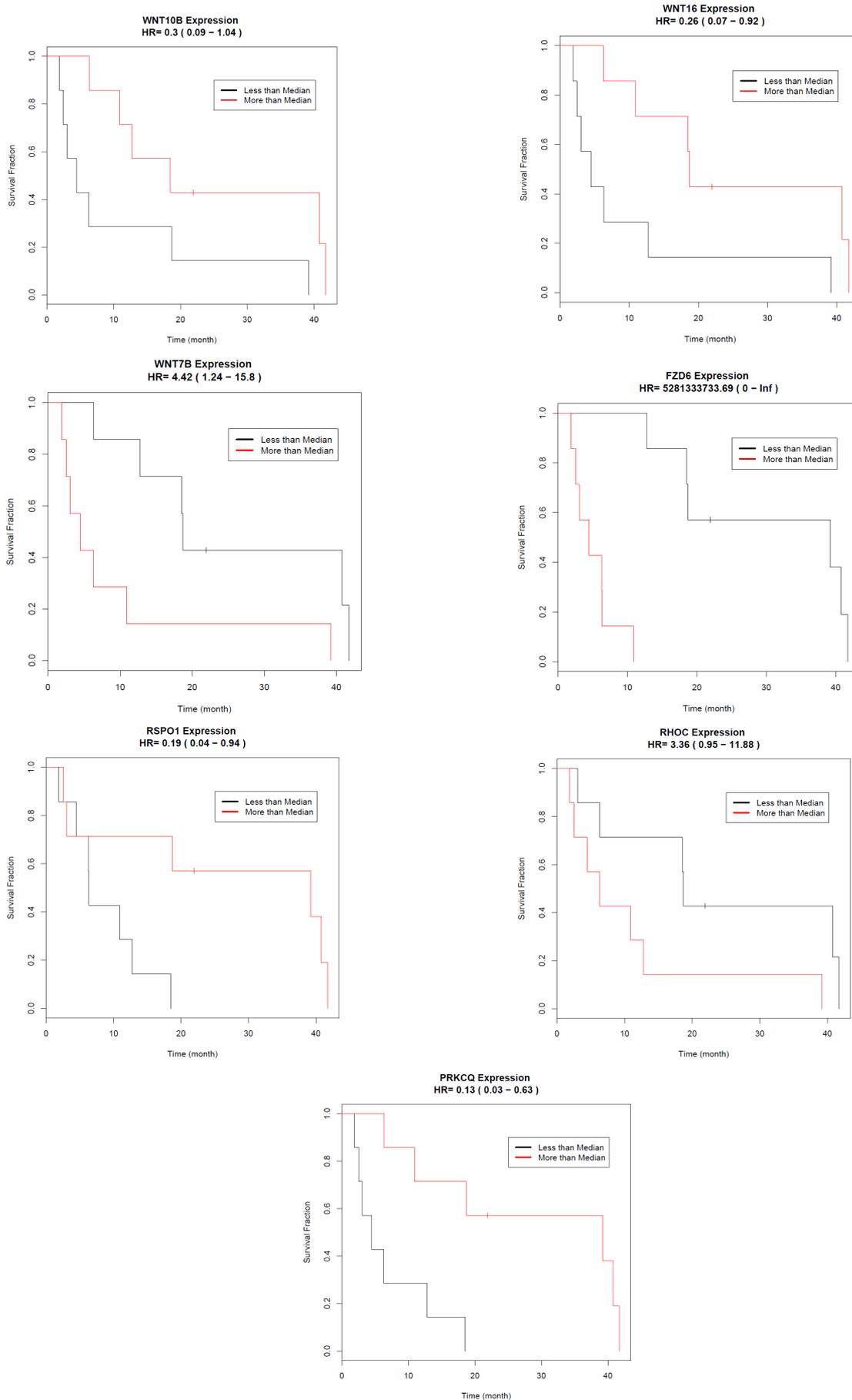
<i>FZD7</i>	$3,3 \times 10^{-5}$	1,4
<i>LEF1</i>	$1,8 \times 10^{-3}$	1,4
<i>FZD6</i>	$2,4 \times 10^{-18}$	1,4
<i>LRP5</i>	$4,0 \times 10^{-11}$	1,3
<i>PYGO1</i>	$3,3 \times 10^{-3}$	1,3
<i>FZD3</i>	$6,3 \times 10^{-6}$	1,3
<i>TCF3</i>	$1,3 \times 10^{-14}$	1,3
<i>LRP6</i>	$3,5 \times 10^{-12}$	1,3
<i>RSPO1</i>	$4,8 \times 10^{-2}$	1,3
<i>FZD4</i>	$2,1 \times 10^{-9}$	1,3
<i>WNT7A</i>	$4,0 \times 10^{-2}$	1,3
<i>FOSL1</i>	$2,9 \times 10^{-4}$	1,3
<i>FZD9</i>	$2,4 \times 10^{-5}$	1,3
<i>HIPK2</i>	$2,3 \times 10^{-9}$	-1,3
<i>CCND2</i>	$3,9 \times 10^{-3}$	1,3
<i>RHOC</i>	$3,4 \times 10^{-11}$	1,2
<i>FBXW11</i>	$4,6 \times 10^{-10}$	-1,2
<i>FHL2</i>	$4,4 \times 10^{-5}$	1,2
<i>WNT7B</i>	$2,9 \times 10^{-7}$	1,2
<i>WNT10B</i>	$1,2 \times 10^{-2}$	1,1
<i>NKD1</i>	$8,5 \times 10^{-4}$	1,1
<i>RUVBL1</i>	$7,1 \times 10^{-15}$	1,1
<i>DAAM1</i>	$7,7 \times 10^{-7}$	-1,1
<i>CAMK2A</i>	$4,7 \times 10^{-2}$	1,1
<i>RHOA</i>	$1,0 \times 10^{-15}$	-1,1

Um panorama de como esses genes estão distribuídos na via de WNT anotada no banco KEGG pode ser visto na figura 4.10.



**Figura 4.10 - Detalhamento dos GDE presentes na via de WNT (KEGG) (caixas em vermelho - genes mais expressos no tumor; caixas em branco - genes não classificados com GDE; caixas em azul - genes mais expressos na mucosa adjacente ao tumor).**

Os tumores foram divididos entre alta e baixa expressão utilizando a mediana da expressão de cada um dos 84 GDE da via de WNT como ponto de corte. A partir dessa classificação o valor prognóstico desse conjunto de genes foi testado. Apenas sete genes apresentaram alguma associação com a sobrevida global dos pacientes, quais sejam *FZD6*, *PRKCQ*, *RSPO1*, *RHOC*, *WNT16*, *WNT7B* e *WNT10B* (figura 4.11).



**Figura 4.11 - Curvas de Kaplan-Meier referentes a sobrevida global dos pacientes, em função da expressão de genes associados a via de WNT. HR: Hazard Ratio bruta.**

Uma segunda etapa na investigação do valor prognóstico de um marcador molecular é a determinação de independência de preditores conhecidos de desfecho como a idade e o estadiamento do tumor. Contudo, não foi possível a obtenção do estadiamento de quatro pacientes, como anteriormente mencionado na tabela 4.1. Desta forma a regressão de Cox foi realizada simulando os 16 arranjos possíveis para os estadiamentos não disponíveis. A tabela 4.9 resume o resultado dessas simulações.

**Tabela 4.9 – Regressão de Cox a partir das simulações de possíveis cenários para os casos com o estadiamento não informado.**

Gene	Nº de cenários onde o valor prognóstico do gene foi dependente do estadiamento	Amplitude dos valores de p encontrados	Amplitude dos HRs ajustados encontrados
<i>WNT10B</i>	9/16	0,01 - 0,07	0,12 - 0,32
<i>RHOC</i>	8/16	0,009 - 0,08	3,2 - 23,1
<i>RSPO1</i>	1/16	0,03 - 0,05	0,16 - 0,19
<i>FZD6</i>	ND	ND	ND
<i>WNT16</i>	1/16	0,01 - 0,06	0,09 - 0,26
<i>PRKCQ</i>	0/16	0,009 - 0,01	0,098 - 0,11
<i>WNT7B</i>	0/16	0,009 - 0,02	5,3 - 19,3

ND - Parâmetros impossíveis de calcular com a metodologia empregada. HR: *Hazard Ratio*

Os genes *WNT10B*, *RHOC* e *RSPO1* foram descartados das próximas etapas de análise devido à probabilidade de seus valores prognósticos não serem independentes do estadiamento (tabela 4.9). Por outro lado, os genes *PRKCQ* e *WNT7B* foram selecionados para serem validados em um grupo diferente de amostras já que em qualquer cenário possível eles manteriam seus valores prognósticos. O gene *FZD6* foi mantido para as próximas etapas apesar da impossibilidade de calcular sua HR ajustada uma vez que ele apresentou a maior distância entre as curvas de sobrevida.

O gene *WNT16* seria descartado assim como *RSPO1*, entretanto por ser um ligante da via WNT, como o gene *WNT7B*, e por ter apresentado uma relação entre o desfecho e a sua expressão gênica contrária aquela vista no *WNT7B*, somado a constatação de que a expressão desses dois genes se correlacionavam de maneira

inversa (Figura 4.12A) e ainda *WNT16* se correlacionava positivamente com a expressão de *MAPK10* (JNK) (Figura 4.12B) levou a formulação da hipótese de ativação diferencial da via de WNT, que será melhor discutida na próxima seção. Portanto *WNT16* foi mantido nas futuras análises.

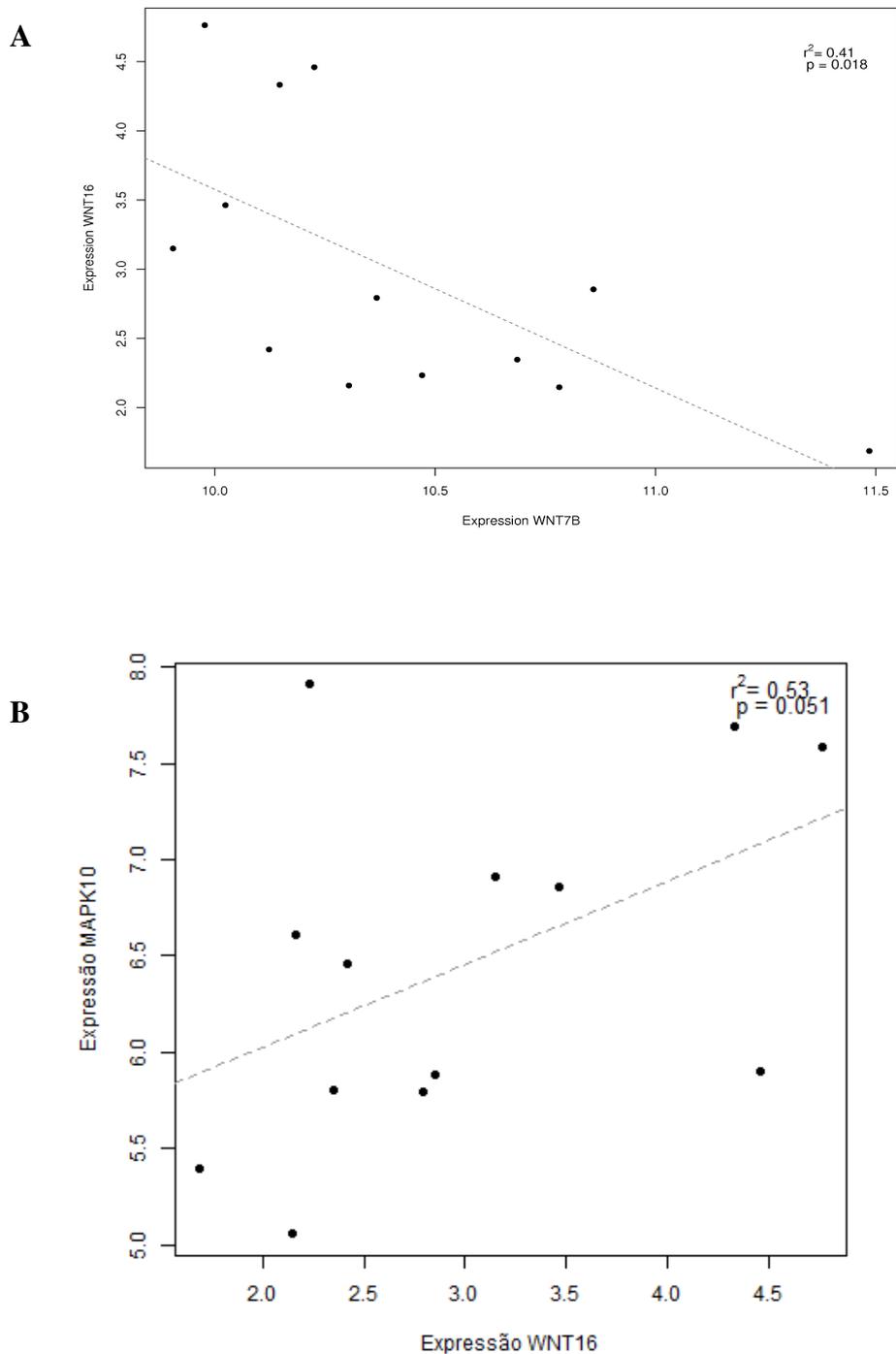


Figura 4.12 - Correlação inversa entre a expressão gênica de *WNT16* e *WNT7B* (A). Correlação positiva entre a expressão de *WNT16* e *MAPK10* (JNK) (B).

Conforme dito anteriormente, os genes *WNT16*, *WNT7B*, *PRKCQ* e *FZD6* foram selecionados para validação em um grupo de amostras independentes (28 pares de amostras, mucosa adjacente ao tumor e tecido tumoral).

Após seleção dos genes citados anteriormente, foi feita a reação de PCRq em 28 pares de amostras de CEE e mucosa adjacente ao tumor para avaliação do gene *WNT7B*. Tanto a maior expressão de *WNT7B* no tumor como seu valor prognóstico foram validados como mostra a figura 4.13.

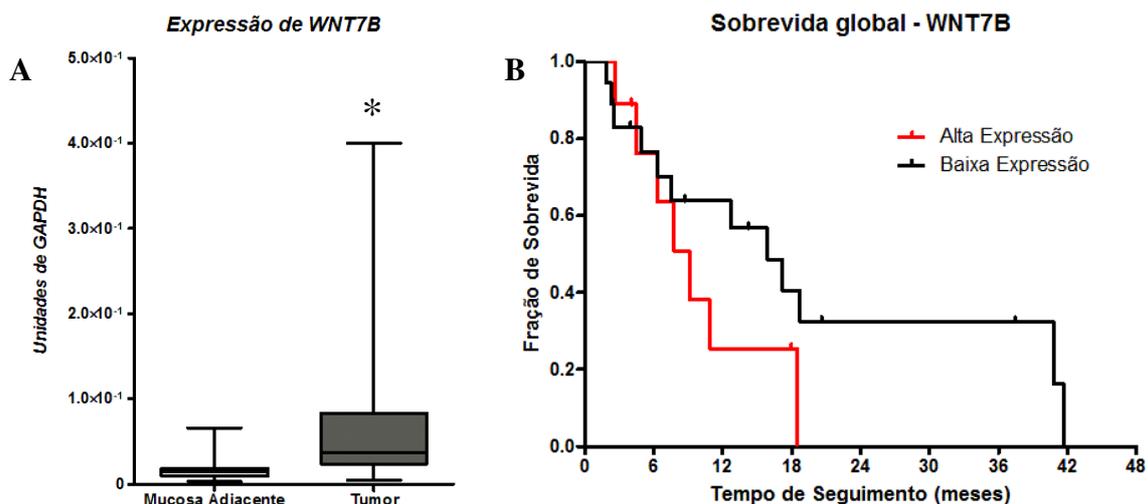
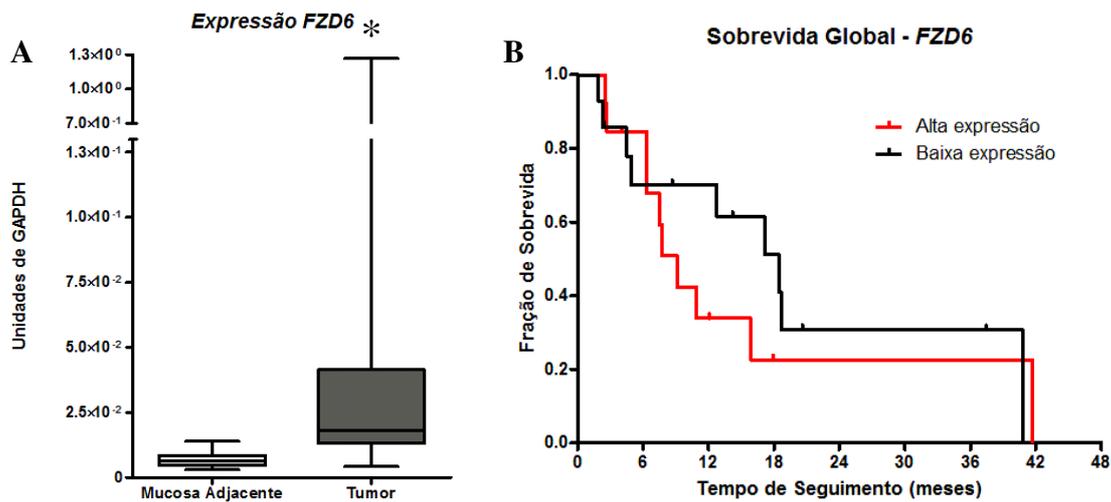


Figura 4.13. Validação do gene *WNT7B* em 28 amostras independentes por PCRq. (A) Comparação entre a expressão no tumor e mucosa normal adjacente mostrando maior expressão no tumor ( $p = 0,0002$ ). (B) Sobrevida global dos pacientes baseada na expressão de *WNT7B* pelo tumor com HR ajustado de 4,32 (1,08 - 17,24).

A validação do gene *FZD6* está descrita na figura 4.14 e embora o aumento da expressão deste gene no tumor em relação à mucosa adjacente tenha sido confirmado, o seu valor prognóstico não se confirmou.

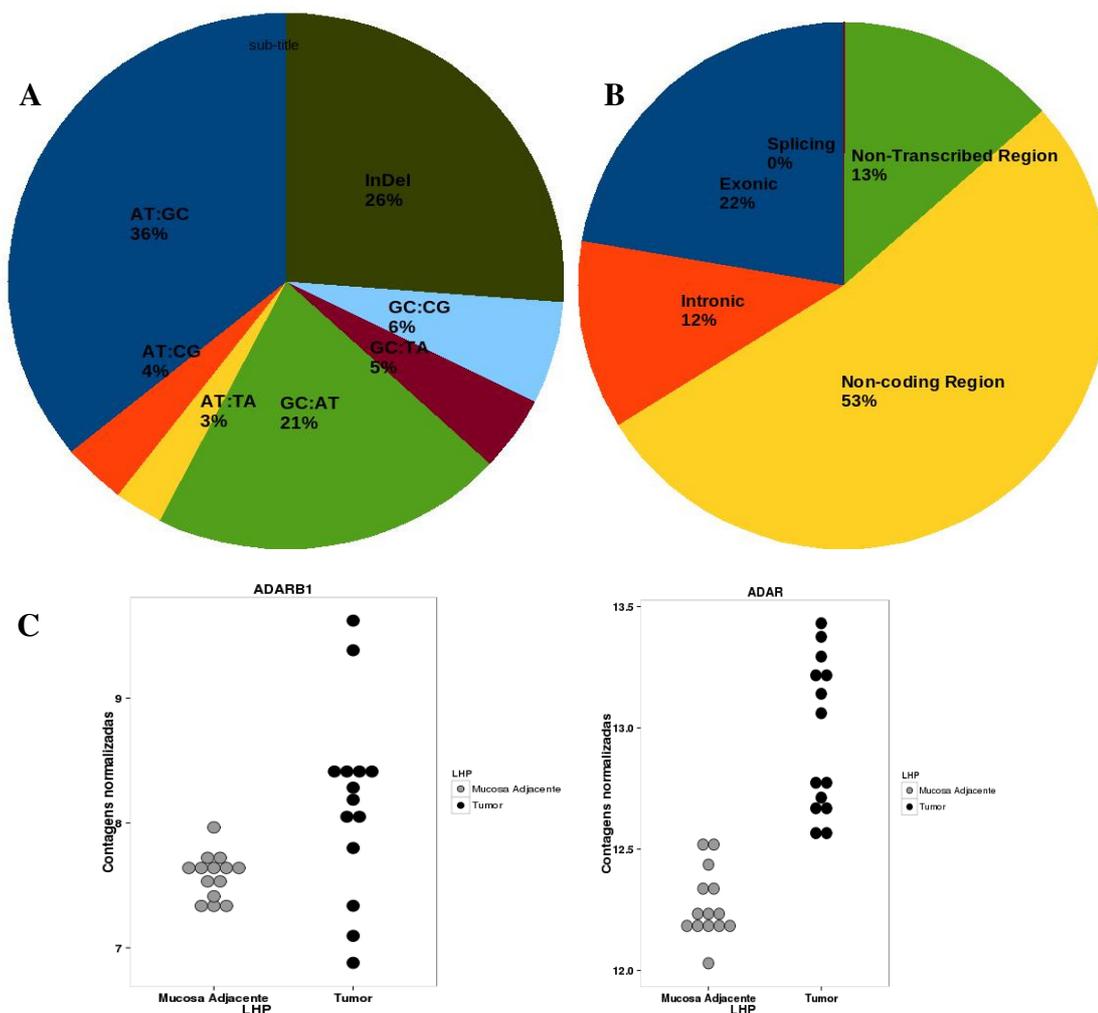


**Figura 4.14.** Validação do gene *FZD6* em 28 amostras independentes por PCRq. (A) Comparação entre a expressão no tumor e mucosa normal adjacente mostrando maior expressão no tumor (\* $p < 0,0001$ ). (B) Sobrevida global dos pacientes baseada na expressão de *FZD6* pelo tumor com HR ajustado de 0,83 (0,28 - 2,48).

Os genes *PRKCCQ*, *WNT16* foram testados por PCRq mas as reações não funcionaram em nenhuma tentativa adotada.

#### 4.5. ANÁLISE DAS SNVs (*single nucleotide variants*)

Primeiramente foram encontradas 21.079 SNVs nos 14 tumores sequenciados. Essas SNVs apresentaram o perfil descrito na figura 4.15.



**Figura 4.15 – Gráfico de setores mostrando o perfil de alterações encontradas nos transcritos expressos pelo tumor (A). Gráfico de setores mostrando a distribuição das alterações encontradas nos transcritos expressos pelo tumor de acordo com a região gênica, após a retirada das alterações do tipo AT:GC e InDels (B). Gráficos em dot plot mostrando a expressão diferencial dos genes que codificam enzimas que editam RNAm (C).**

Considerando apenas as alterações em regiões exônicas e sendo retiradas as alterações do tipo AT:GC e InDels a mediana das SNVs ao longo dos tumores foi de 65 variando de 33 a 199. As alterações do tipo AT:GC foram desconsideradas pois há uma grande probabilidade de serem proveniente de edição no RNAm feitas por enzimas da família ADAR (*Adenosine Deaminases acting on RNA*) que estão super-expressas no tumor (figura 15C).

Os genes mais frequentemente mutados dentre os tumores estão representados na figura 4.5.2, com destaque para o gene *TP53* (78,6%) e o *LOC389831* (42,9%). Os outros genes mutados não atingiram uma frequência maior que 20% chegando a 883 genes mutados.

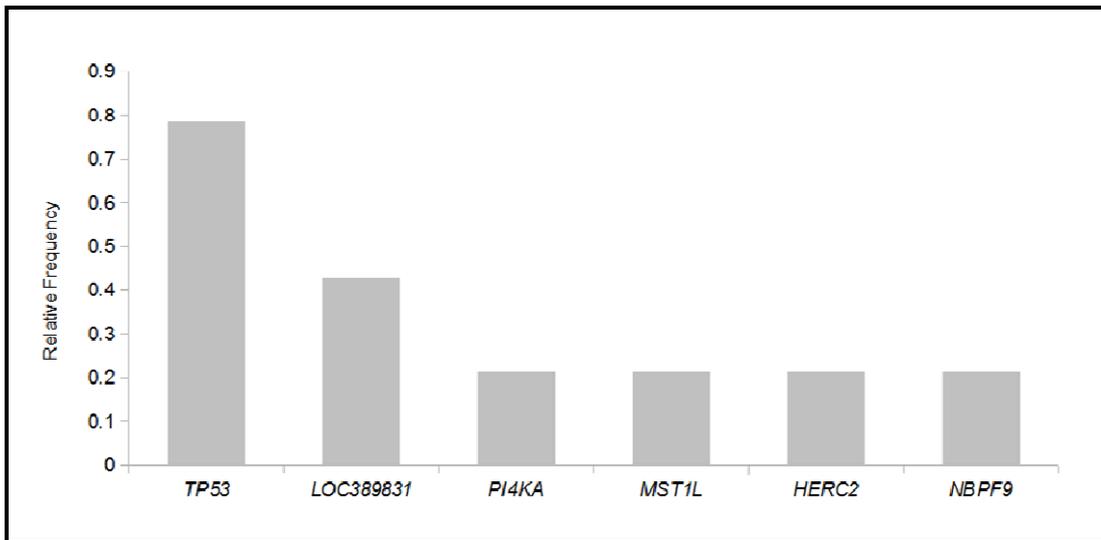


Figura 4.16 – Gráfico em barras mostrando a frequência dos genes mais mutados dentre as amostras de CEE, após a retirada das alterações do tipo AT:GC e InDels.

O gene *TP53* foi sequenciado em um conjunto independente de 42 tumores de esôfago onde 29 (69,0%) apresentaram mutação neste gene endossando a frequência encontrada no RNA-seq (p=0,51). No total foram encontradas 39 mutações em sua maioria levando a troca de aminoácido. Vale ressaltar que foram encontradas 10 mutações em íntrons, mas todo paciente com mutação em um íntron também apresentava mutação em algum éxon. A figura 4.17 ilustra o perfil mutacional, em que a maioria das mutações foi de AT>GC (32%), bem como a proporção de cada consequência das mutações, em que 60% representam mutações com troca de aminoácidos.

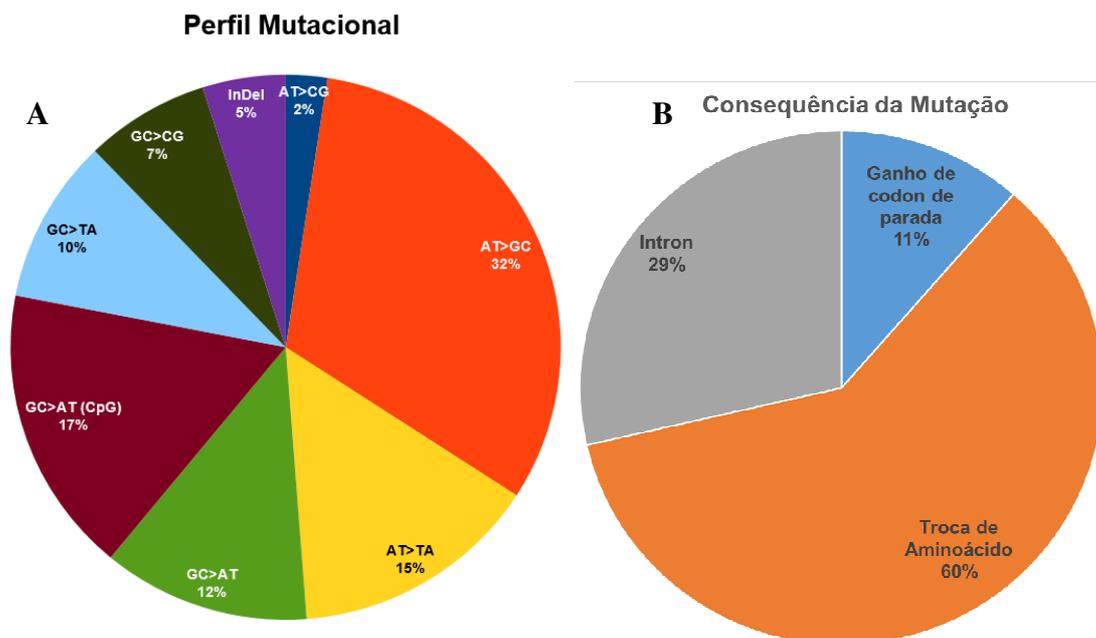
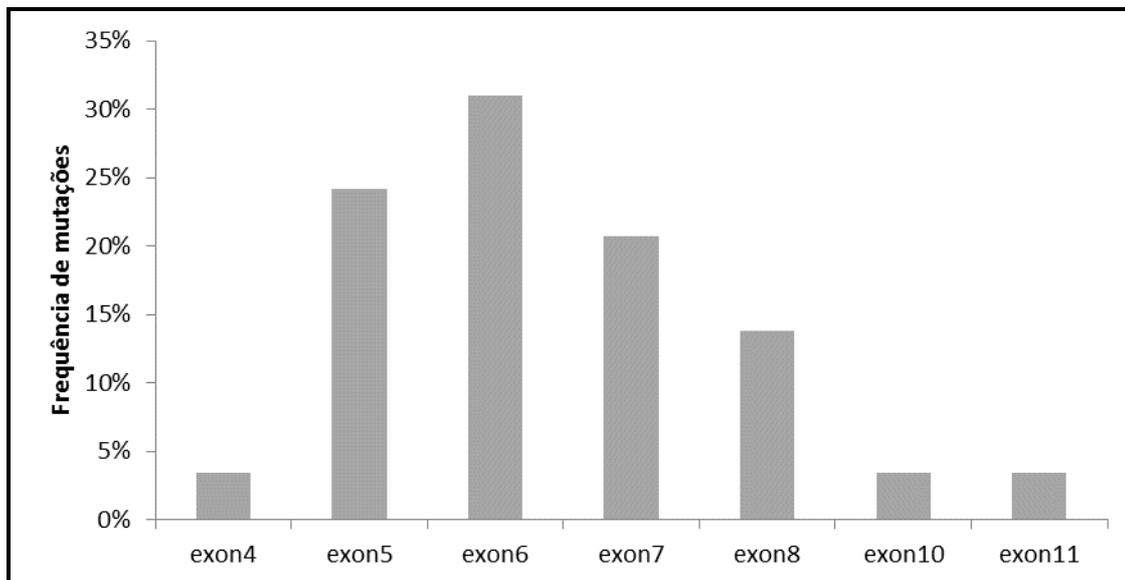


Figura 4.17 – Gráfico de setores ilustrando o perfil mutacional (A) bem como a proporção de cada consequência das mutações (B) no gene *TP53* analisadas por NGS em 42 tumores.

Como pode ser visto na figura 4.18 os éxons 5 a 8 somam mais de 90% das mutações com impacto na proteína. É importante ressaltar que não foi encontrada mutação nos exons 1, 2, 3 e 9 apesar de terem sido sequenciados .



**Figura 4.18 – Gráfico em barras mostrando a frequência de mutações encontradas por exon do gene *TP53*.**

Uma investigação acerca da associação do status mutacional do gene *TP53* com características clínico-patológicas foi realizada e nenhuma característica apresentou significância. Estas características foram: sobrevida global ( $p = 0,66$ ), estadiamento do tumor ( $p = 0,39$ ), grau de diferenciação do tumor ( $p = 1,00$ ) e presença de metástase linfonodal ou à distância ( $p = 0,07$ ).

#### **4.6. ANÁLISE DE TRANSCRITOS QUIMÉRICOS**

A busca por transcritos quiméricos que poderiam revelar fusões gênicas mostrou que dentre os 14 tumores avaliados apenas sete apresentavam algum evento fusional totalizando 21 eventos fusionais. Vale ressaltar que nenhum transcrito quimérico foi encontrado em mais de um tumor como pode ser visto na tabela 4.10.

**Tabela 4.10 – Análise de fusão gênica. A tabela mostra os 21 transcritos quiméricos encontrados nos tumores usando o *software* FusionMap.**

ID_amostra	Cobertura	Cromossomos envolvidos	Transcrito fundidos
s1t	8	chrX-chr17	PLS3->IKZF3
s1t	6	chrX-chr6	SLC25A43->IMPG1
s2t	14	chr5-chr5	CCDC127->AHRR
s4t	4	chr12-chr12	RAB3IP->CCT2
s4t	13	chr14-chr14	HNRNPC->CHD8
s4t	3	chr17-chr17	POLR2A->NLGN2
s4t	25	chr2-chr2	NDUFA10->MYEOV2
s4t	5	chr3-chr3	MAP4->SPINK8
s4t	19	chr3-chr3	MAP4->SPINK8
s4t	37	chr3-chr3	FLNB->SLMAP
s4t	42	chr7-chr7	ZNF3->TAF6
s7t	24	chr17-chr17	SPAG9->CA10
s13t	4	chr5-chr5	NDUFAF2->ZSWIM6
s14t	5	chr10-chr10	SGMS1->CACNB2
s14t	5	chr10-chr17	CTBP2->EFCAB5
s14t	6	chr11-chr4	LOC100133315->DEFB131
s14t	18	chr8-chr13	YWHAZ->DACH1
s14t	5	chr14-chr17	TRAF3->CDK12
s14t	44	chr9-chr9	PHF2->C9orf102
s15t	21	chr1-chr1	CTTNBP2NL->ST7L
s15t	22	chr8-chr8	SFRP1->SLC20A2

Uma vez que metade da amostra estava livre de eventos fusionais algumas características foram testadas para verificar quais estariam associadas a esse fenômeno. Primeiramente foi testado o valor prognóstico da presença das fusões e verificou-se que as curvas de sobrevida praticamente se sobrepunham ( $p = 0,55$ ). Além disso, também foi refutada a ideia de que os eventos fusionais poderiam estar associados ao estadiamento do tumor ( $p = 0,17$ ).

Um possível processo capaz de originar fusões gênicas são os problemas no mecanismo de reparo de quebras no DNA, principalmente *Non-homologous End joining* (NHEJ). A partir do RNA-seq foram testados 14 genes que compõe esse mecanismo de reparo. Os genes testados foram: *XRCC4* ( $p = 0,93$ ), *XRCC5* ( $p = 0,63$ ), *XRCC6* ( $p = 0,77$ ), *PRKDC* ( $p = 0,99$ ), *LIG4* ( $p = 0,38$ ), *NHEJ1* ( $0,90$ ), *DCLRE1C* ( $p = 0,42$ ), *POLM* ( $p = 0,62$ ), *POLL* ( $p = 0,88$ ), *PNKP* ( $p = 0,15$ ), *APTX* ( $p = 0,19$ ), *APLF* ( $p = 0,16$ ), *ATM* ( $p = 0,67$ ) e *ATR* ( $0,03$ ). O único com valor estatisticamente significativo foi o *ATR*, que está ilustrado na figura 4.6.1.

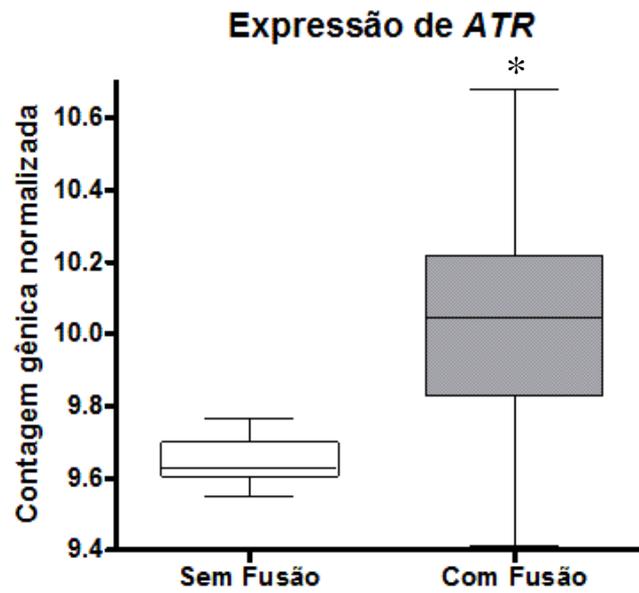


Figura 4.19 - Comparação entre a expressão genica do gene *ATR* mostrando maior expressão desse gene em amostras tumorais que apresentaram fusão comparadas as que não apresentaram fusão (\*p = 0,03).

## 5. DISCUSSÃO

Neste trabalho foram avaliadas alterações quantitativas (abundância) e qualitativas (sequência) do RNA de 14 pares de amostras de CEE em comparação a mucosa adjacente ao tumor além de oito mucosas de esôfago de indivíduos sem câncer através da técnica de RNA-seq. Quanto as alterações nos níveis de expressão dos RNAm, foi mostrado que quase 7.000 genes foram considerados diferencialmente expressos, mas apenas 1.226 genes apresentavam consistente variação em pelo menos 75% dos tumores em relação a mucosa adjacente. A análise referente as mutações presentes nos transcritos revelou que em média havia 50 mutações por tumor variando de 20 a 120 mutações, sendo *TP53* o gene mais frequentemente mutado (em 78% das amostras) seguido por *LOC389831* (com 40% das amostras mutadas) enquanto os demais genes mutados foram frequentes em 20% ou menos das amostras de CEE avaliadas. Também foi investigada a presença de transcritos quiméricos e foi visto um total de 21 fusões distribuídas em sete tumores sem que nenhuma fusão fosse encontrada em mais de um tumor.

Uma das métricas mais comumente utilizadas para se avaliar a qualidade dos dados gerados por sequenciamento é a atribuição de um índice *Q* (*Q score*) a cada base sequenciada. Os índices de qualidade medem a probabilidade que uma base é chamada (*base calling*) incorretamente e sua pontuação de qualidade é feita através de um algoritmo *Phred-like*, semelhante ao desenvolvido originalmente para os dados de sequenciamento utilizando a plataforma Sanger. Pontuações baixas de *Q* podem levar a uma baixa acurácia das sequências, resultando em determinação de variantes incorretas, conclusões errôneas e custos mais elevados nos experimentos de validação.

Nesse trabalho, o sequenciamento utilizando a plataforma Illumina HiSeq 2500 gerou *reads* com a grande maioria das bases com uma pontuação maior ou

igual a Q30, o que significa um erro a cada 1000 bases sequenciadas. Este nível de precisão é ideal para uma variedade de aplicações, incluindo estudos por RNA-Seq assim como sugere a nota técnica (*Technote Q-score*) da Illumina. Os artigos mencionados na seção 1.5 desta tese, que também avaliaram CEE por RNA-seq, omitem em suas metodologias o filtro de qualidade aplicado, mas alguns citam o programa utilizado, desta forma assumimos que o corte usado para o filtro de qualidade foi o padrão dos programas que é de 20 na escala *Phred*. Também foi possível observar que, em geral, houve uma queda nos valores de qualidade na porção 3' das *reads*, um viés da plataforma de sequenciamento utilizada, fato que pode ser atribuído ao decaimento da fluorescência após 50 ciclos devido ao aumento no tamanho da sequência, como já discutido por Liu e colaboradores (2012).

Após o rigoroso controle de qualidade, cerca de 80% do total de *reads* geradas foram mapeadas corretamente no genoma de referência. Esta porcentagem é superior àquelas encontradas em outros trabalhos acima citados que obtiveram uma taxa de mapeamento de 60 a 75% das *reads* (Ma *et al*, 2012, Tong *et al*, 2012 e Jiang *et al*, 2014). Algumas razões podem ser responsáveis pelos não mapeamentos de uma *read* no genoma de referência, por exemplo, se a *read* em questão apresentar mais de 2 *erros de pareamento* em relação a sequência de referência ou se a *read* for mapeada em mais de uma região genômica (Kim *et al*, 2012).

Dentre os genes mais expressos na mucosa adjacente em relação à mucosa de indivíduos saudáveis há a predominância de genes ligados direta ou indiretamente ao processo inflamatório, denotando a importância deste *hallmark* do câncer (Hanahan e Weinberg, 2011) para o desenvolvimento do carcinoma epidermóide de esôfago. Abaixo serão discutidos os genes mais alterados observados nesta análise.

O gene *DES* (desmina) é um conhecido marcador de transição epitélio-mesênquima e já foi demonstrado que células de linhagem de câncer de pulmão

umentam a expressão de desmina quando desafiadas com TGF- $\beta$  (Chen *et al.*, 2013). Assim como a desmina o gene *ACTG2* também é um alvo de TGF- $\beta$  (Untergrasser *et al.*, 2005) e também foi implicado na transição epitélio-mesênquima em carcinoma hepatocelular (Benzoubir *et al.*, 2015).

A proteína produzida pelo gene *BIPFB1* foi encontrado mais abundante em escarro de indivíduos com doença pulmonar obstrutiva crônica (uma doença com forte componente inflamatório) em relação ao escarro de fumantes assintomáticos (Titz *et al.*, 2015). Além disso, Yang e colaboradores (2013) mostraram que a superexpressão do gene *BIPFB1* em linhagens celulares de carcinoma nasal diminuiu o crescimento celular, bloqueou a evolução do ciclo celular e tornou-as menos tumorigênicas em camundongos através da inibição da via de MAP quinase.

O gene *PLA2G4E* é da família das Fosfolipases A2 que está envolvido na produção do ácido araquidônico que é um precursor de eicosanoides, como a prostaglandina E2, fundamentais para a resposta inflamatória. Uma mutação que levou a perda de função da proteína codificada por *PLA2G4E* foi implicado com melhor prognóstico de pacientes com câncer oral após o tratamento (Biswas *et al.*, 2014).

A expressão do RNAm de *MUC5AC* está aumentada em epitélio respiratório desafiado por fuligem seja na forma de pó de carvão ou fumaça de cigarro (Kania *et al.*, 2014 e Haswell *et al.*, 2010). Além disso, este gene também foi encontrado super expresso no esôfago de Barrett (Arul *et al.*, 2000). Sendo assim, o *MUC5AC* poderia ter um efeito protetor sobre a mucosa esofágica contra a irritação provocada pelo cigarro ou pelo refluxo gástrico.

No que se refere aos genes menos expressos na mucosa adjacente em relação à mucosa sadia, o gene *SCARNA7* foi o de menor logFC. Este gene produz um RNA não codificante que junto com a proteína fibrilarina são responsáveis pela metilação do RNA ribossômico, que é uma modificação essencial para sua função. O *SCARNA7* já foi descrito como aumentado em tumores de mama murinos e humanos (Su *et al.*, 2014).

O *FOSB* é um membro da família dos fatores de transcrição AP1 e está relacionado com diferenciação da epiderme (Eckert *et al.*, 2013). Já foi demonstrado que o pH ácido é capaz de diminuir a expressão de *FOSB* em linhagem imortalizada de células do epitélio esofágico (Duggan *et al.*, 2006).

Cistationa-beta-sintase é uma enzima codificada pelo gene *CBS*, que participa do ciclo de transmetilação e da síntese de cisteína. Durante a conversão de

homocisteína em cisteína a cistationa-beta-sintase produz H<sub>2</sub>S que já foi descrito como responsável pelo aumento da proliferação celular em câncer de cólon (Szabo *et al*, 2013).

O *TAS1R1* codifica uma subunidade de um receptor de sabor (Jiang *et al*, 2012). Até o momento não há nenhum trabalho que mostre qualquer associação entre a expressão desse gene e câncer. Contudo, *TAS1R1* foi relatado aumentar a influxo do cátion Ca<sup>+</sup> em espermatozoides (Meyer *et al*, 2012) e sabe-se que o aumento da concentração de Ca<sup>+</sup> intracelular pode levar a ativação de vias de proliferação (Sahmoun e Singh, 2010).

Por fim, a expressão do gene *PEG10* foi associado com a redução da apoptose em carcinoma hepatocelular (Wang *et al*, 2008).

Infelizmente os trabalhos presentes na literatura que investigam a expressão de genes no CEE por RNA-seq não dispõe de comparação entre mucosa sadia e mucosa adjacente ao tumor. No entanto, nosso grupo realizou análise semelhante por microarranjo de DNA encontrando a mesma prevalência alta de procesos biológicos relacionados a inflamação enriquecidos (Nicolau, 2014).

A comparação TxN revelou um número de 1.226 GDEs em 14 pares de tumor e mucosa adjacente ao tumor. Este número é compatível com o relatado anteriormente por Tong e colaboradores (2012) que encontraram 1.730 GDEs em 12 amostras. Os genes com maior logFC nessa comparação foram *ZIC5*, *ADAMT20*, *HOXD13*, *MMP3* e *PRAME*. Os genes com menor logFC foram *PLA2G2A*, *LOC390705*, *CRISP3*, *DPCR1* e *KRT4*.

O gene *ZIC5* codifica uma proteína *zinc finger* que está associada a diversas malformações embrionárias ligadas a defeitos na crista neural (Grinberg e Millen, 2005). Além disso, sua expressão foi encontrada aumentada em diversos tipos de tumores de cérebro (Aruga *et al*, 2010).

O gene *ADAMTS20* produz uma metaloprotease que não foi associada com câncer até o momento. Contudo, *ADAMTS20* está relacionado com a transição epitélio-mesênquima das células da crista neural durante o desenvolvimento embrionário (Ismat, Cheshire e Andrew, 2013). Uma vez que tanto *ZIC5* quanto *ADAMTS20* apresentam papel no desenvolvimento da crista neural, e ambos estão superexpressos no CEE, pode haver uma correção destes dois genes de forma a atuarem em conjunto seja em processos fisiológicos, seja em processos patológicos.

*HOXD13* é um gene relacionado com o desenvolvimento embrionário do esôfago (Warot et al, 1997) e a proteína codificada por esse gene foi encontrada 3 vezes mais expressa em CEE (Cantile et al, 2009).

A Metaloproteinase-3 (MMP3) teve sua atividade relacionada à invasão tecidual em câncer cervical (Chen et al, 2014) e de próstata (Slavin et al, 2014). Além disso, a expressão de MMP3 está relacionada com o fenótipo invasor de linhagens celulares derivadas de tumores de esôfago (Uraoka et al, 2014).

PRAME é uma proteína que regula negativamente os receptores de ácido retinóico (RAR) (Bullinger et al, 2013). Até o momento, a expressão do gene *PRAME* não havia sido mostrada em câncer de esôfago, mas sabe-se que a baixa expressão de RAR está relacionada com a invasão linfonodal em CEE (Fujishima et al, 2011). Como a perda de diferenciação de células tumorais esofágicas é um fenômeno comumente observado neste tumor, a perda de RAR pode ser um dos motivos que leva a este fenômeno, uma vez que sua expressão leva a diferenciação de células epiteliais esofágicas (Song et al, 2009).

*LOC390705* é um pseudogene e não existem trabalhos relacionados a este pseudogene em tumores.

O CRISP3 é uma proteína secretada em vários fluidos humanos, como saliva, líquido seminal, suor e sangue (Udby et al, 2002). A função dessa proteína ainda não está muito clara, mas em câncer de próstata a sua alta expressão foi associada ao aumento na recidiva (Bjarteli et al, 2006) e em tumores de cavidade oral foi descrita como menos expressa que no tecido normal adjacente (Ko et al, 2012).

O gene *DPCR1* é importante no desenvolvimento de uma doença respiratória conhecida como pambronquiolite. Até o momento, foi mostrado que a presença de um polimorfismo nesse gene apresentou proteção no desenvolvimento do CEE (Shen et al, 2014).

A Citoqueratina 4 (KRT4) já foi mostrada ser um excelente marcador para o CEE, uma vez que a célula neoplásica não apresenta marcação dessa proteína enquanto que células displásicas têm pouca expressão e células normais expressam muito KRT4 (Takikita et al, 2011).

Integrando as análises NxH e TxN foram identificados 125 GDEs em comum entre as duas comparações. Os genes *FBP4*, *MCAM*, *SLC2A6*, *NEK6* e o *IL8* tiveram expressão aumentada na mucosa adjacente e ainda maior no tumor.

O gene *FBP4* produz uma proteína de mesmo nome que tem função de ligar ácidos graxos. A superexpressão dessa proteína foi relatada em carcinoma oral (Lee

*et al*, 2014). Além disso, no co-cultivo de adipócitos com células de câncer de próstata foi possível observar que a expressão de *FBP4* pelo adipócito foi capaz de aumentar a proliferação e a capacidade de invasão da célula neoplásica (Herroon *et al*, 2013). Um outro fenômeno relevante relacionado à *FBP4* é a relação entre sua expressão e a angiogênese em células endoteliais dependente e independente de VEGF (Harjes *et al*, 2014).

O gene *MCAM* codifica a proteína CD146 que parece estar envolvida no processo de metástase tumoral. Tumores de nasofaringe apresentaram uma alta expressão de CD146 em metástase, mas não no sítio primário (Lin *et al*, 2014). Em três modelos de tumores enxertados em camundongos, o uso do anticorpo monoclonal anti-CD146 (mAb-AA98) impediu a metástase dos tumores em até 80% quando comparados aos camundongos não tratados com o anticorpo. Além disso, o mAb-AA98 diminuiu em até 70% a densidade dos microvasos formados nos tumores (Yan *et al*, 2003).

Alterações no metabolismo energético do tumor levam ao aumento da captação de glicose pelas células neoplásicas (Warburg, Wind e Negelein, 1927). O gene *SLC2A6* codifica uma proteína transportadora de glicose, a GLUT6 cujo aumento de expressão foi relatada em diversos tumores, como adenocarcinoma de pâncreas, carcinoma ductal invasivo de mama e leiomioma de útero (Godoy *et al*, 2006). Além disso, GLUT6 foi o transportador de glicose mais expresso em câncer de endométrio tendo sua expressão correlacionada com a captação de glicose, taxa de glicólise aumentada e sobrevivência das células (Byrne *et al*, 2014).

NEK é uma família de 11 genes cuja expressão já foi relacionada com a progressão do ciclo celular, formação do fuso mitótico e sinalização mitogênica (Fry *et al*, 2012). A superexpressão do gene *NEK6* em linhagens celulares de câncer de bexiga e pulmão impediu a parada de ciclo celular induzida por p53 pela manutenção dos níveis de ciclina B e cdc2 (Jee *et al*, 2010).

Alguns genes se encaixaram num perfil exatamente inverso ao anterior, ou seja, genes que tiveram expressão na mucosa adjacente menor que na mucosa saudável e expressão no tumor menor que em ambas. Dos 27 genes encontrados com esse perfil, quatro são discutidos abaixo. São eles: *IL12A*, *NOG*, *PARD6B* e *LPAR4*.

A Interleucina 12 (IL-12) é uma citocina pró-inflamatória com potencial de induzir resposta Th1 bem como estimular células NK (Linfócitos *natural killers*) (Del Vecchio *et al*, 2007). Ardolino e colaboradores (2014) demonstraram que o

tratamento com IL-12 foi capaz de aumentar a sobrevivência de camundongos enxertados com tumores provenientes de células derivadas de linfomas que não expressavam moléculas de MHC-I e que este aumento na sobrevivência estava relacionado com a capacidade da IL-12 estimular células NK a produzir IFN- $\gamma$ . Além disso, já foi relatado que pacientes com CEE possuem níveis séricos de IL-12 menores que pessoas saudáveis e que um polimorfismo no gene *IL12A*, que codifica a proteína IL-12, aumentou a chance de desenvolver CEE de 1,1 a 1,9 vezes (Tao *et al*, 2012).

A baixa expressão do gene *NOG* quando combinada com alta expressão de *BMP6* foi associada a um prognóstico ruim em CEE (Yuen *et al*, 2012). Além disso, o tratamento de células oriundas de câncer de pulmão com a proteína codificada por esse gene (Noggin) foi capaz de impedir a fosforilação de Akt (Liu *et al*, 2014), que é uma via importante para o desenvolvimento do CEE. Também foi observado que a sua expressão diminuiu o volume de tumores enxertados de células de câncer de pulmão em camundongos atímicos (Kang, Gupta e Bogdanov, 2012).

O gene *PARD6B* codifica uma proteína da família Par6, que juntamente com Par3 e PKC atípica formam um complexo essencial para polarização e divisão das células epiteliais (Bose e Wrana, 2006). Apesar de já ter sido relatado que a superexpressão de Par6 leva ao aumento da proliferação em linhagem celular proveniente de câncer de mama (MCF-10A) (Nolan *et al*, 2008), em uma outra linhagem também oriunda de câncer de mama (NMuMG) a ativação de Par6 foi essencial para indução de apoptose (Avery-Cooper *et al*, 2014). Sendo assim, em CEE parece ser mais provável que Par6 leve a apoptose como mostrado em células NMuMG.

O ácido lisofosfatídico (LPA) é uma molécula sinalizadora derivada do lipídeo de membrana lisofosfatidilcolina. Os receptores de LPA estão normalmente associados a proteína G que desencadeiam sinais de proliferação e migração celular (Moolenaar, 2002). Lee e colaboradores (2008), investigando os receptores de LPA em células de neuroblastoma, concluíram que enquanto o receptor LPA1 promovia migração e invasão, o receptor LPA4 antagonizava esse efeito. Além disso, em uma linhagem celular derivada de carcinoma de laringe, que expressa constitutivamente LPA1 e não LPA4, quando estimulada com LPA, foi observada a indução de proliferação e migração celular, porém quando *LPA4* foi transfectado nessa célula, esta parou de responder ao estímulo com LPA (Matayoshi *et al*, 2013).

Dentre esses 125 genes na interseção entre NxH e TxN o gene *PLA2G2A* foi o único a se mostrar aumentado no tecido adjacente em relação ao sadio e diminuído em relação ao tumor. O *PLA2G2A* é um gene intimamente relacionado com o processo pró-inflamatório. Durante a progressão da carcinogênese, o tecido adjacente ao tumor, sofrendo os insultos dos agentes carcinogênicos, induziria um sinal pró-inflamatório, já o tumor propriamente dito seria prejudicado pela expressão desse gene. Já foi relatado em CEE que a expressão de *PLA2G2A* correlaciona-se negativamente com a capacidade do tumor invadir e conseqüentemente piora o prognóstico da doença (Ren *et al.*, 2013).

Os genes *WNK2*, *BST1* e *CBS* são genes que foram encontrados menos expressos na mucosa normal adjacente ao tumor quando comparados com a mucosa de indivíduos sadios, mas foram mais expressos no tumor do que em ambas as mucosas correspondendo ao terceiro perfil (perfil 3).

O gene *BST1* pertence à família das NADase/ADP-ribosil ciclase (Malavasi *et al.*, 2008). A superexpressão da proteína codificada por este gene, a CD157, modificou a morfologia e a aderência de células de câncer de ovário conferindo um fenótipo invasor a essas células (Morone *et al.*, 2012).

O gene *WNK2* será mencionado mais à frente e gene *CBS* já foi discutido anteriormente.

O enriquecimento pelo GO mostrou processos relacionados a inflamação e a análise pelo STRING apontou a IL8 como a proteína com mais interações, portanto o gene *IL8* foi avaliado por PCRq e foi validado o aumento de sua expressão no tumor em relação a mucosa adjacente. Em 2009, Ribeiro e colaboradores observaram que a expressão deste gene estava aumentada no tumor em relação a mucosa adjacente, mostrando que *IL8* tem um papel importante na carcinogênese esofágica. Além disso, Ogura e colaboradores (2013) correlacionaram o aumento de expressão de *IL8* e do seu receptor com uma pior sobrevida global, com uma maior profundidade do tumor e com invasão linfonodal o que não foi observado no presente trabalho.

A análise de enriquecimento pelo KEGG está de acordo com a encontrada por Tong e colaboradores (2012) que, inclusive, encontrou a via de WNT sobrerrepresentada.

O gene *WNK2* codifica uma proteína cinase citoplasmática que regula positivamente a via WNT/ $\beta$ -catenina (Serysheva *et al.*, 2013) e negativamente a via de EGF via ERK (Moniz *et al.*, 2007). O gene *MAPK3* (ERK1) foi encontrado

diminuído no tumor e sua expressão correlacionou negativamente com a de *WNK2*. Isso poderia explicar porque tumores epidermóides de esôfago não têm ativação da via de EGF (Gonzaga *et al*, 2012) como os epitélios que apresentam fatores etiológicos comuns, como os tumores de cabeça e pescoço (Morgan e Grandis, 2009) e pulmão (Sequist *et al*, 2006).

Os genes *MAPK3* e *WNK2* foram selecionados para validação por PCRq. A menor expressão no tumor do gene *MAPK3* validou os achados no RNA-seq. Contudo, dois pares (7%) tiveram uma expressão maior no tumor. Este evento corrobora com o achado anterior do nosso grupo onde menos de 10% das amostras analisadas apresentam superexpressão de receptores da família HER (Gonzaga *et al.*, 2012).

Diversos ligantes WNT e receptores FZD da via de WNT foram encontrados superexpressos nos tumores. Recentemente, Deng e colaboradores (2015) mostraram que 55% dos CEE tinham marcação positiva para Wnt2 enquanto que não houve marcação para essa proteína na mucosa adjacente ao tumor. No presente trabalho o gene *WNT2* foi o que apresentou maior *fold-change* (75 vezes) dentre os ligantes WNT, ou seja, amostras de pacientes com CEE possuem maior expressão de *WNT2* quando comparadas com o tecido adjacente ao tumor.

A alta expressão de *WNT16* foi associada a uma sobrevida mais longa nos pacientes com CEE. Teh e colaboradores (2007) relatam que *WNT16* ativa a via não canônica de WNT que leva a ativação de JNK. Nesse trabalho, o gene *MAPK10*, que codifica uma proteína JNK, apresentou expressão aumentada no tumor e essa expressão se correlacionou positivamente com a expressão de *WNT16*. O possível mecanismo para explicar o bom prognóstico de pacientes com CEE que expressam mais *WNT16* poderia estar em sua capacidade de induzir *MAPK10* cuja expressão foi capaz de aumentar a radiosensibilidade de linhagens de CEE (Wakita *et al.*, 2015) e aumentou a sensibilidade a cisplatina em uma linhagem de tumor papilífero de tireóide (Lapouge *et al.*, 2005). Tanto a radioterapia quanto a cisplatina são usualmente utilizados no tratamento do CEE.

Um outro ligante de WNT com impacto validado na sobrevida dos pacientes foi o gene *WNT7B* cuja alta expressão indicou mau prognóstico. A produção de *WNT7B* por células do microambiente tumoral aumentou a vascularização do tumor e a sua inativação em modelos murinos diminuiu a capacidade metastática de tumores de mama (Yeo *et al.*, 2014).

O gene que codifica o receptor de WNT, Frizzled-6 teve o aumento da expressão no tumor validada por PCRq. Em 2009, Wu, Zierold e Ranheim mostraram que a perda da redulação de *FZD6* é um evento preponderante para o desenvolvimento de leucemia de célula B. A suspeita do valor prognóstico de *FZD6* no RNA-seq não se confirmou no grupo de validação.

Por fim, os genes *WNT16*, *PRKCQ* e *WNK2* não amplificaram na maioria das amostras testadas ou apresentaram Cqs nas triplicatas muito diferentes entre si. Tal fato sugere que a detecção por PCRq destes genes por SYBR green não seja possível, o que pode ser explicado pela limitação de detecção do método, uma vez que genes pouco expressos (com menos de 11 cópias de seu RNAm transcrito) podem exibir uma grande variabilidade na quantificação de cada cópia analisada (BERNARDO *et al.*, 2013).

Todos esses dados em conjunto sugerem que a via de WNT desempenha um papel chave na carcinogênese do esôfago.

Além da expressão gênica o RNA-seq possibilita investigar variações na sequência do RNA. No presente trabalho a maior proporção de variações encontradas foi do tipo AT>GC o mesmo foi anteriormente reportado por Xu e colaboradores (2013). Existe uma grande possibilidade dessas alterações serem fruto de edição do RNA e não serem mutações somáticas. A proteína ADAR é responsável pela conversão de uma adenosina em inosina em RNAs. Durante a síntese do cDNA, a inosina será pareada a uma citosina (Brenda, 2002), provocando uma alteração de AT>GC. Assim como no presente trabalho, já havia sido relatado o aumento da expressão de ADAR em CEE (Qin *et al.*, 2014). Esse conjunto de informações nos levou a desconsiderar essas alterações da análise por serem de baixa confiabilidade. Além disso, as alterações do tipo INDEL também foram retiradas da análise uma vez que o protocolo escolhido para o presente trabalho foi o proposto pelo *Broad Institute* e este não recomenda a avaliação de INDEL.

Como foram analisados apenas as variações presente nos transcritos apenas aquelas provenientes de *reads* que alinharam em exons foram consideradas. Além disso, como o interesse era em mutações somáticas todas as alterações presentes em bancos públicos de SNPs foram excluídas. Deste modo foram considerados 883 genes mutados, praticamente o mesmo número de genes (884) encontrados por Agrawal e colaboradores (2012), único trabalho, antes desse, que avaliou amostras de CEE por RNA-seq em população ocidental.

O gene com maior frequência de mutação entre as amostras foi o *TP53* com quase 80% das amostras mutadas. Por essa razão esse gene foi validado em outras amostras por sequenciamento do gene inteiro por NGS, alcançando cerca de 70% das amostras mutadas que foi menor que a frequência reportada por Zhang e colaboradores (2015) e Sawada e colaboradores (2016), mas foi igual a frequência relatada por Agrawal e colaboradores (2012), Lin e colaboradores (2014) e a disponível no *International Cancer Genome Consortium* (ICGC). Os exons mais mutados foram os de 5 a 8 (mais de 90%) que correspondem ao domínio de ligação com ao DNA (Okorokov *et al.*, 2009). As semelhanças com os dados do ICGC para o *TP53* uma vez que os outros genes apontados na figura 4.16 não estão nem entre os 20 mais frequentes no ICGC, demonstrando a grande heterogeneidade intertumoral do CEE.

Como dito anteriormente outras possibilidades além da expressão gênica são passíveis de serem avaliadas, como a presença de transcritos quiméricos. Esses podem advir de fusões gênicas e no presente trabalho foram encontrados 21 transcritos quiméricos em sete tumores, não tendo sido identificados nenhum evento na mucosa adjacente ao tumor. Além disso, o mesmo transcrito quimérico não foi encontrado em mais de um tumor. A fusão gênica reportada por Jiang e colaboradores (2014) não figurava entre as 21 encontradas neste trabalho. Além disso, nenhuma das fusões foram encontradas no *FusionCancer Database* (Wang *et al.*, 2015) e no *Mitelman Database of Chromosome Aberrations and Gene Fusions in Cancer* (Mitelman, Johansson e Mertens, 2016), dois bancos públicos de fusões gênicas.

A expressão do gene *ATR* foi maior entre os tumores que apresentavam algum transcrito quimérico. A proteína codificada pelo gene *ATR* é uma proteína sensora de dano no DNA, mais especificamente aqueles que geram bloqueio da forquilha de replicação (López-Contreras, Fernandez-Capetillo, 2010). Isso sugere que as amostras que contém fusões estão provavelmente sob maior estresse genotóxico.

Analisados em conjunto esses resultados sugerem que o fenômeno de fusão gênica seja consequência da tumorigênese e não o inverso.

Portanto, nesse trabalho foi possível notar que com a metodologia de sequenciamento de RNA (RNA-seq) foi possível identificar 522 GDEs em mucosa não-tumoral adjacente ao tumor quando comparada com a mucosa de indivíduos sem câncer, assim como foi possível identificar 1.226 GDEs em tumores de esôfago

quando comparados com as mucosas adjacentes ao tumor. Dentre as principais vias contendo esses genes diferencialmente expressos, destacaram-se as vias de inflamação e de sinalização de WNT. Importante ressaltar que alguns membros da via de WNT apresentaram valor prognóstico para indivíduos com CEE. Além disso foi feita a análise de mutações nos transcritos e o gene com maior frequência de mutação em CEE foi o *TP53* com quase 70% das amostras mutadas, após validação. Após avaliação das fusões gênicas, não foram encontrados eventos relevantes e o gene *ATR* foi associado a presença de fusões gênicas em CEE. Em conjunto esses dados mostram que ainda há um grande caminho a se percorrer no entendimento da tumorigênese da mucosa esofágica.

## 6 . CONCLUSÕES

- Foram encontrados 522 genes classificados como diferencialmente expressos (GDEs) entre mucosa não tumoral adjacente ao CEE e a mucosa de indivíduos sadios e 1.226 GDEs entre o CEE e a mucosa não tumoral adjacente ao tumor;

- Os principais processos enriquecidos encontrados no GO e em KEGG foram os processos relacionados a inflamação e a via de sinalização por WNT;

- Foram encontrados 883 genes mutados exclusivamente no tumor sendo o *TP53* o mais frequentemente mutado entre as amostras (78,6%) e o segundo mais mutado foi um pseudogene *LOC389831* (42,9%) ;

- Dos 7 genes selecionados para validação por PCRq em 28 pares de amostras independentes, o gene *MAPK3* foi validado com expressão reduzida no tumor em comparação com a mucosa adjacente enquanto os genes *WNT7B*, *IL8* e *FZD6* tiveram suas expressões aumentadas no tumor quando comparadas a mucosa adjacente. Os genes *WNT16*, *PRKCQ* e *WNK2* não puderam ser validados ;

- O gene *TP53* foi selecionado para validação em 42 amostras de CEE independentes e foi classificado como mutado em 69% das amostras;

- Foram encontrados 21 fusões em 7 tumores de CEE e nenhuma fusão foi encontrada em mucosa adjacente ao tumor;

- Após análise de sobrevida, a maior expressão do gene *WNT7B* apresentou pior prognóstico em CEE.

## 7. REFERÊNCIAS

Agrawal N, Jiao Y, Bettegowda C, Hutfless SM, Wang Y, David S, et al. Comparative genomic analysis of esophageal adenocarcinoma and squamous cell carcinoma. *Cancer Discov* [Internet]. 2012;2:899–905. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/22877736>

Anders S, Pyl PT, Huber W. HTSeq--a Python framework to work with high-throughput sequencing data. *Bioinformatics* [Internet]. 2015;31:166–9. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/25260700>

Ardolino M, Azimi CS, Iannello A, Trevino TN, Horan L, Zhang L, et al. Cytokine therapy reverses NK cell anergy in MHC-deficient tumors. *J Clin Invest* [Internet]. 2014;124:4781–94. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/25329698>

Aruga J, Nozaki Y, Hatayama M, Odaka YS, Yokota N. Expression of ZIC family genes in meningiomas and other brain tumors. *BMC Cancer* [Internet]. 2010;10:79. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/20199689>

Avery-Cooper G, Doerr M, Gilbert RW, Youssef M, Richard A, Huether P, et al. Par6 is an essential mediator of apoptotic response to transforming growth factor beta in NMuMG immortalized mammary cells. *Cancer Cell Int* [Internet]. 2014;14:19. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/24581220>

Baselga J. Targeting the epidermal growth factor receptor with tyrosine kinase inhibitors: small molecules, big hopes. *J Clin Oncol* [Internet]. 2002;20:2217–9. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/11980990>

Bass BL. RNA E <scp>DITING BY</scp> A <scp>DENOSINE</scp> D <scp>EAMINASES</scp> T <scp>HAT</scp> A <scp>CT ON</scp> RNA. *Annu Rev Biochem* [Internet]. 2002;71:817–46. Available from: <http://www.annualreviews.org/doi/abs/10.1146/annurev.biochem.71.110601.135501>

Benzoubir N, Mussini C, Lejamtel C, Dos Santos A, Guillaume C, Desterke C, et al. Gamma-smooth muscle actin expression is associated with epithelial-mesenchymal transition and stem-like properties in hepatocellular carcinoma. *PLoS One* [Internet]. 2015;10:e0130559. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/26110787>

Biswas NK, Das S, Maitra A, Sarin R, Majumder PP. Somatic mutations in arachidonic acid metabolism pathway genes enhance oral cancer post-treatment disease-free survival. *Nat Commun* [Internet]. 2014;5:5835. Available from: <http://www.nature.com/doi/abs/10.1038/ncomms6835>

Bjartell A, Johansson R, Björk T, Gadaleanu V, Lundwall A, Lilja H, et al.

Immunohistochemical detection of cysteine-rich secretory protein 3 in tissue and in serum from men with cancer or benign enlargement of the prostate gland. *Prostate* [Internet]. 2006;66:591–603. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/16388501>

Bose R, Wrana JL. Regulation of Par6 by extracellular signals. *Curr Opin Cell Biol* [Internet]. 2006;18:206–12. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/16490351>

Bras J, Guerreiro R, Hardy J. Use of next-generation sequencing and other whole-genome strategies to dissect neurological disease. *Nat Rev Neurosci* [Internet]. 2012;13:453–64. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/22714018>

Bullinger L, Schlenk RF, Götz M, Botzenhardt U, Hofmann S, Russ AC, et al. PRAME-induced inhibition of retinoic acid receptor signaling-mediated differentiation—a possible target for ATRA response in AML without t(15;17). *Clin Cancer Res* [Internet]. 2013;19:2562–71. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/23444226>

Byrne FL, Poon IKH, Modesitt SC, Tomsig JL, Chow JDY, Healy ME, et al. Metabolic vulnerabilities in endometrial cancer. *Cancer Res* [Internet]. 2014;74:5832–45. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/25205105>

Cantile M, Franco R, Tschan A, Baumhoer D, Zlobec I, Schiavo G, et al. HOX D13 expression across 79 tumor tissue types. *Int J cancer* [Internet]. 2009;125:1532–41. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/19488988>

Capdevila J, Elez E, Macarulla T, Ramos FJ, Ruiz-Echarri M, Tabernero J. Anti-epidermal growth factor receptor monoclonal antibodies in cancer treatment. *Cancer Treat Rev* [Internet]. 2009;35:354–63. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/19269105>

Castelletto R, Castellsague X, Muñoz N, Iscovich J, Chopita N, Jmelnitsky A. Alcohol, tobacco, diet, mate drinking, and esophageal cancer in Argentina. *Cancer Epidemiol Biomarkers Prev* [Internet]. 1994;3:557–64. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/7827585>

Castellsagué X, Muñoz N, De Stefani E, Victora CG, Castelletto R, Rolón PA. Influence of mate drinking, hot beverages and diet on esophageal cancer risk in South America. *Int J cancer* [Internet]. 2000;88:658–64. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/11058886>

Castellsagué X, Muñoz N, De Stefani E, Victora CG, Castelletto R, Rolón PA, et al. Independent and joint effects of tobacco smoking and alcohol drinking on the risk of esophageal cancer in men and women. *Int J cancer* [Internet]. 1999;82:657–64. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/10417762>

Chen B, Zhang C, Dong P, Guo Y, Mu N. Molecular regulation of cervical cancer growth and invasion by VEGFa. *Tumour Biol* [Internet]. 2014;35:11587–93. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/25135429>

Chen H, Zhou X, Shi Y, Yang J. Roles of p38 MAPK and JNK in TGF- $\beta$ 1-induced Human Alveolar Epithelial to Mesenchymal Transition. *Arch Med Res* [Internet]. 2013;44:93–8. Available from: <http://linkinghub.elsevier.com/retrieve/pii/S0188440913000386>

Cohen DJ, Ajani J. An expert opinion on esophageal cancer therapy. *Expert Opin Pharmacother* [Internet]. 2011;12:225–39. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/21226634>

Del Vecchio M, Bajetta E, Canova S, Lotze MT, Wesa A, Parmiani G, et al. Interleukin-12: biological properties and clinical application. *Clin Cancer Res* [Internet]. 2007;13:4677–85. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/17699845>

Deng F, Zhou K, Cui W, Liu D, Ma Y. Clinicopathological significance of wnt/ $\beta$ -catenin signaling pathway in esophageal squamous cell carcinoma. *Int J Clin Exp Pathol* [Internet]. 2015;8:3045–53. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/26045816>

Duggan SP, Gallagher WM, Fox EJP, Abdel-Latif MM, Reynolds J V, Kelleher D. Low pH induces co-ordinate regulation of gene expression in oesophageal cells. *Carcinogenesis* [Internet]. 2006;27:319–27. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/16113055>

Eckert RL, Adhikary G, Young CA, Jans R, Crish JF, Xu W, et al. AP1 transcription factors in epidermal differentiation and skin cancer. *J Skin Cancer* [Internet]. 2013;2013:537028. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/23762562>

Ferlay J, Soerjomataram I, Dikshit R, Eser S, Mathers C, Rebelo M, et al. Cancer incidence and mortality worldwide: sources, methods and major patterns in GLOBOCAN 2012. *Int J cancer* [Internet]. 2015;136:E359–86. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/25220842>

Freeman RK, Ascoti AJ, Mahidhara RJ. Palliative Therapy for Patients with Unresectable Esophageal Carcinoma. *Surg Clin North Am* [Internet]. 2012;92:1337–51. Available from: <http://linkinghub.elsevier.com/retrieve/pii/S0039610912001296>

Fry AM, O'Regan L, Sabir SR, Bayliss R. Cell cycle regulation by the NEK family of protein kinases. *J Cell Sci* [Internet]. 2012;125:4423–33. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/23132929>

Ge H, Liu K, Juan T, Fang F, Newman M, Hoeck W. FusionMap: detecting fusion genes from next-generation sequencing data at base-pair resolution. *Bioinformatics* [Internet]. 2011;27:1922–8. Available from: <http://bioinformatics.oxfordjournals.org/cgi/doi/10.1093/bioinformatics/btr310>

Gonzaga IM, Soares-Lima SC, de Santos PTS, Blanco TCM, de Reis BSB, Quintella DC, et al. Alterations in epidermal growth factor receptors 1 and 2 in esophageal squamous cell carcinomas. *BMC Cancer* [Internet]. 2012;12:569. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/23207070>

Lapouge G, Millon R, Muller D, Abecassis J, Eber M, Bergerat JP, et al. Cisplatin-induced genes as potential markers for thyroid cancer. *Cell Mol Life Sci* [Internet]. 2005;62:53–64. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/15619007>

Levy D, de Almeida LM, Szklo A. The Brazil SimSmoke Policy Simulation Model: The Effect of Strong Tobacco Control Policies on Smoking Prevalence and Smoking-Attributable Deaths in a Middle Income Nation. Samet JM, editor. *PLoS Med* [Internet]. 2012;9:e1001336. Available from: <http://dx.plos.org/10.1371/journal.pmed.1001336>

López-Contreras AJ, Fernandez-Capetillo O. The ATR barrier to replication-born DNA damage. *DNA Repair (Amst)* [Internet]. 2010;9:1249–55. Available from: <http://linkinghub.elsevier.com/retrieve/pii/S1568786410003174>

Love MI, Huber W, Anders S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol* [Internet]. 2014;15:550. Available from: <http://genomebiology.com/2014/15/12/550>

Qin Y-R, Qiao J-J, Chan THM, Zhu Y-H, Li F-F, Liu H, et al. Adenosine-to-Inosine RNA Editing Mediated by ADARs in Esophageal Squamous Cell Carcinoma. *Cancer Res* [Internet]. 2014;74:840–51. Available from: <http://cancerres.aacrjournals.org/cgi/doi/10.1158/0008-5472.CAN-13-2545>

Reed CE. Comparison of different treatments for unresectable esophageal cancer. *World J Surg* [Internet]. 1995;19:828–35. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/8553674>

Ren P, Zhang J-G, Xiu L, Yu Z-T. Clinical significance of phospholipase A2 group IIA (PLA2G2A) expression in primary resected esophageal squamous cell carcinoma. *Eur Rev Med Pharmacol Sci* [Internet]. 2013;17:752–7. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/23609358>

Song Y, Li L, Ou Y, Gao Z, Li E, Li X, et al. Identification of genomic alterations in oesophageal squamous cell cancer. *Nature* [Internet]. 2014;509:91–5. Available

from: <http://www.ncbi.nlm.nih.gov/pubmed/24670651>

Teh M-T, Blaydon D, Ghali LR, Edmunds S, Pantazi E, Barnes MR, et al. Role for WNT16B in human epidermal keratinocyte proliferation and differentiation. *J Cell Sci* [Internet]. 2007;120:330–9. Available from: <http://jcs.biologists.org/cgi/doi/10.1242/jcs.03329>

Titz B, Sewer A, Schneider T, Elamin A, Martin F, Dijon S, et al. Alterations in the sputum proteome and transcriptome in smokers and early-stage COPD subjects. *J Proteomics* [Internet]. 2015;128:306–20. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/26306861>

Tong M, Chan KW, Bao JYJ, Wong KY, Chen J-N, Kwan PS, et al. Rab25 is a tumor suppressor gene with antiangiogenic and anti-invasive activities in esophageal squamous cell carcinoma. *Cancer Res* [Internet]. 2012;72:6024–35. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/22991305>

Untergasser G, Gander R, Lilg C, Lepperdinger G, Plas E, Berger P. Profiling molecular targets of TGF-beta1 in prostate fibroblast-to-myofibroblast transdifferentiation. *Mech Ageing Dev* [Internet]. 2005;126:59–69. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/15610763>

Wakita A, Motoyama S, Sato Y, Koyota S, Usami S, Yoshino K, et al. REG Iα activates c-Jun through MAPK pathways to enhance the radiosensitivity of squamous esophageal cancer cells. *Tumour Biol* [Internet]. 2015;36:5249–54. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/25656613>

Wang K, Li M, Hakonarson H. ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data. *Nucleic Acids Res* [Internet]. 2010;38:e164–e164. Available from: <http://nar.oxfordjournals.org/lookup/doi/10.1093/nar/gkq603>

Wang Y, Wu N, Liu J, Wu Z, Dong D. FusionCancer: a database of cancer fusion genes derived from RNA-seq data. *Diagn Pathol* [Internet]. 2015;10:131. Available from: <http://www.diagnosticpathology.org/content/10/1/131>

Wu Q-L, Zierold C, Ranheim EA. Dysregulation of Frizzled 6 is a critical component of B-cell leukemogenesis in a mouse model of chronic lymphocytic leukemia. *Blood* [Internet]. 2009;113:3031–9. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/19179304>

Xu X, Zhu K, Liu F, Wang Y, Shen J, Jin J, et al. Identification of somatic mutations in human prostate cancer by RNA-Seq. *Gene* [Internet]. 2013;519:343–7. Available from: <http://linkinghub.elsevier.com/retrieve/pii/S0378111913001194>

Yang Y, Liao Q, Wei F, Li X, Zhang W, Fan S, et al. LPLUNC1 inhibits

nasopharyngeal carcinoma cell growth via down-regulation of the MAP kinase and cyclin D1/E2F pathways. *PLoS One* [Internet]. 2013;8:e62869. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/23650533>

Yeo E-J, Cassetta L, Qian B-Z, Lewkowich I, Li J -f., Stefater JA, et al. Myeloid WNT7b Mediates the Angiogenic Switch and Metastasis in Breast Cancer. *Cancer Res* [Internet]. 2014;74:2962–73. Available from: <http://cancerres.aacrjournals.org/cgi/doi/10.1158/0008-5472.CAN-13-2421>

Arul GS, Moorghen M, Myerscough N, Alderson DA, Spicer RD, Corfield AP. Mucin gene expression in Barrett's oesophagus: an in situ hybridisation and immunohistochemical study. *Gut* [Internet]. 2000;47:753–61. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/11076872>

Bernardo V, Ribeiro Pinto LF, Albano RM. Gene expression analysis by real-time PCR: experimental demonstration of PCR detection limits. *Anal Biochem* [Internet]. 2013;432:131–3. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/23026780>

Chen LQ, Hu CY, Ghadirian P, Duranceau A. Early detection of esophageal squamous cell carcinoma and its effects on therapy: an overview. *Dis Esophagus* [Internet]. 1999;12:161–7. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/10631905>

da Costa NM, Soares Lima SC, de Almeida Simão T, Ribeiro Pinto LF. The potential of molecular markers to improve interventions through the natural history of oesophageal squamous cell carcinoma. *Biosci Rep* [Internet]. 2013;33. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/23837802>

Dawsey SM, Shen Q, Nieberg RK, Liu SF, English SA, Cao J, et al. Studies of esophageal balloon cytology in Linxian, China. *Cancer Epidemiol Biomarkers Prev* [Internet]. 1997;6:121–30. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/9037563>

Dobin A, Davis CA, Schlesinger F, Drenkow J, Zaleski C, Jha S, et al. STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* [Internet]. 2013;29:15–21. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/23104886>

Fu L, Zhang C, Zhang L-Y, Dong S-S, Lu L-H, Chen J, et al. Wnt2 secreted by tumour fibroblasts promotes tumour progression in oesophageal cancer by activation of the Wnt/ $\beta$ -catenin signalling pathway. *Gut* [Internet]. 2011;60:1635–43. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/21672941>

Fujishima F, Suzuki T, Nakamura Y, Taniyama Y, Ono K, Sugawara A, et al. Retinoid receptors in human esophageal squamous cell carcinoma: retinoid X receptor as a potent prognostic factor. *Pathol Int* [Internet]. 2011;61:401–8. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/21707843>

Gao Y-B, Chen Z-L, Li J-G, Hu X-D, Shi X-J, Sun Z-M, et al. Genetic landscape of esophageal squamous cell carcinoma. *Nat Genet* [Internet]. 2014;46:1097–102. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/25151357>

Godoy A, Ulloa V, Rodríguez F, Reinicke K, Yañez AJ, García M de los A, et al. Differential subcellular distribution of glucose transporters GLUT1-6 and GLUT9 in human cancer: ultrastructural localization of GLUT1 and GLUT5 in breast tumor tissues. *J Cell Physiol* [Internet]. 2006;207:614–27. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/16523487>

Grinberg I, Millen KJ. The ZIC gene family in development and disease. *Clin Genet* [Internet]. 2005;67:290–6. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/15733262>

Grosshans H, Filipowicz W. Molecular biology: the expanding world of small RNAs. *Nature* [Internet]. 2008;451:414–6. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/18216846>

Hanahan D, Weinberg RA. Hallmarks of cancer: the next generation. *Cell* [Internet]. 2011;144:646–74. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/21376230>

Harari PM. Epidermal growth factor receptor inhibition strategies in oncology. *Endocr Relat Cancer* [Internet]. 2004;11:689–708. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/15613446>

Harjes U, Bridges E, McIntyre A, Fielding BA, Harris AL. Fatty acid-binding protein 4, a point of convergence for angiogenic and metabolic signaling pathways in endothelial cells. *J Biol Chem* [Internet]. 2014;289:23168–76. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/24939870>

Haswell LE, Hewitt K, Thorne D, Richter A, Gaça MD. Cigarette smoke total particulate matter increases mucous secreting cell numbers in vitro: a potential model of goblet cell hyperplasia. *Toxicol In Vitro* [Internet]. 2010;24:981–7. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/20060463>

Herroon MK, Rajagurubandara E, Hardaway AL, Powell K, Turchick A, Feldmann D, et al. Bone marrow adipocytes promote tumor growth in bone via FABP4-dependent mechanisms. *Oncotarget* [Internet]. 2013;4:2108–23. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/24240026>

Hongo M, Nagasaki Y, Shoji T. Epidemiology of esophageal cancer: Orient to Occident. Effects of chronology, geography and ethnicity. *J Gastroenterol Hepatol* [Internet]. 2009;24:729–35. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/19646015>

Ismat A, Cheshire AM, Andrew DJ. The secreted AdamTS-A metalloprotease is required for collective cell migration. *Development* [Internet]. 2013;140:1981–93. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/23536567>

Jee HJ, Kim AJ, Song N, Kim H-J, Kim M, Koh H, et al. Nek6 overexpression antagonizes p53-induced senescence in human cancer cells. *Cell Cycle* [Internet]. 2010;9:4703–10. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/21099361>

Jemal A, Bray F, Center MM, Ferlay J, Ward E, Forman D. Global cancer statistics. *CA Cancer J Clin* [Internet]. 2011;61:69–90. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/21296855>

Jemal A, Siegel R, Ward E, Murray T, Xu J, Smigal C, et al. Cancer statistics, 2006. *CA Cancer J Clin* [Internet]. 2006;56:106–30. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/16514137>

Jiang P, Josue J, Li X, Glaser D, Li W, Brand JG, et al. Major taste loss in carnivorous mammals. *Proc Natl Acad Sci U S A* [Internet]. 2012;109:4956–61. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/22411809>

Jiang Y-Z, Li Q-H, Zhao J-Q, Lv J-J. Identification of a novel fusion gene (HLA-E and HLA-B) by RNA-seq analysis in esophageal squamous cell carcinoma. *Asian Pac J Cancer Prev* [Internet]. 2014;15:2309–12. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/24716975>

Kang H-W, Gupta S, Bogdanov A. Orthotopic expression of noggin protein in cancer cells inhibits human lung carcinoma growth in vivo. *Mol Imaging Biol* [Internet]. 2012;14:480–8. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/21913026>

Kania N, Setiawan B, Widjadjanto E, Nurdiana N, Widodo MA, Kusuma HMSC. Subchronic inhalation of coal dust particulate matter 10 induces bronchoalveolar hyperplasia and decreases MUC5AC expression in male Wistar rats. *Exp Toxicol Pathol Off J Gesellschaft für Toxikologische Pathol* [Internet]. 2014;66:383–9. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/24975055>

Kim D, Pertea G, Trapnell C, Pimentel H, Kelley R, Salzberg SL. TopHat2: accurate alignment of transcriptomes in the presence of insertions, deletions and gene fusions. *Genome Biol* [Internet]. 2013;14:R36. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/23618408>

Ko W-C, Sugahara K, Sakuma T, Yen C-Y, Liu S-Y, Liaw G-A, et al. Copy number changes of CRISP3 in oral squamous cell carcinoma. *Oncol Lett* [Internet]. 2012;3:75–81. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/22740859>

Lee D, Wada K, Taniguchi Y, Al-Shareef H, Masuda T, Usami Y, et al. Expression of fatty acid binding protein 4 is involved in the cell growth of oral squamous cell carcinoma. *Oncol Rep* [Internet]. 2014;31:1116–20. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/24425381>

Lee Z, Cheng C-T, Zhang H, Subler MA, Wu J, Mukherjee A, et al. Role of LPA4/p2y9/GPR23 in negative regulation of cell motility. *Mol Biol Cell* [Internet]. 2008;19:5435–45. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/18843048>

Li B, Li J, Xu WW, Guan XY, Qin YR, Zhang LY, et al. Suppression of esophageal tumor growth and chemoresistance by directly targeting the PI3K/AKT pathway. *Oncotarget* [Internet]. 2014;5:11576–87. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/25344912>

Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, et al. The Sequence Alignment/Map format and SAMtools. *Bioinformatics* [Internet]. 2009;25:2078–9. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/19505943>

Lijinsky W, Kovatch RM, Saavedra JE. Carcinogenesis and mutagenesis by N-nitroso compounds having a basic center. *Cancer Lett* [Internet]. 1992;63:101–7. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/1562986>

Lima SCS, Hernández-Vargas H, Simão T, Durand G, Krueel CDP, Le Calvez-Kelm F, et al. Identification of a DNA methylome signature of esophageal squamous cell carcinoma and potential epigenetic biomarkers. *Epigenetics* [Internet]. 2011;6:1217–27. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/21946330>

Lin D-C, Hao J-J, Nagata Y, Xu L, Shang L, Meng X, et al. Genomic and molecular characterization of esophageal squamous cell carcinoma. *Nat Genet* [Internet]. 2014;46:467–73. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/24686850>

Lin J-C, Chiang C-F, Wang S-W, Wang W-Y, Kwan P-C, Wu G-J. Significance of expression of human METCAM/MUC18 in nasopharyngeal carcinomas and metastatic lesions. *Asian Pac J Cancer Prev* [Internet]. 2014;15:245–52. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/24528033>

Liu C-W, Li C-H, Peng Y-J, Cheng Y-W, Chen H-W, Liao P-L, et al. Snail regulates Nanog status during the epithelial-mesenchymal transition via the Smad1/Akt/GSK3 $\beta$  signaling pathway in non-small-cell lung cancer. *Oncotarget* [Internet]. 2014;5:3880–94. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/25003810>

Liu L, Li Y, Li S, Hu N, He Y, Pong R, et al. Comparison of next-generation sequencing systems. *J Biomed Biotechnol* [Internet]. 2012;2012:251364. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/22829749>

Liu Q-W, Fu J-H, Luo K-J, Yang H-X, Wang J-Y, Hu Y, et al. Identification of EGFR and KRAS mutations in Chinese patients with esophageal squamous cell carcinoma. *Dis Esophagus* [Internet]. 2011;24:374–80. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/21615826>

Livak KJ, Schmittgen TD. Analysis of relative gene expression data using real-time quantitative PCR and the 2<sup>(-Delta Delta C(T))</sup> Method. *Methods* [Internet]. 2001;25:402–8. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/11846609>

Ma S, Bao JYJ, Kwan PS, Chan YP, Tong CM, Fu L, et al. Identification of PTK6, via RNA sequencing analysis, as a suppressor of esophageal squamous cell carcinoma. *Gastroenterology* [Internet]. 2012;143:675–86.e1–12. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/22705009>

Makino T, Yamasaki M, Miyata H, Yoshioka S, Takiguchi S, Fujiwara Y, et al. p53 Mutation status predicts pathological response to chemoradiotherapy in locally advanced esophageal cancer. *Ann Surg Oncol* [Internet]. 2010;17:804–11. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/19885698>

Malavasi F, Deaglio S, Funaro A, Ferrero E, Horenstein AL, Ortolan E, et al. Evolution and function of the ADP ribosyl cyclase/CD38 gene family in physiology and pathology. *Physiol Rev* [Internet]. 2008;88:841–86. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/18626062>

Marioni JC, Mason CE, Mane SM, Stephens M, Gilad Y. RNA-seq: an assessment of technical reproducibility and comparison with gene expression arrays. *Genome Res* [Internet]. 2008;18:1509–17. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/18550803>

Matayoshi S, Chiba S, Lin Y, Arakaki K, Matsumoto H, Nakanishi T, et al. Lysophosphatidic acid receptor 4 signaling potentially modulates malignant behavior in human head and neck squamous cell carcinoma cells. *Int J Oncol* [Internet]. 2013;42:1560–8. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/23467751>

Meyer D, Voigt A, Widmayer P, Borth H, Huebner S, Breit A, et al. Expression of Tas1 taste receptors in mammalian spermatozoa: functional role of Tas1r1 in regulating basal Ca<sup>2+</sup> and cAMP concentrations in spermatozoa. *PLoS One* [Internet]. 2012;7:e32354. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/22427794>

Moniz S, Veríssimo F, Matos P, Brazão R, Silva E, Kotelevets L, et al. Protein kinase WNK2 inhibits cell proliferation by negatively modulating the activation of MEK1/ERK1/2. *Oncogene* [Internet]. 2007;26:6071–81. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/17667937>

Moolenaar WH. Lysophospholipids in the limelight: autotaxin takes center stage. *J*

Cell Biol [Internet]. 2002;158:197–9. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/12135981>

Morgan S, Grandis JR. ErbB receptors in the biology and pathology of the aerodigestive tract. *Exp Cell Res* [Internet]. 2009;315:572–82. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/18778701>

Morone S, Lo-Buono N, Parrotta R, Giacomino A, Nacci G, Brusco A, et al. Overexpression of CD157 contributes to epithelial ovarian cancer progression by promoting mesenchymal differentiation. *PLoS One* [Internet]. 2012;7:e43649. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/22916288>

Nolan ME, Aranda V, Lee S, Lakshmi B, Basu S, Allred DC, et al. The polarity protein Par6 induces cell proliferation and is overexpressed in breast cancer. *Cancer Res* [Internet]. 2008;68:8201–9. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/18922891>

Ogura M, Takeuchi H, Kawakubo H, Nishi T, Fukuda K, Nakamura R, et al. Clinical significance of CXCL-8/CXCR-2 network in esophageal squamous cell carcinoma. *Surgery* [Internet]. 2013;154:512–20. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/23972657>

Okorokov AL, Orlova E V. Structural biology of the p53 tumour suppressor. *Curr Opin Struct Biol* [Internet]. 2009;19:197–202. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/19286366>

Olivier M, Eeles R, Hollstein M, Khan MA, Harris CC, Hainaut P. The IARC TP53 database: new online mutation analysis and recommendations to users. *Hum Mutat* [Internet]. 2002;19:607–14. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/12007217>

Pace-Asciak CR. Hepoxilins in cancer and inflammation--use of hepoxilin antagonists. *Cancer Metastasis Rev* [Internet]. 2011;30:493–506. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/22005952>

Ribeiro FS, Simão TA, Amoêdo ND, Andreollo NA, Lopes LR, Acatauassu R, et al. Evidence for increased expression of tissue factor and protease-activated receptor-1 in human esophageal cancer. *Oncol Rep* [Internet]. 2009;21:1599–604. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/19424642>

Rolón PA, Castellsagué X, Benz M, Muñoz N. Hot and cold mate drinking and esophageal cancer in Paraguay. *Cancer Epidemiol Biomarkers Prev* [Internet]. 1995;4:595–605. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/8547825>

Rossini A, de Almeida Simão T, Marques CB, Soares-Lima SC, Herbster S, Rapozo

DCM, et al. TP53 mutation profile of esophageal squamous cell carcinomas of patients from Southeastern Brazil. *Mutat Res* [Internet]. 2010;696:10–5. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/19944185>

Sahmoun AE, Singh BB. Does a higher ratio of serum calcium to magnesium increase the risk for postmenopausal breast cancer? *Med Hypotheses* [Internet]. 2010;75:315–8. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/20371155>

Sawada G, Niida A, Uchi R, Hirata H, Shimamura T, Suzuki Y, et al. Genomic Landscape of Esophageal Squamous Cell Carcinoma in a Japanese Population. *Gastroenterology* [Internet]. 2016;150:1171–82. Available from: <http://linkinghub.elsevier.com/retrieve/pii/S0016508516001347>

Schmieder R, Edwards R. Quality control and preprocessing of metagenomic datasets. *Bioinformatics* [Internet]. 2011;27:863–4. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/21278185>

Sequist L V, Joshi VA, Jänne PA, Bell DW, Fidias P, Lindeman NI, et al. Epidermal growth factor receptor mutation testing in the care of lung cancer patients. *Clin Cancer Res* [Internet]. 2006;12:4403s – 4408s. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/16857818>

Sergina N V, Moasser MM. The HER family and cancer: emerging molecular mechanisms and therapeutic targets. *Trends Mol Med* [Internet]. 2007;13:527–34. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/17981505>

Serysheva E, Berhane H, Grumolato L, Demir K, Balmer S, Bodak M, et al. Wnk kinases are positive regulators of canonical Wnt/ $\beta$ -catenin signalling. *EMBO Rep* [Internet]. 2013;14:718–25. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/23797875>

Shen F-F, Yue W-B, Zhou F-Y, Pan Y, Zhao X-K, Jin Y, et al. Variations in the MHC region confer risk to esophageal squamous cell carcinoma on the subjects from high-incidence area in northern China. *PLoS One* [Internet]. 2014;9:e90438. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/24595008>

SLAUGHTER DP, SOUTHWICK HW, SMEJKAL W. Field cancerization in oral stratified squamous epithelium; clinical implications of multicentric origin. *Cancer* [Internet]. 1953;6:963–8. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/13094644>

Slavin S, Yeh C-R, Da J, Yu S, Miyamoto H, Messing EM, et al. Estrogen receptor  $\alpha$  in cancer-associated fibroblasts suppresses prostate cancer invasion via modulation of thrombospondin 2 and matrix metalloproteinase 3. *Carcinogenesis* [Internet]. 2014;35:1301–9. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/24374826>

Song S, Guan B, Men T, Hoque A, Lotan R, Xu X-C. Antitumor effect of retinoic acid receptor-beta2 associated with suppression of cyclooxygenase-2. *Cancer Prev Res (Phila)* [Internet]. 2009;2:274–80. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/19258542>

Stein HJ, Brücher BL, Sendler A, Siewert JR. Esophageal cancer: patient evaluation and pre-treatment staging. *Surg Oncol* [Internet]. 2001;10:103–11. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/11750229>

Su H, Xu T, Ganapathy S, Shadfan M, Long M, Huang TH-M, et al. Elevated snoRNA biogenesis is essential in breast cancer. *Oncogene* [Internet]. 2014;33:1348–58. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/23542174>

Szabo C, Coletta C, Chao C, Módis K, Szczesny B, Papapetropoulos A, et al. Tumor-derived hydrogen sulfide, produced by cystathionine- $\beta$ -synthase, stimulates bioenergetics, cell proliferation, and angiogenesis in colon cancer. *Proc Natl Acad Sci U S A* [Internet]. 2013;110:12474–9. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/23836652>

Takikita M, Hu N, Shou J-Z, Giffen C, Wang Q-H, Wang C, et al. Fascin and CK4 as biomarkers for esophageal squamous cell carcinoma. *Anticancer Res* [Internet]. 2011;31:945–52. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/21498718>

Tao Y-P, Wang W-L, Li S-Y, Zhang J, Shi Q-Z, Zhao F, et al. Associations between polymorphisms in IL-12A, IL-12B, IL-12R $\beta$ 1, IL-27 gene and serum levels of IL-12p40, IL-27p28 with esophageal cancer. *J Cancer Res Clin Oncol* [Internet]. 2012;138:1891–900. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/22740240>

Trapnell C, Pachter L, Salzberg SL. TopHat: discovering splice junctions with RNA-Seq. *Bioinformatics* [Internet]. 2009;25:1105–11. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/19289445>

Udby L, Cowland JB, Johnsen AH, Sørensen OE, Borregaard N, Kjeldsen L. An ELISA for SGP28/CRISP-3, a cysteine-rich secretory protein in human neutrophils, plasma, and exocrine secretions. *J Immunol Methods* [Internet]. 2002;263:43–55. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/12009203>

Uraoka N, Oue N, Sakamoto N, Sentani K, Oo HZ, Naito Y, et al. NRD1, which encodes nardilysin protein, promotes esophageal cancer cell invasion through induction of MMP2 and MMP3 expression. *Cancer Sci* [Internet]. 2014;105:134–40. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/24168165>

Victora CG, Muñoz N, Day NE, Barcelos LB, Peccin DA, Braga NM. Hot beverages and oesophageal cancer in southern Brazil: a case-control study. *Int J cancer* [Internet]. 1987;39:710–6. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/3583451>

Wang C, Xiao Y, Hu Z, Chen Y, Liu N, Hu G. PEG10 directly regulated by E2Fs might have a role in the development of hepatocellular carcinoma. *FEBS Lett* [Internet]. 2008;582:2793–8. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/18625225>

Warburg O, Wind F, Negelein E. THE METABOLISM OF TUMORS IN THE BODY. *J Gen Physiol* [Internet]. 1927;8:519–30. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/19872213>

Warot X, Fromental-Ramain C, Fraulob V, Chambon P, Dollé P. Gene dosage-dependent effects of the Hoxa-13 and Hoxd-13 mutations on morphogenesis of the terminal parts of the digestive and urogenital tracts. *Development* [Internet]. 1997;124:4781–91. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/9428414>

Xuan W, Qu Q, Zheng B, Xiong S, Fan G-H. The chemotaxis of M1 and M2 macrophages is regulated by different chemokines. *J Leukoc Biol* [Internet]. 2015;97:61–9. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/25359998>

Yamasaki M, Miyata H, Fujiwara Y, Takiguchi S, Nakajima K, Nishida T, et al. p53 genotype predicts response to chemotherapy in patients with squamous cell carcinoma of the esophagus. *Ann Surg Oncol* [Internet]. 2010;17:634–42. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/19941080>

Yan X, Lin Y, Yang D, Shen Y, Yuan M, Zhang Z, et al. A novel anti-CD146 monoclonal antibody, AA98, inhibits angiogenesis and tumor growth. *Blood* [Internet]. 2003;102:184–91. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/12609848>

Yang M, Liu R, Sheng J, Liao J, Wang Y, Pan E, et al. Differential expression profiles of microRNAs as potential biomarkers for the early diagnosis of esophageal squamous cell carcinoma. *Oncol Rep* [Internet]. 2013;29:169–76. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/23124769>

Yuen H-F, McCrudden CM, Grills C, Zhang S-D, Huang Y-H, Chan K-K, et al. Combinatorial use of bone morphogenetic protein 6, noggin and SOST significantly predicts cancer progression. *Cancer Sci* [Internet]. 2012;103:1145–54. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/22364398>

Zhang L, Zhou Y, Cheng C, Cui H, Cheng L, Kong P, et al. Genomic Analyses Reveal Mutational Signatures and Frequently Altered Genes in Esophageal Squamous Cell Carcinoma. *Am J Hum Genet* [Internet]. 2015;96:597–611. Available from: <http://linkinghub.elsevier.com/retrieve/pii/S0002929715001007> (1–39,39–57,57–65,65–136)

