

**ESTUDO EM LARGA ESCALA DE ALTERAÇÕES
TRANSCRICIONAIS RELACIONADAS À
OCORRÊNCIA DE METÁSTASES NO CÂNCER ORAL**

FREDERICO OMAR GLEBER NETTO

**Tese apresentada à Fundação Antônio
Prudente para a obtenção do título de Doutor
em Ciências**

Área de concentração: Oncologia

Orientador: Dr. Emmanuel Dias-Neto

**São Paulo
2015**

FICHA CATALOGRÁFICA

Preparada pela Biblioteca da Fundação Antônio Prudente

Gleber Netto, Frederico Omar

Estudo em larga escala de alterações transcricionais relacionadas à ocorrência de metástases no câncer oral / Frederico Omar Gleber Netto - São Paulo, 2015.

169p.

Tese (Doutorado)-Fundação Antônio Prudente.

Curso de Pós-Graduação em Ciências - Área de concentração: Oncologia.

Orientador: Emmanuel Dias-Neto

Descritores: 1. NEOPLASIAS BUCAIS/complicações. 2. METÁSTASE NEOPLÁSICA. 3. CARCINOMA DE CÉLULAS ESCAMOSAS. 4. EXPRESSÃO GÊNICA. 5. SEQUENCIAMENTO DE NUCLEOTÍDEOS EM LARGA ESCALA

DEDICATÓRIA

Dedico este trabalho aos indivíduos acometidos pelo câncer de boca, que duelam não só com a expectativa da morte que é destino comum a todos, mas, sobremaneira, com a perspectiva da transfiguração decorrente do tratamento. Para estes a cura nem sempre representa a esperança de uma vida plena.

“A infeliz padecia de um modo cru, porque cancro é indiferente às virtudes do sujeito; quando rói, rói; roer é o seu officio.”

Memórias Póstumas de Brás Cubas - Machado de Assis

AGRADECIMENTOS

Aos pacientes que doaram voluntariamente seu material biológico, mesmo sabendo que os frutos deste trabalho não lhe trariam benefícios diretos.

Ao meu orientador, Dr. Emmanuel Dias-Neto e à Dra. Diana Noronha Nunes pela oportunidade de me dedicar a este trabalho, por terem me ajudado na realização do doutorado sanduíche e por terem oportunizado meu prosseguimento na carreira acadêmica.

Ao Dr. Luiz Paulo Kowalski por ter marcado a minha formação científica de forma rara, tanto na transmissão do conhecimento, quanto ao me oferecer diversas oportunidades de produção e crescimento.

À Ivana Márcia Alves Diniz por ter me incentivado à desbravar a estrada da ciência participando de cada decurso, dividindo o êxito e o revés, e assegurando que o caminho fosse trilhado com serenidade e sem ressaltos. Devo cada vitória na vida acadêmica ao seu companheirismo, altruísmo e estima.

À Melissa Pool Pizzi por ser uma excepcional companheira, assegurando minha resiliência mesmo no momento mais árduo de produção deste trabalho e da minha vida. Agradeço especialmente por me despertar o desejo de continuidade e a perseverança.

Às amigas Maria Galli de Amorim, Sheila Garcia, Bruna Barros, Ana Flávia de Mattos Costa por terem representado um importante alicerce no caminho trilhado durante o doutorado, provendo companheirismo, alegria e cooperação. E também a todos os demais colegas do CIPE pela agradável companhia.

Ao **Dr. Rodrigo Drummond, Dr. Israel Silva, Renan Valieris, Dr. Jorge Souza** e pelo trabalho incomensurável nas análises de bioinformática.

Aos colegas de laboratório **Andrew Thomas, Dra. Juliana Laino, Gabriela Branco e Dr. Gustavo Ribeiro.**

Ao **Dr. Hugo Campos, Eloisa Olivieri e Louise Mota** do Biobanco do *AC Camargo Cancer Center* pelo imenso auxílio na obtenção e processamento de amostras biológicas.

À **Dra. Dirce Carraro** por revisar e contribuir na evolução deste trabalho no decorrer do doutorado.

À **Dra. Elisa Napolitano** pelo grande auxílio na execução dos experimentos de preparação de bibliotecas e sequenciamento.

A todos membros do **GENCAPO** que participaram da elaboração deste projeto e permitiram a execução do mesmo da melhor forma possível. Agradeço especialmente à **Dra. Eloiza Tajara** por ter se esforçado ao máximo para que tudo necessário para realização do projeto estivesse acessível.

Ao **Dr. Clóvis Pinto, Dra. Cláudia Camillo e Ivete Fabelo** pelo grande auxílio nas questões referentes à patologia.

Aos funcionários da Pós-Graduação do *AC Camargo Cancer Center* e do CIPE, em nome da **Vanuza Rodrigues e Vanessa Dantas** que sempre estiveram prontas a ajudar.

Ao **Dr. David T. W. Wong** da *University of California Los Angeles*, por me receber em seu laboratório e permitir o desenvolvimento e a liderança de tantos projetos importantes.

Aos amigos do *Center for Oral/Head & Neck Oncology Research* **Dr. Shigeo Ishikawa, Dra. Maha Yakob, Dra. Julie Kanjanapangka, Blanca Majem, David Akin, Samantha Chiang, Xinmin Yan, Jie Sun, Michael Tu, Dr. Yong Zhang e Dr. Noe Gómez.**

Ao **Dr. Elmer Fernandez** e à **Gabriela Merino** da *Universidad Católica de Córdoba* que participaram das etapas de análises de dados.

À **Dra. Maria** e à **Dra. Nadir** por terem me dado suporte essencial ao desenvolvimento deste trabalho.

À **FAPESP** que subsidiou este trabalho através dos processos 2011/13315-4 e 2013/09142-2.

Em especial, agradeço à minha família, **Celso, Vera, Juliana, Lúcia, Emerson, Elizabeth e Elaine**, por não medirem esforços ao me prover todos os meios para que eu alcançasse o crescimento intelectual e profissional. Sem estas pessoas esta tese não seria possível.

RESUMO

Gleber-Netto FO. **Estudo em larga escala de alterações transcricionais relacionadas à ocorrência de metástases no câncer oral.** São Paulo; 2015. [Tese de Doutorado-Fundação Antônio Prudente]

O carcinoma epidermóide de boca (CEB) é uma neoplasia agressiva, causadora de alta morbidade e mortalidade. Uma de suas características clínicas mais importantes é a disseminação tumoral levando a uma doença metastática nos linfonodos regionais. Entretanto, limitações relacionadas às tecnologias diagnósticas atualmente disponíveis fazem com que a detecção da metástase linfonodal ainda seja um grande desafio clínico. Assim, melhorias na capacidade de predição de potencial metastático e/ou na detecção precoce e precisa da doença linfonodal possibilitariam determinar quais pacientes, de fato, se beneficiariam da abordagem cirúrgica cervical. Ao mesmo tempo, preservaria os portadores de tumores menos agressivos (não-metastáticos), reduzindo a morbidade e levando à maior sobrevida. Neste estudo investigamos padrões de expressão gênica tumoral determinados a partir do sequenciamento em larga escala do transcriptoma, comparando, por RNA-Seq, 10 CEB pequenos, mas já metastáticos (T1N+), com 10 CEB grandes, mas não-metastáticos (T3/T4N0). Análises de bioinformática sugeriram 150 genes diferencialmente expressos (DE) com potencial de separar os dois grupos de estudo. O valor preditivo destes genes foi desafiado *in silico* com dados do *The Cancer Genome Atlas* (TCGA), seguido de validação técnica nas mesmas amostras sequenciadas e, finalmente, validação biológica a partir de amostras independentes de CEB. Nossos resultados indicaram o *NR3C2* como um potencial biomarcador de metástase nodal e sugerem que a regulação de canais iônicos pode ter um papel importante no processo metastático do CEB. Embora mais estudos sejam necessários, nossos resultados indicam que no futuro marcadores transcricionais poderão ser usados no diagnóstico e predição da ocorrência de doença metastática linfonodal em pacientes com CEB, promovendo melhorias na abordagem terapêutica.

SUMMARY

Gleber-Netto FO. **[Large scale study of transcriptional alterations related to lymph node metastasis in OSCC]**. São Paulo; 2015. [Tese de Doutorado-Fundação Antônio Prudente]

Oral squamous cell carcinoma (OSCC) is an aggressive disease related to high morbidity and mortality. Lymph node metastasis is one of the main prognostic factors for OSCC patients, but no reliable diagnostic methods are available to accurately diagnose metastatic disease. Improvements in diagnosis of metastasis would enhance the indication of neck surgery for lymph nodes resection, sparing patients with locally restricted disease. In this study, we aimed to investigate if differences in the pattern of gene expression would help to diagnose metastatic OSCC. For this, we sequenced the transcriptome of 10 small (T1) metastatic and 10 large (T3-T4) non-metastatic OSCC aiming to determine differentially expressed transcripts (DET). Our results showed 150 DET with potential to separate metastatic and non-metastatic OSCC cases. This set of 150 genes were challenged through an *in silico* approach using data from The Cancer Genome Atlas (TCGA). Validated genes were then confirmed by RT-qPCR using the same set of sequenced samples followed by another confirmation in an independent set of OSCC cases. The transcript *NR3C2* was found as a putative biomarker for lymph node metastasis. Pathway analysis indicated that transcriptional alterations related to ion channels pathways may be involved in the biology of metastasis in OSCC. Although our results are not definitive, it indicates that transcriptional markers would be useful to diagnose or predict lymph node metastatic disease in OSCC patients, potentially improving therapeutic approaches.

LISTA DE FIGURAS

Figura 1	Estimativas de incidência de CEB a cada 100 mil habitantes para o ano de 2014.....	3
Figura 2	Incidência de CEB entre os anos de 2000 e 2005 na cidade de São Paulo de acordo com o gênero e faixa etária ao diagnóstico.....	5
Figura 3	Estruturas anatômicas da cavidade bucal	6
Figura 4	Fluxo de surgimento e progressão do CEB	24
Figura 5	Dados de sobrevida relativa em 5 anos obtidos do SEER-NIH no período entre 1975 e 2007 para CEB de língua, assoalho bucal e outros sítios da mucosa oral.....	26
Figura 6	Eletroferogramas demonstrando o processamento de uma amostra de RNA durante a construção de bibliotecas.....	56
Figura 7	Estrutura final da biblioteca de cDNA construída a partir do <i>Total RNA Seq Kit</i>	57
Figura 8	Representação esquemática da ePCR.....	59
Figura 9	Química de sequenciamento da plataforma de sequenciamento SOLiD.....	61
Figura 10	Representação do esquema de sequenciamento na plataforma SOLiD.....	62
Figura 11	Fórmulas utilizadas para cálculo dos níveis de expressão gênica.....	67
Figura 12	Distribuição do número de fragmentos sequenciados por amostra de acordo com o grupo de estudo.....	74

Figura 13	Classificação em grupos dos 18815 transcritos identificados no sequenciamento.....	76
Figura 14	Classificação em subgrupos dos 18815 transcritos identificados no estudo de acordo com a classificação do Ensembl.....	78
Figura 15	<i>Heatmap</i> representando a expressão relativa dos 150 genes diferencialmente expressos de acordo com as amostras sequenciadas.....	83
Figura 16	Classificação em grupos dos 150 transcritos diferencialmente expressos entre tumores metastáticos e tumores não metastáticos.....	84
Figura 17	Classificação em subgrupos dos 150 genes diferencialmente expressos entre os grupos de estudo.....	84
Figura 18	Distribuição dos 150 transcritos diferencialmente expressos de acordo com o cromossomo de origem.....	85
Figura 19	Distribuição dos 150 genes diferencialmente expressos por região cromossômica de acordo com o <i>fold-change</i>	87
Figura 20	PCA utilizando o padrão de expressão dos 150 transcritos DE.....	94
Figura 21	Distribuição de genes DE deste estudo em relação às seleções TCGA_1, TCGA_2 e TCGA_3.....	97
Figura 22	Distribuição de genes DE deste estudo em relação às seleções TCGAn1 e TCGAn2.....	99
Figura 23	<i>Boxplots</i> referentes aos genes validados na seleção TCGAn1.....	101
Figura 24	PCA utilizando o padrão de expressão dos 7 genes validados <i>in silico</i>	102

Figura 25	Curvas ROC utilizando padrão de expressão dos 7 genes validados <i>in silico</i>	104
Figura 26	PCA utilizando padrão de expressão dos 3 genes com maior AUC.....	106
Figura 27	Padrão de expressão dos genes com maior AUC nos grupos N+ e N0.....	107
Figura 28	Correlação entre valores de expressão gênica obtidos no sequenciamento e no qRT-PCR.....	108
Figura 29	Padrão de expressão dos transcritos candidatos em uma amostra independente de CEB.....	110
Figura 30	Curvas ROC utilizando o padrão de expressão dos 3 genes candidatos em amostras independentes.....	112
Figura 31	Curvas de Kaplan-Meier geradas a partir da categorização dos pacientes do TCGAn1 de acordo com o status linfonodal e a expressão de <i>NR3C2</i>	115
Figura 32	Curvas de Kaplan-Meier geradas a partir da categorização dos pacientes da seleção TCGAn1 combinando o status linfonodal e a expressão de <i>NR3C2</i> e seus respectivos desmembramentos de acordo com o status linfonodal.....	117

LISTA DE TABELAS

Tabela 1	Genes diferencialmente expressos validados na seleção TCGA_3.....	98
Tabela 2	Genes diferencialmente expressos validados na seleção TCGAn1.....	100
Tabela 3	Análise ROC multivariada baseada nos 7 transcritos validados <i>in silico</i>	105
Tabela 4	Valores de sensibilidade e especificidade do teste utilizando o transcrito <i>NR3C2</i> de acordo com diferentes pontos de corte de expressão gênica medida em <i>fold change</i>	113
Tabela 5	Análise ROC multivariada baseada nos 3 transcritos candidatos nas amostras de validação biológica.....	114
Tabela 6	Vias biológicas hiperativadas em tumores metastáticos relacionadas ao transporte de íons pela membrana celular.....	123
Tabela 7	Vias biológicas hipoativadas em tumores metastáticos relacionadas ao transporte de íons pela membrana celular.....	124

LISTA DE QUADROS

Quadro 1	Classificação clínica do CEB.....	30
Quadro 2	Estadiamento clínico do CEB baseado no TNM.....	31
Quadro 3	Estudos que buscaram a determinação de transcritos relacionados à metástase nodal em CEB.....	42
Quadro 4	Transcritos diferencialmente hiperexpressos em tumores metastáticos encontrados em diferentes estudos.....	46
Quadro 5	Transcritos diferencialmente hipoexpressos em tumores metastáticos encontrados em diferentes estudos.....	47
Quadro 6	Ensaio <i>Taqman</i> utilizados para mensuração da expressão dos genes candidatos a biomarcadores e do controle endógeno.....	66
Quadro 7	Características clínicas dos pacientes cujas amostras foram avaliadas neste estudo.....	70
Quadro 8	Características patológicas dos tumores utilizados no sequenciamento.....	71
Quadro 9	Dados globais das corridas de sequenciamento.....	75
Quadro 10	Classificação dos transcritos sequenciados neste estudo.....	77
Quadro 11	Classes de transcritos mais e menos representadas nos 20 casos de CEB.....	80
Quadro 12	Genes DE deste estudo também observados em outros trabalhos.....	93
Quadro 13	Critérios de seleção para formação dos grupos de validação <i>in silico</i>	96

Quadro 14	Amostras utilizadas na validação técnica.....	107
Quadro 15	Características clínicas das amostras utilizadas na validação biológica.....	111

LISTA DE SIGLAS E ABREVIATURAS

ACCCC	<i>AC Camargo Cancer Center</i>
AJCC	<i>American Joint Committee on Cancer</i>
AUC	Área abaixo da curva
cDNA	DNA complementar
CEB	Carcinoma epidermóide de boca
CECP	Carcinoma epidermóide de cabeça e pescoço
CEP	Comitê de Ética em Pesquisa
Chr	Cromossomo
CID-10	Classificação Internacional de Doenças - 10
cpm	<i>Counts</i> por milhão
Ct	<i>Cycle threshold</i>
DE	Diferencialmente expressos
DPM	Desordens potencialmente malignas
ePCR	PCR em emulsão
FC	<i>Fold-change</i>
fdr	<i>False discovery ratio</i>
HPV	Papiloma vírus humano
HR	Risco relativo
IC	Intervalo de confiança
INCA	Instituto Nacional do Câncer
LNC	Não-codificador de proteína longo
M	Doença metastática à distância
MEC	Matriz extracelular
miRNA	MicroRNA
miscRNA	<i>Miscellaneous</i> RNA
mRNA	RNA mensageiro
N	Status dos linfonodos cervicais
N+	Linfonodo positivo
N0	Linfonodo negativo

ng	Nanogramas
NIH	<i>National Institutes of Health</i>
nt	Nucleotídeos
OR	<i>Odds ratio</i>
OSCC	<i>Oral squamous cell carcinoma</i>
PC	Codificador de proteína
PCA	Análise de Componentes Principais
PCR	Reação em cadeia da polimerase
PsG	Pseudogene
qRT-PCR	PCR quantitativo em tempo real
RE	Retículo endoplasmático
RIN	<i>RNA integrity number</i>
RM	Receptor mineralocorticoide
RNASeq	Sequenciamento massivo em paralelo do RNA
rRNA	RNA ribossomal
SEER	<i>Surveillance, Epidemiology and End Results Program</i>
SNC	Não-codificador de proteína curto
T	Tamanho do tumor primário
TCGA	<i>The Cancer Genome Atlas</i>
TEM	Transição epitélio-mesênquima
tRNA	RNA transportador
UCLA	<i>University of California Los Angeles</i>
UICC	<i>Union for International Cancer Control</i>

ÍNDICE

1	INTRODUÇÃO	1
1.1	Carcinoma epidermóide de boca.....	1
1.2	Epidemiologia	2
1.3	Características clínicas.....	5
1.4	Etiologia	8
1.5	Patogênese e progressão.....	15
1.6	Prognóstico.....	24
1.7	Metástase linfonodal.....	31
1.8	Análise transcricional de CEB metastáticos	40
2	OBJETIVOS	49
3	MATERIAIS E MÉTODOS	50
3.1	Aspectos Éticos	50
3.2	Coleta de amostras.....	51
3.3	Crterios de inclusão e exclusão de amostras para sequenciamento..	51
3.4	Extração de ácidos nucléicos	52
3.5	Preparo de bibliotecas de cDNA.....	53
3.6	PCR em emulsão e sequenciamento	57
3.7	Análise de bioinformática.....	62
3.8	Validação in silico	63
3.9	Determinação de genes candidatos	65
3.10	PCR em tempo real quantitativo.....	66
3.11	Análise estatística.....	67
4	RESULTADOS E DISCUSSÃO	68
4.1	Características clínico-patológicas das amostras sequenciadas	68
4.2	Resultados do sequenciamento do transcriptoma.....	72
4.3	Determinação dos genes diferencialmente expressos	80

4.4	Curva ROC dos genes candidatos	102
4.5	Validação Técnica	106
4.6	Validação biológica.....	109
5	CONCLUSÃO.....	125
6	PERSPECTIVAS FUTURAS.....	126
7	DOUTORADO SANDUÍCHE.....	128
8	REFERÊNCIAS BIBLIOGRÁFICAS	158

ANEXOS

Anexo 1 Carta de aprovação do Comitê de Ética em Pesquisa do
A.C. Camargo Cancer Center

Anexo 2 Trabalhos publicados durante o doutoramento

PREFÁCIO

Esta tese, apresentada ao Programa de Pós-Graduação em Oncologia da Fundação Antônio Prudente, foi desenvolvida tanto no Laboratório de Genômica Médica do Centro Internacional de Pesquisa A.C. Camargo Cancer Center (ACCCC), como também na *School of Dentistry, University of California Los Angeles* (UCLA) durante período de realização estágio de doutoramento sanduíche.

Inicialmente serão abordados os resultados gerados no ACCCC entre o período de agosto de 2011 - dezembro de 2013 e janeiro a agosto de 2015. Esses dados referem-se ao objetivo central desta tese que é a determinação de marcadores moleculares para o diagnóstico das metástases nodais em carcinomas de células escamosas de boca.

Em seguida, serão apresentados resultados referentes a um dos projetos desenvolvidos no *Center for Oral/Head & Neck Oncology Research* sob orientação do Dr. David Wong que dedica seus estudos ao desenvolvimento de biomarcadores salivares para diagnóstico de doenças humanas. Este trabalho será apresentado na forma de um artigo já submetido a um periódico. Tal trabalho objetiva a determinação de marcadores salivares para diagnóstico precoce do carcinoma epidermóide de boca.

Desta forma, serão apresentados nesta tese dois trabalhos que possuem como objetivo geral desenvolver biomarcadores de utilidade clínica para dois dos principais problemas relacionados ao câncer de boca: o diagnóstico precoce e a disseminação metastática linfonodal.

1 INTRODUÇÃO

1.1 CARCINOMA EPIDERMÓIDE DE BOCA

O carcinoma epidermóide de boca (CEB) é o tumor maligno mais comum da região de cabeça e pescoço. Esta neoplasia se desenvolve a partir da transformação do epitélio de revestimento que recobre a mucosa bucal, levando à proliferação dos ceratinócitos que rompem a lâmina basal e invadem as camadas teciduais subjacentes. O tumor se infiltra em padrões que variam de lençóis, cordões e ilhotas de ceratinócitos que exibem variado grau de atipia e podem invadir capilares sanguíneos ou linfáticos, além de permear estruturas neuronais. Tal padrão de crescimento leva ao envolvimento de múltiplas estruturas teciduais e, frequentemente, culmina na disseminação metastática para os linfonodos cervicais. Por tais características, o tratamento desta lesão é agressivo, exigindo o envolvimento de margens cirúrgicas amplas e regimes de radioterapia associados ou não à quimioterapia. Entretanto, mesmo empregando esta terapêutica radical que causa significativa morbidade, as taxas de recorrência são altas e a sobrevida desses pacientes ainda é baixa (WÜNSCH-FILHO 2002).

Deste modo, hoje existe uma grande necessidade de melhoria nos índices de sucesso no tratamento dos pacientes com CEB, o que justifica a

busca de um maior conhecimento sobre as particularidades da biologia deste tumor.

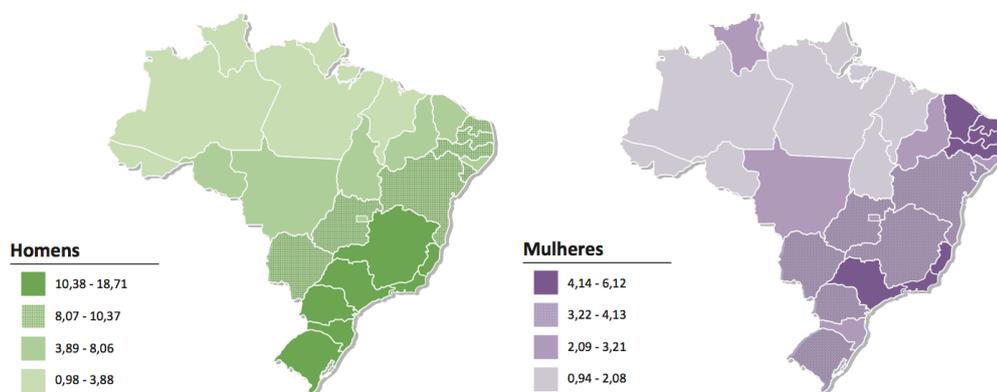
Neste trabalho buscamos avançar no conhecimento de aspectos moleculares do CEB visando melhorar a indicação terapêutica e diminuir a morbidade e mortalidade desses pacientes através do aperfeiçoamento na metodologia diagnóstica das disseminações metastáticas para os linfonodos cervicais.

1.2 EPIDEMIOLOGIA

O CEB representa cerca de 3% dos casos de câncer no mundo, sendo o oitavo tumor mais comum entre os homens. Estimam-se mais de 100.000 mortes causadas pelo CEB por ano e a maioria desses casos ocorre em países em desenvolvimento. O Brasil apresenta uma das maiores incidências deste tumor, principalmente devido ao aumento do etilismo e o alto consumo de tabaco (WÜNSCH-FILHO 2002; WARNAKULASURIYA 2009a; DE CAMARGO et al. 2010; JOHNSON et al. 2011).

De acordo com as estimativas do Instituto Nacional do Câncer (INCA) para 2014 cerca de 15290 novos casos de câncer de boca ocorreram no Brasil, representando o quinto câncer mais incidente entre os homens (excluindo carcinoma de pele não melanoma). No nosso país, o estado de São Paulo apresenta uma das maiores incidências, tanto entre os homens quanto entre as mulheres, representando cerca de 29% de todos os novos casos de CEB. Estimou-se que em 2014, 16,49 novos casos de CEB fossem

surgir a cada 100 mil homens e 4,74 novos casos a cada 100 mil mulheres (Figura 1) (Ministério da Saúde 2014).



Fonte: Ministério da Saúde (2014)

Figura 1 - Estimativas de incidência de CEB a cada 100 mil habitantes por estado da federação para o ano de 2014 de acordo com o gênero.

O gênero masculino é o mais afetado por este tumor (JOHNSON et al. 2011; SABA et al. 2011). PIRES et al. (2013) avaliaram 346 casos de CEB atendidos na Universidade do Estado do Rio de Janeiro e observaram que os homens representavam 67%, revelando uma relação de 2 homens acometidos para cada mulher. Dados do INCA de 2014 indicam que o risco estimado de desenvolvimento de CEB entre os homens é de 11,54 novos casos a cada 100 mil indivíduos, enquanto que entre as mulheres este risco é de 3,92 novos casos a cada 100 mil. Entretanto, WÜNSCH-FILHO (2002) ressalta que o consumo de tabaco entre mulheres brasileiras tem aumentado mais do que entre os homens, e possíveis mudanças neste cenário podem ocorrer no futuro. Neste sentido, WEIJERS et al. (2011),

avaliaram a incidência de casos CEB em uma instituição holandesa em dois intervalos distintos de tempo (entre 1980 e 1984 e entre 2000 e 2004), observando um aumento significativo da incidência de CEB entre mulheres nos anos mais recentes, provavelmente em decorrência da maior exposição recente das mulheres a fatores de risco.

Os principais indivíduos afetados são de meia idade e idosos, principalmente após a quinta década de vida (JOHNSON et al. 2011). Entre brasileiros, foi relatado que a idade média ao diagnóstico é em torno de 62 anos, semelhante ao observado em outras regiões do mundo (SABA et al. 2011; WIEJERS et al. 2011; JOHNSON et al. 2011; PIRES et al. 2013). Quando avalia-se os gêneros, individualmente, observa-se que homens são acometidos mais precocemente (59 anos x 67 anos) (PIRES et al. 2013). Essa tendência pode ser observada nos dados de incidência de CEB entre os anos de 2000 e 2005 na cidade de São Paulo (INCA) (**Figura 2**). Mais homens são diagnosticados entre a quinta e sétima décadas de vida, enquanto as mulheres apresentam picos de incidência entre a sexta e a oitava décadas de vida. Destaca-se que a incidência deste tumor entre a oitava e nona décadas é maior entre o gênero feminino.

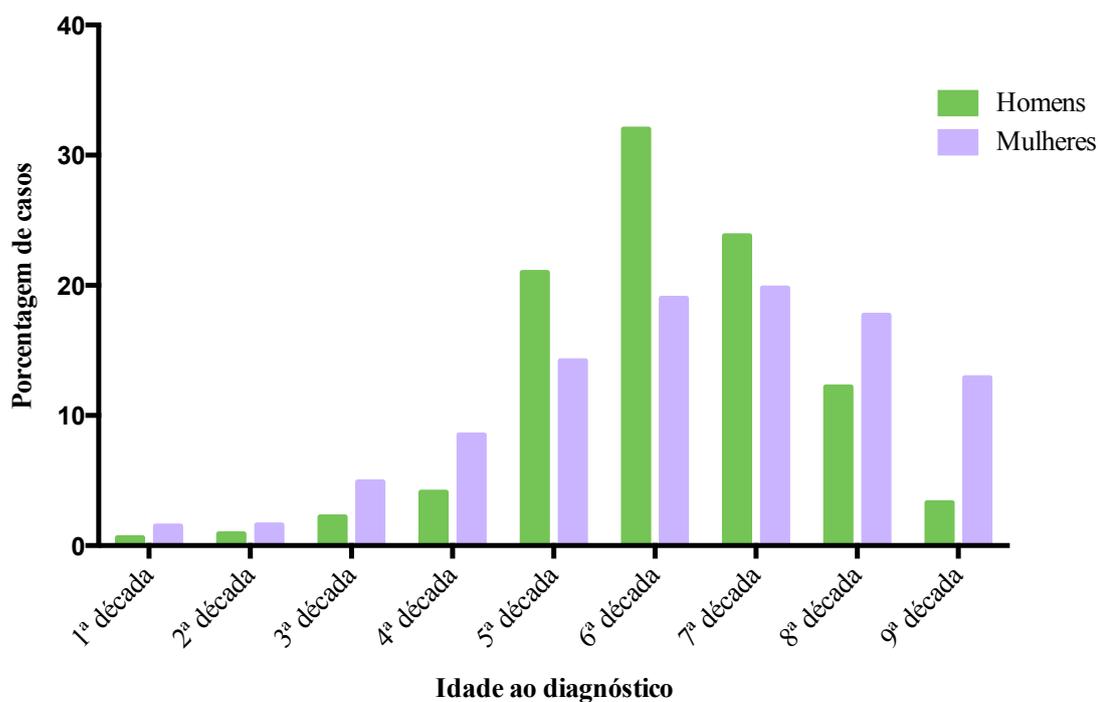


Figura 2 - Incidência de CEB entre os anos de 2000 e 2005 na cidade de São Paulo de acordo com o gênero e faixa etária ao diagnóstico.

1.3 CARACTERÍSTICAS CLÍNICAS

Considerando aspectos epidemiológicos, etiológicos e clinico-patológicos, os carcinomas epidermóides que afetam as estruturas que compõem a cavidade oral são divididos em três grupos, sendo eles: carcinomas do vermelhão do lábio, carcinomas da orofaringe e carcinoma de boca (CEB) (NEVILLE e DAY 2002). A precisa separação desses grupos de tumores é de fundamental importância prática. Isto se deve ao fato de que a etiologia, o comportamento biológico e a resposta à terapia variam significativamente, mesmo entre tumores de mesmo aspecto histopatológico, se compararmos lesões situadas nesses três diferentes sítios (LAMBERT et al. 2011).

Os CEBs são aqueles que envolvem as estruturas anatômicas da língua oral (excluindo a base de língua), assoalho bucal, gengiva, palato duro, mucosa alveolar, mucosas da bochecha e dos lábios (região não exposta ao sol) (NEVILLE e DAY 2002). De acordo com a Classificação Internacional de Doenças - 10 (CID-10 Versão 2015) a codificação dos sítios envolvidos pelo CEB é: C02 - Neoplasia maligna que afeta todas as faces da língua oral; C03 - Neoplasia maligna que afeta todas as faces das gengivas; C04 - Neoplasia maligna que afeta todas as faces do assoalho bucal; C05 - Neoplasia maligna que afeta o palato duro; C06 - Neoplasia maligna que afeta a mucosa jugal, mucosa alveolar, região retromolar e demais áreas da cavidade oral (**Figura 3**).

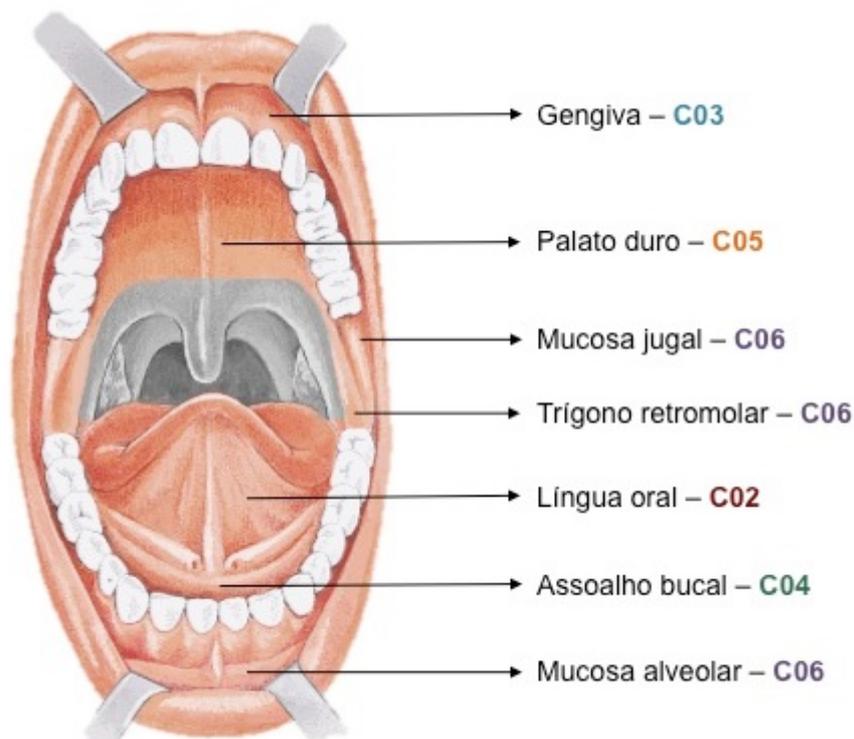


Figura 3 - Estruturas anatômicas da cavidade bucal. Estruturas anatômicas afetadas pelo CEB e suas respectivas codificações de acordo com o CID-10. As áreas em cinza representam as estruturas anatômicas da orofaringe.

Neste trabalho iremos abordar exclusivamente os aspectos relacionados aos tumores que afetam estas regiões, já que os mesmos apresentam características biológicas semelhantes.

O sítio mais acometido pelo CEB é a língua e o assoalho de boca, representando mais de 50% de todos os casos (BAGAN et al. 2010; RIKARDBSEN et al. 2014). Entre pacientes brasileiros, PIRES et al. (2013) observaram que a borda da língua (C02) foi o sítio mais afetado (37%) seguido da mucosa alveolar e gengiva (20%) e assoalho da boca e ventre da língua (19%).

Em estágios iniciais o quadro clínico do CEB é bastante inespecífico, podendo variar de pequenas ulcerações indolores à diminutas áreas leucoplásicas ou eritroplásicas. O diagnóstico neste estágio é dependente da realização de biópsia para avaliação histopatológica. Já em estágios avançados, as características clínicas do tumor tornam-se mais óbvias, sendo comum observar ulcerações de bordas endurecidas com padrão de crescimento endofítico, ou massas tumorais exofíticas firmes à palpação e com limites imprecisos. A presença de sangramento e dor são características comuns a estas lesões. Parestesia, disfagia e dislalia são características frequentemente relatadas, especialmente em tumores que envolvem a língua. Ao exame extra-oral é comum observar tumorações cervicais caracterizadas por linfonodos enfiados que, em muitos casos, apresentam-se fixos à palpação (BAGAN et al. 2010).

1.4 ETIOLOGIA

A etiologia do CEB ainda não é completamente conhecida. Entretanto, há evidências de maior risco de desenvolvimento de CEB relacionados à exposição a diversos fatores. Estes fatores não são necessariamente causais, mas sem dúvida estão associados de alguma forma à doença (DÖBROSSY 2005). Nesta tese, denominarei estes fatores como fatores de risco, e é importante ressaltar que muitos desses fatores de risco apresentam algum grau de associação entre si, podendo representar fatores de confusão.

Considerando apenas os CEBs, que estão restritos às localizações já descritas anteriormente neste texto (CID: C02, C03, C04, C05 e C06), os principais fatores de risco na população brasileira são o tabagismo e o etilismo. Na maioria dos casos esses hábitos estão associados, e têm se mostrado que o efeito sinérgico desses hábitos aumenta significativamente o risco de desenvolvimento do CEB (DÖBROSSY 2005; RADOÏ e LUCE 2013).

O consumo de tabaco é o maior fator de risco independente (JOHNSON et al. 2011). O tabaco possui inúmeras moléculas reconhecidamente carcinogênicas que estão relacionadas ao estresse oxidativo. Ocasionalmente essas moléculas podem causar danos ao DNA do epitélio de revestimento da mucosa bucal e promover mutações. Estas mutações podem se acumular e culminar na transformação maligna deste epitélio. O risco de desenvolvimento de CEB aumenta de modo proporcional

com a frequência, quantidade e o tempo de consumo do tabaco. Entretanto, o abandono do hábito pode reduzir o risco de desenvolvimento do tumor, podendo praticamente desaparecer após 20 anos de abandono do tabagismo. A idade de início de consumo parece ser também importante, já que indivíduos que começam a fumar mais jovens possuem um risco maior de desenvolvimento do tumor que aqueles que iniciaram o hábito em idades avançadas. O uso de cigarros com ou sem filtro não altera o risco de desenvolvimento do tumor. Um estudo brasileiro mostrou que o risco de desenvolvimento de carcinoma epidermóide de língua (sítio mais comum entre fumantes) foi similar entre indivíduos que consumiam cigarro industrializado (OR 16.3 IC95% 2,1-125,0) ou cigarro de palha (OR 16,9 IC95% 2,2-131,0) (FRANCO et al. 1989). Para consumidores de tabaco em outras formulações, como charutos e cachimbo, o risco de carcinoma epidermóide é maior para o palato mole, assoalho de boca, gengiva e o restante da mucosa bucal do que para a língua. A exposição involuntária ao tabaco (em casa e/ou trabalho), denominado de tabagismo passivo, não foi relacionada ao aumento do risco de desenvolvimento para o CEB, embora o seja para outros tumores (RADOÏ e LUCE 2013).

O risco atribuído exclusivamente ao etilismo no desenvolvimento de CEB é baixo quando comparado ao tabagismo, representando em torno de 18% dos casos. Esse risco aumenta de acordo com a quantidade de álcool ingerido por dia, frequência do consumo e o tempo de consumo durante a vida (RADOÏ e LUCE 2013). Uma meta-análise demonstrou que o maior consumo de álcool esteve relacionado a um risco relativo de

desenvolvimento de CEB e de carcinoma epidermóide de faringe de 5.13 (IC 95% 4,31-6,10), enquanto que o consumo baixo ou moderado esteve relacionado a um risco relativo de 1,13 (IC95% 1,00-1,26) e 1,83 (IC95% 1,62-2,07), respectivamente. É importante ressaltar que o risco relativo para desenvolvimento de CEB e CEF entre consumidores de grandes quantidades de álcool foi maior do que para tumores de qualquer outra região do corpo (BAGNARDI et al. 2015).

O efeito carcinogênico do álcool parece estar relacionado ao acetaldeído. Este é um dos produtos da metabolização do álcool no organismo humano e possui ação mutagênica sobre o DNA. O álcool também pode favorecer a atividade carcinogênica de outras moléculas a partir da degradação das membranas celulares fosfolipídicas. Essa degradação pode aumentar a permeabilidade tecidual aos agentes carcinogênicos do tabaco, e, além disso, pode catalisar a ação de tais moléculas. Outro aspecto importante é que o etilismo crônico está relacionado à deficiências nutricionais e danos ao sistema de metabolização hepáticos diminuindo a efetividade deste órgão em eliminar moléculas com potencial ação carcinogênica (JOHNSON et al. 2011), além de influenciar significativamente na composição da microbiota da cavidade oral (THOMAS et al. 2014).

O abandono do consumo do álcool parece diminuir o risco de desenvolvimento do CEB, sendo que o risco diminui progressivamente com o tempo de abandono do hábito. Os sítios que parecem ser mais sensíveis aos efeitos carcinogênicos do etilismo são o assoalho de boca e a região do

trígono retromolar. Embora muitos estudos tenham discutido que tipo de bebidas alcoólicas estariam associadas a um maior risco para o desenvolvimento do CEB, os resultados parecem estar mais relacionados à regionalismos, e a quantidade de álcool consumida parece ter um papel muito mais importante do que o tipo de bebida ingerida (RADOÏ e LUCE 2013).

Entretanto, ainda é motivo de discussão na literatura o papel do etilismo em não fumantes como fator etiológico importante, especialmente para tumores cavidade oral. De qualquer forma, o etilismo em não fumantes parece estar relacionado a um risco de desenvolvimento de CEB significativamente menor do que em tabagistas (RADOÏ e LUCE 2013).

Deficiências alimentares também parecem ter um papel importante na modulação do risco de desenvolvimento de CEB. Uma dieta pobre em frutas e vegetais parece ser um fator de risco para desenvolvimento de tumores da cavidade oral, especialmente quando compara-se ao risco de desenvolvimento de tumores da orofaringe, hipofaringe e laringe em indivíduos mal nutridos (JOHNSON et al. 2011; RADOÏ e LUCE 2013). Um hábito alimentar importante, especialmente na região sul do Brasil, é o consumo de bebida quente preparada a partir da erva-mate (*Ilex paraguariensis*), na forma de “chimarrão”. Foi observado que o consumo desta bebida pode aumentar o risco de desenvolvimento de CEB em 2,11 vezes (IC95% 1,39-3,19) (WARNAKULASURIYA 2009b; RADOÏ e LUCE 2013). Consequência de uma alimentação inadequada, o baixo índice de

massa corporal também parece ser um fator de risco para o desenvolvimento de CEB (RADOÏ e LUCE 2013).

Higiene oral deficiente e saúde oral ruim estão associadas a um maior risco de desenvolvimento de CEB. HOMANN et al. (2001) sugerem que o aumento da placa bacteriana na cavidade oral possui um papel sinérgico com o consumo excessivo de álcool no aumento do risco de desenvolvimento do CEB. Foi observado que a produção de acetaldeído em indivíduos etilistas com pior higiene oral foi significativamente maior do que em indivíduos com melhores hábitos de higiene. Esse risco também parece aumentar em usuários de próteses dentárias mal ajustadas relacionadas à lesões traumáticas orais, especialmente para carcinoma epidermóide da língua (OR 3,2 IC95% 1,4-7,6) (RADOÏ e LUCE 2013). Sugere-se que essa relação se deva à irritação crônica causada pela prótese que pode diminuir a proteção da barreira mucosa contra carcinógenos, agindo como um cofator, especificamente naqueles indivíduos com alta exposição a outros fatores de risco como o tabagismo (WARNAKULASURIYA 2009b). Relata-se também um maior risco de desenvolvimento de CEB entre indivíduos com doença periodontal (RADOÏ e LUCE 2013). Entretanto, o tabagismo é um dos principais agentes etiológicos para as duas patologias. Além disso, o risco de desenvolvimento de CEB em pacientes com doença periodontal é maior entre tabagistas do que entre não tabagistas, indicando um provável fator de confusão nesta associação.

A condição socioeconômica têm sido muito discutida como um importante fator de risco de desenvolvimento de CEB, sendo este maior

entre categorias profissionais desfavorecidas e com menor nível de educação (RADOÏ e LUCE 2013). Entretanto, o desfavorecimento econômico também pode estar diretamente relacionado a outros fatores de risco, como alimentação deficiente, hábitos de higiene precários, tabagismo e etilismo.

Condições de imunossupressão congênitas ou adquiridas como as observadas em pacientes transplantados e pacientes portadores da síndrome da imunodeficiência adquirida (AIDS) também agregam maior risco de desenvolvimento de CEB (RADOÏ e LUCE 2013). Outras condições patológicas como o Líquen plano bucal erosivo e anemia por deficiência de ferro levando ao desenvolvimento da síndrome de Plummer-Vinson também são considerados fatores de risco para o surgimento do CEB (NEVILLE e DAY 2002).

Em culturas orientais, principalmente no sudeste asiático, o consumo de masca de betel é importante agente etiológico para desenvolvimento do CEB (JOHNSON et al. 2011). Entretanto, este não parece ser um fator etiológico importante na nossa comunidade.

O papel da infecção pelo papiloma vírus humano (HPV) no desenvolvimento de carcinomas epidermóides está bem estabelecido para tumores da orofaringe, o que compreende as tonsilas palatinas, seus arcos, palato mole e base da língua (RAMPIAS et al. 2009). Entretanto, para os demais sítios da cavidade oral, ainda não é claro se o HPV possui algum papel na patogênese tumoral (HÜBBERS e AKGÜL 2015).

O HPV é um vírus de DNA que possui tropismo por ceratinócitos. Existem diversos tipos de HPV, sendo que alguns possuem maior potencial de indução de transformação maligna do epitélio que outros. Estes são denominados de HPVs de alto risco, sendo que os principais membros desse grupo são os HPVs 16 e 18. A simples detecção de partículas virais de HPV não indica que esteja ocorrendo um processo de transformação maligna do epitélio. A tradução das proteínas E6 e E7 virais são essenciais para a efetivação deste processo (BOCCARDO 2010; TOMMASINO 2014). Essas proteínas são capazes de inibir importantes genes supressores de tumor no hospedeiro, como a pRb e a p53. A inibição destas proteínas acarreta na alteração do controle do ciclo celular, inibição da apoptose e instabilidade genética, podendo culminar no desenvolvimento de uma neoplasia epitelial maligna. A ligação da proteína viral E7 à proteína pRb do hospedeiro leva à indução da transcrição da proteína p16, que torna-se altamente expressa no epitélio do hospedeiro com infecção ativa por HPV de alto risco (RAMPIAS et al. 2009; LEEMANS et al. 2011). Dessa forma, o aumento da expressão de p16 associado as outras alterações moleculares descritas anteriormente, são considerados importantes marcadores de infecção ativa por HPV (HÜBBERS e AKGÜL 2015). No CEB, a detecção do DNA viral varia bastante, sendo mais comum no assoalho de boca e língua. A prevalência do DNA viral em pacientes com CEB é maior do que na população em geral, mas significativamente menor do que em casos de carcinoma da orofaringe. Entretanto, apesar da detecção do DNA viral, nem sempre é possível estabelecer uma relação entre a infecção e o tumor, já

que a expressão das proteínas E6/E7 em casos CEB é usualmente baixa, e este é o principal marcador de atividade viral no tecido do hospedeiro. Além disso, muitos desses casos não apresentavam aumento da expressão da proteína p16 (LINGEN et al. 2013; REUSCHENBACH et al. 2013). Dessa forma, ainda não é possível estabelecer se a infecção por HPV de alto risco em casos de CEB possui papel importante na etiologia deste grupo de tumores (HÜBBERS e AKGÜL 2015).

1.5 PATOGÊNESE E PROGRESSÃO

Como já mencionado, o surgimento do CEB se dá a partir da transformação maligna dos ceratinócitos de revestimento da mucosa bucal, e evidências indicam que o acúmulo de alterações genéticas seja responsável pela evolução gradual das células epiteliais normais em células neoplásicas. Este modelo é conhecido como carcinogênese em múltiplos passos, sendo o mais aceito para a compreensão da patogênese do CEB **(Figura 4)**.

SLAUGHTER et al. em 1953 observaram, através de análise histopatológica, que casos de carcinoma epidermóide de cabeça e pescoço (CECP) usualmente exibiam, no epitélio adjacente ao tumor, variados graus de atipia celular. Entretanto, clinicamente o epitélio mostrava-se sem alterações que pudessem ser relacionadas ao quadro microscópico. Esses autores relacionaram as frequentes alterações microscópicas no epitélio normal à grande frequência de tumores múltiplos em pacientes com histórico

de tumores primários de cabeça e pescoço. Dessa forma, propuseram uma teoria de que pacientes com CECP apresentavam uma mucosa comprometida e altamente susceptível à formação neoplasias malignas.

A teoria de que a mucosa oral podia esconder alterações que aumentavam o risco de desenvolvimento neoplásico foi confirmada por CALIFANO et al. (1996). Foi verificado que o acúmulo de alterações genéticas no epitélio poderia acarretar em alterações clínicas e histopatológicas na mucosa bucal que, ocasionalmente, culminariam no desenvolvimento de uma neoplasia maligna. De acordo com este trabalho, a quantidade de mutações observadas, muitas vezes, se correlacionava ao grau de alterações morfológicas verificadas microscopicamente, e que para o tumor se desenvolver, um número significativo de alterações no DNA deveria ocorrer. Dessa forma, concluiu-se que quanto maior o número de mutações acumuladas por uma célula, maior a probabilidade desta célula gerar uma neoplasia.

Entretanto, as mutações no epitélio também podem ocorrer na ausência de manifestações clínicas, como demonstrado por TABOR et al. (2001). Esses autores verificaram que alterações genéticas estavam presentes no epitélio das margens cirúrgicas de pacientes com CECP mesmo na ausência de sinais clínicos ou evidências microscópicas de atipia celular. Isso mostrou que, em grande parte dos casos, uma mucosa geneticamente alterada permanece nos pacientes com CECP, o que justificaria a alta taxa de recidivas observadas nesses pacientes. Essa mucosa com maior susceptibilidade à transformação maligna devido ao

acúmulo de mutações no DNA foi denominada de campo de cancerização (BRAAKHUIS et al. 2003).

Acredita-se que o campo de cancerização se forme a partir do seguinte fluxo: inicialmente ocorre uma ou mais mutações em células tronco epiteliais do epitélio normal. Ao se proliferar, esta célula levará a formação de uma população clones mutados. A partir da dinâmica de crescimento fisiológico do epitélio, um campo de células mutadas se disseminaria lateralmente através da superfície mucosa. A ocorrência de novos eventos mutacionais poderia culminar tanto na formação de lesões pré-neoplásicas quanto no próprio desenvolvimento da neoplasia maligna. Estes eventos poderiam ocorrer em células tronco epiteliais de diferentes regiões da mucosa, levando a ocorrência de formação de dois ou mais campos de cancerização em um mesmo indivíduo (BRAAKHUIS et al. 2005a e b; LEEMANS et al. 2011). Assim, podemos refinar o conceito de campo de cancerização considerando como um grupo de células com alterações genéticas, que não apresentam um crescimento invasivo, mas que sob efeito de novos eventos mutacionais podem formar uma neoplasia maligna (BRAAKHUIS et al. 2005b).

A partir desses achados estabeleceu-se as seguintes conclusões: o câncer bucal pode se desenvolver em múltiplas áreas que sofreram alterações pré-neoplásicas (campos de cancerização); tumores são comumente circundados por epitélio atípico; o CEB muitas vezes é constituído por múltiplas lesões independentes que coalescem; e a

persistência de epitélio atípico após a cirurgia de remoção do tumor primário pode explicar o surgimento de tumores múltiplos (BRAAKHUIS et al. 2005b).

Essas descobertas foram essenciais para a compreensão não só do desenvolvimento do CEB, como também no entendimento do surgimento de lesões que acometem a mucosa bucal e estão diretamente relacionadas à carcinogênese bucal. Sugere-se que, em sua maioria, os casos de CEB são precedidos por condições assintomáticas que podem ou não exibir sinais clínicos. Alterações que se apresentam clinicamente visíveis são denominadas de desordens potencialmente malignas (DPM) (previamente chamadas de lesões pré-neoplásicas) e apresentam um maior risco de desenvolvimento do CEB do que a mucosa normal (DIONNE et al. 2015).

São consideradas DPMs as seguintes afecções bucais: a leucoplasia e eritroplasia, a ceratose palatina por fumo invertido, o líquen plano oral, a fibrose submucosa, o lúpus eritematoso discoide, a disceratose congênita e a epidermólise bolhosa (NAPIER e SPEIGHT 2008).

As DPMs de maior significado clínico são a leucoplasia e a eritroplasia, principalmente pela alta prevalência e pela maior taxa de transformação maligna quando comparadas às demais. Embora o conhecimento sobre as DPMs seja limitado, sabe-se que o desenvolvimento da leucoplasia e da eritroplasia pode estar diretamente relacionado aos mesmos fatores de risco relacionados ao desenvolvimento do CEB (NAPIER e SPEIGHT 2008).

A leucoplasia é caracterizada por placas esbranquiçadas que ocorrem na mucosa oral e que histopatologicamente pode revelar hiperqueratose,

hiperplasia e variado grau de displasia arquitetural e celular. Já a eritroplasia, possui diversas semelhanças com a leucoplasia, embora se caracterize por uma lesão atrófica eritematosa que ao exame histopatológico, além de poder exibir variados graus de displasia epitelial, revela também atrofia dos estratos celulares epiteliais. Em alguns casos, são observadas lesões com características leucoplásicas e eritroplásicas, sendo então denominadas de eritroleucoplasias. Apesar dessas características, o diagnóstico dessas condições é de exclusão, ou seja, é realizado após eliminação de qualquer outra possibilidade diagnóstica, já que não há sinais patognomônicos destas entidades (DIONNE et al. 2015).

Embora o maior risco relacionado à leucoplasia e à eritroplasia seja o desenvolvimento do CEB, nem todas essas alterações progridem para transformação maligna. Além da progressão neoplásica, essas alterações podem persistir inalteradas durante toda a vida do paciente, podem progredir em extensão, ou mesmo, podem reduzir de tamanho e até mesmo desaparecer (NAPIER e SPEIGHT 2008).

A descrição da taxa de transformação maligna dessas das DPMs varia significativamente na literatura. Em estudos realizados em países em desenvolvimento, a porcentagem de DPMs que se transformam em tumores malignos varia de 0,13% a 2,2% por ano. Entretanto, ao considerar trabalhos realizados em países desenvolvidos esses números saltam de 1,1% a 17,5% (NAPIER e SPEIGHT 2008). Apesar da baixa porcentagem de transformação maligna, as DPMs representam o principal sinal de risco para o desenvolvimento do câncer de boca.

A determinação de quais lesões irão progredir para o fenótipo maligno ainda não é possível. Entretanto, algumas características são indicativas de maior propensão de malignização. Sabe-se que há maior chance de transformação maligna dentro de 5 anos após o diagnóstico, observando uma tendência de queda no risco após este período. Embora a maioria das leucoplasias afetarem a mucosa jugal, aquelas que surgem no assoalho da boca e na borda da língua são as de maior risco para desenvolvimento neoplásico. Lesões não homogêneas com áreas nodulares ou verrucosas, exibindo uma mistura de padrões leucoplásicos e eritroplásicos (eritroleucoplasia) são de alto risco. Essas características clínicas associadas a um crescimento progressivo caracterizam a leucoplasia verrucosa proliferativa, entidade nosológica que agrega a maior taxa de desenvolvimento do CEB (60% a 100% dos casos). O diagnóstico da leucoplasia em pacientes que não apresentam exposição a fatores de risco, como o tabagismo, também representam um sinal de alerta para progressão neoplásica. Esta condição denominada de leucoplasia idiopática, principalmente no gênero feminino, apresenta maior conversão tumoral do que leucoplasias que ocorrem em indivíduos tabagistas (NAPIER e SPEIGHT 2008). Características histopatológicas, como por exemplo o grau de envolvimento epitelial por displasia celular, têm sido sugeridas como relevantes na predição do surgimento do CEB. Entretanto, essas medidas são altamente subjetivas e com potencial prognóstico ainda incerto (DIONNE et al. 2015).

O surgimento do CEB a partir das modificações genéticas graduais e cumulativas no epitélio, levando ou não à formação de lesões visíveis clinicamente (DPM), é precedido por um estágio denominado de carcinoma *in situ*. Nesta etapa da transformação tecidual, observa-se alterações arquiteturais e pronunciada atipia celular em todos os estratos celulares viáveis do epitélio de revestimento, compatíveis com os aspectos morfológicos de uma neoplasia epitelial maligna. Entretanto, esta lesão não apresenta invasão da lâmina basal, sendo assim restrita ao componente epitelial (WARNAKULASURIYA et al. 2008).

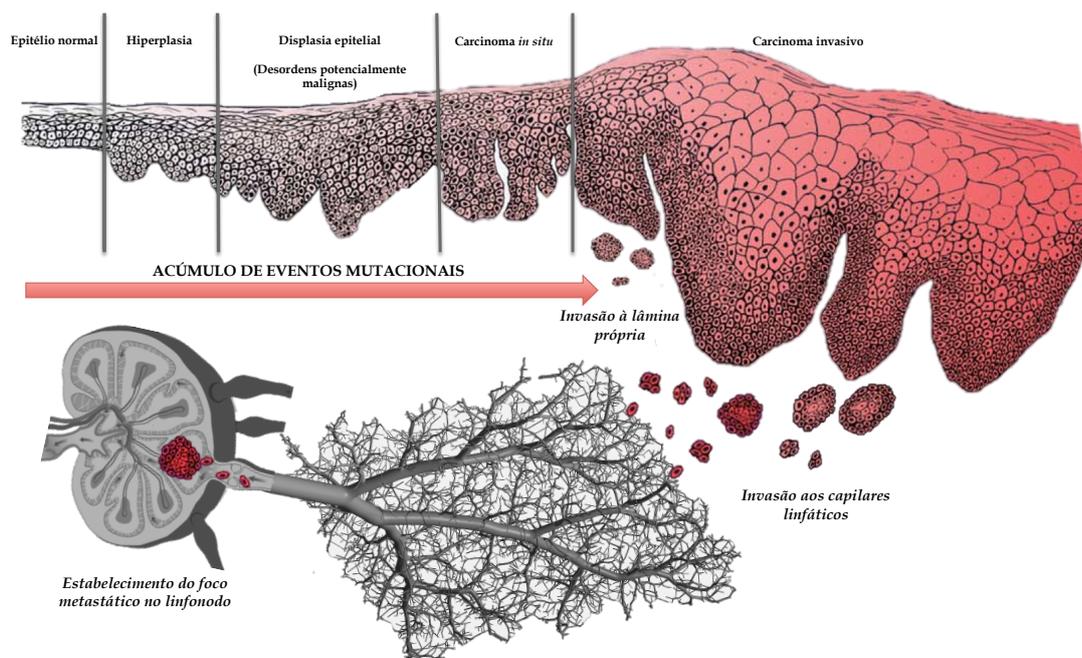
A transformação de uma lesão *in situ* para uma lesão invasiva se dá a partir da aquisição de fenótipos relacionados à degradação da membrana basal, permitindo a invasão epitelial na lâmina própria da mucosa bucal. O processo inicial de invasão se caracteriza por mudanças na interação molecular célula-célula e célula-matriz extracelular, degradação proteolítica do tecido circunjacente e aumento da motilidade permitindo um deslocamento espacial da célula tumoral pelo tecido (STEEG 2006). Uma vez que as células neoplásicas invadam o tecido conjuntivo, diversos padrões de crescimento poderão ser estabelecidos, assim como grande variação no padrão de diferenciação das celular pode ser observado. O CEB pode estabelecer padrões de crescimento em grandes blocos celulares que usualmente conservam características fenotípicas de células epiteliais diferenciadas. Nesses casos observa-se intensa ceratinização no interior da massa tumoral (pérolas de ceratina) associada à manutenção das características morfológicas do epitélio de revestimento. Esses tumores são

classificados como bem diferenciados, e assim considerados de menor agressividade. Entretanto, em muitos casos, observa-se uma significativa perda de diferenciação epitelial, intenso pleomorfismo celular e alto grau de atipia celular e nuclear. Esses tumores pouco diferenciados, podem exibir um padrão de crescimento mais invasivo, caracterizado por pequenos grupos celulares que invadem profundamente o tecido em cordões de células pouco coesas (NEVILLE et al. 2009).

Uma característica dos tumores epiteliais é que o grau de diferenciação celular está comumente relacionada à características biológicas da doença. Tanto é que, para a maioria dos tumores, modificações morfológicas relacionadas à perda de diferenciação epitelial são consideradas na gradação histopatológica, indicando a possibilidade de um comportamento mais agressivo (HANAHAN e WEINBERG 2011). As alterações moleculares relacionadas à perda de diferenciação não só modificam a morfologia celular, como também influenciam nas interações destas células com a matriz extracelular (MEC). Além de um padrão morfológico mais fusiforme, observa-se a perda de junções aderentes, expressão de proteínas que degradam a MEC, aumento de motilidade e resistência à apoptose em tumores agressivos. Essas modificações caracterizam o fenômeno transição epitélio-mesênquima (TEM) e estão comumente associadas ao maior grau de infiltração tecidual e à ocorrência de metástase. Em tumores que passam pela TEM observa-se um aumento da expressão de proteínas envolvidas no processo de migração celular, comumente aumentadas no processo inflamatório ou durante a

embriogênese (HANAHAN e WEINBERG 2011). Essas alterações são o estopim para o desenvolvimento de múltiplas mudanças sequenciais e dramáticas que poderão culminar no último passo de desenvolvimento do CEB que é a geração da metástase tumoral.

Resumidamente, os passos envolvidos no processo de metastatização são constituídos por desenvolvimento da capacidade de invasão local ao tecido adjacente, intravasamento para os capilares sanguíneos ou linfáticos, capacidade de sobreviver ao estresse físico do ambiente circulatório, evadir à ação de células dos sistema imune, extravasamento através da parede endotelial para um novo sítio e indução de formação de um ambiente tecidual adequado para o sucesso do nicho metastático (STEEG 2006).



Fonte: Adaptado de DIONNE et al. (2015)

Figura 4 - Fluxo de surgimento e progressão do CEB. O acúmulo de mutações genéticas forma, gradativamente, um epitélio cada vez mais susceptível ao desenvolvimento de lesões com potencial de transformação neoplásica. Sob efeito de novos eventos mutacionais, esse epitélio pode sofrer transformação maligna, que se caracteriza pela invasão à lâmina própria da mucosa bucal. Novas mutações no genoma de células neoplásicas podem culminar no surgimento de um ou de múltiplos clones com fenótipo metastático, caracterizado pela capacidade de invasão ao leito capilar que sustenta o tumor, sobrevivência no ambiente vascular, capacidade de extravasar e se estabelecer no linfonodo cervical.

1.6 PROGNÓSTICO

O CEB é caracterizado por alta morbidade e mortalidade em decorrência do seu padrão agressivo de crescimento e sua capacidade de gerar metástases. Essas características levam à ocorrência de altos índices de recidivas locais ou loco-regionais, que representam os principais fatores relacionados às baixas taxas de sobrevivência nesses pacientes. Os carcinomas

epidermóides que afetam a cavidade oral são especialmente agressivos e com pior prognóstico quando comparados aos tumores dos demais sítios da cabeça e pescoço (LEEMANS et al. 1994; WÜNSCH-FILHO 2002; PATEROMICHELAKIS et al. 2005; KADEMANI 2007; OW e MYERS 2011).

Apesar da sua agressividade, dados do *Surveillance, Epidemiology and End Results Program* (SEER) do *National Institute of Health* (NIH) dos Estados Unidos, mostram que a sobrevida em pacientes com CEB aumentou nos últimos anos, principalmente em pacientes com tumores da língua oral (CID-C02) (39,7% em 1975 e 64,8% em 2004) e demais sítios da cavidade oral (CID-C06) (51,8% em 1975 e 61,5% em 2004) (**Figura 5**). Essa tendência também foi observada em um estudo holandês (BRAAKHUIS et al. 2014), onde a sobrevida em 5 anos de pacientes com CEB entre 1989 - 1994 subiu de 56% para 62% em 2007, assim como em um outro estudo europeu (KARIM-KOS et al. 2008). De acordo com SHAH e GIL (2009), fatores como o diagnóstico precoce, melhor conhecimento das características do tumor e do paciente que podem influenciar na decisão terapêutica, tratamento adequado das disseminações tumorais para os linfonodos cervicais e implementação da radioterapia adjuvante e/ou quimioterapia associada à radioterapia, são os principais responsáveis por essa melhora. No entanto, a sobrevida no Brasil ainda é pobre, decorrente principalmente do atraso no diagnóstico já que os pacientes são usualmente diagnosticados com tumores avançados (WÜNSCH-FILHO 2002; OLIVEIRA et al. 2006).

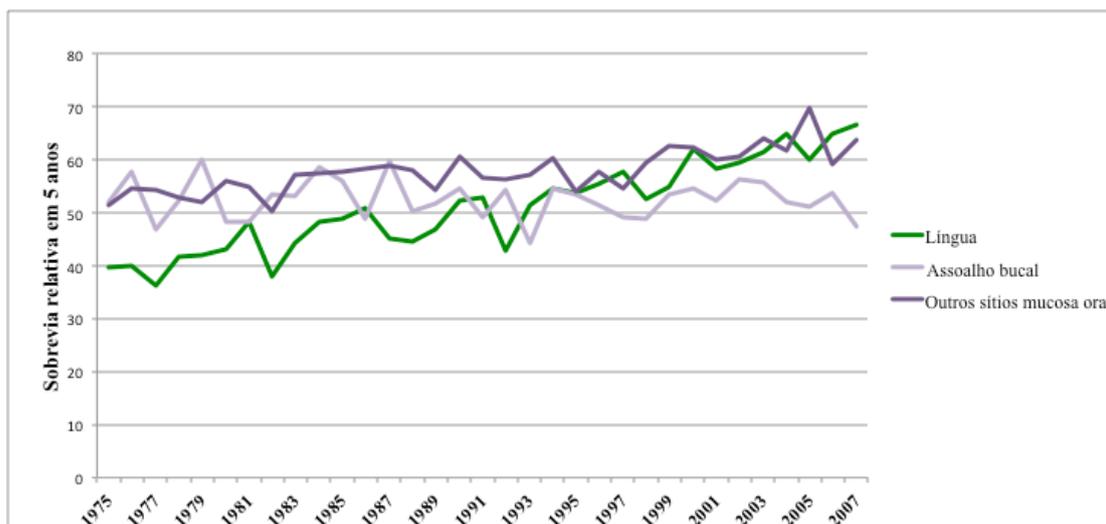


Figura 5 - Dados de sobrevivência relativa em 5 anos obtidos do SEER-NIH no período entre 1975 e 2007 para CEB de língua, assoalho bucal e outros sítios da mucosa oral. Observa-se um pequeno incremento na sobrevivência de pacientes com tumores de língua e outros sítios da mucosa oral. Já tumores do assoalho bucal não exibem a mesma tendência.

A determinação do prognóstico de pacientes com CEB é feita a partir de observações relacionadas ao tumor e ao próprio paciente. Os principais fatores de risco relacionados à mortalidade são associados ao tumor, como o seu tamanho, sua localização e seu padrão de disseminação local e à distância (SHAH e GIL 2009; BELCHER et al. 2014; MONTERO et al. 2014). Características do paciente, como presença de comorbidades, aceitação e cooperação ao tratamento, abandono de fatores de risco, como o tabagismo e o etilismo, e aspectos socioeconômicos são importantes na determinação terapêutica e, dessa forma, também podem influenciar no prognóstico (SHAH e GIL 2009).

O tamanho tumoral é um importante fator na determinação prognóstica de pacientes com CEB. Tumores grandes são tratados de forma

muito mais invasiva que tumores menores, o que aumenta significativamente a morbidade e, conseqüentemente, influencia na mortalidade pós-cirurgia. Além disso, o tamanho tumoral possui relação direta com a ocorrência de metástases nodais, que representam o principal fator prognóstico nesses pacientes.

A localização tumoral também é importante, influenciando na morbidade decorrente da cirurgia e acessibilidade à radioterapia. Além disso, enquanto tumores do palato duro e das gengivas superiores apresentam um comportamento relativamente melhor, tumores que afetam a língua oral, o assoalho de boca e gengivas inferiores apresentam comportamento mais agressivo (SHAH e GIL 2009). Essas características evidenciam que múltiplos aspectos devem ser considerados para a compreensão do curso da doença e determinação das opções de tratamento. Assim, para a categorização de pacientes com CEB para fins de determinação terapêutica e prognóstica, utiliza-se critérios de mensuração da extensão da doença baseado em características anatômicas do tumor primário e do status de disseminação para linfonodos regionais e sítios à distância. Essa classificação é denominada de TNM (T - descreve o tamanho do tumor primário; N - descreve o envolvimento linfonodal por doença metastática; M - descreve a presença de doença metastática à distância) e é regulada pelo *Union for International Cancer Control (UICC)* e pelo *American Joint Committee on Cancer (AJCC)*. Esta classificação quando realizada a partir de critérios obtidos durante o exame clínico recebe o prefixo "c", sendo denominado cTNM. Entretanto, a mesma avaliação da doença pode ser feita

durante o exame anatomopatológico e, dessa forma, recebe o prefixo “p” (pTNM). Outras variações da classificação TNM existem para indicar outras particularidades de importância clínica. O sufixo “(m)” é usado após a designação do tamanho do tumor indicando a presença de múltiplos tumores primários em um único sítio (cT(m)NM). O prefixo “y” é empregado quando se mensura a doença durante ou após uma terapia multimodal inicial (ycTNM ou ypTNM). O prefixo “r” indica um tumor recorrente diagnosticado após um intervalo livre de doença (rTNM) e o prefixo “a” é utilizado após determinação da extensão do tumor durante a autópsia (PATEL e SHAH 2005; BELCHER et al. 2014). Os critérios utilizados para qualquer uma das variações dessa classificação estão descritos no Quadro 1.

A partir da classificação TNM determina-se o estadiamento da doença, combinando cada um dos critérios de extensão tumoral, como descrito na Quadro 2. O estadiamento é um dos principais fatores prognósticos para pacientes com CEB e é uma importante ferramenta para determinação das opções terapêuticas (SHAH e GIL 2009). O estágio clínico ao diagnóstico é o fator mais importante para a determinação do prognóstico, já que tumores de estágio iniciais apresentam altas taxas de cura. Entretanto, uma vez estabelecidas metástases nos linfonodos cervicais um significativo declínio na sobrevida é observado (SHAH e GIL 2009; BELCHER et al. 2014). A casuística de pacientes tratados no *Memorial Sloan-Kettering Cancer Center*, mostra a relação entre o aumento do estadiamento clínico e a diminuição da sobrevida em 5 anos para pacientes com CEB. Dentre aqueles com estágio clínico I, a sobrevida foi de 94%, para

o estágio clínico II 80%, estágio clínico III 66% e estágio clínico IV 58% (SHAH e GIL 2009).

Outras características do tumor que não estão incluídas no sistema TNM também são sugeridas como importantes na determinação prognóstica, como a avaliação histopatológica. Embora a gradação histopatológica do CEB reflita a agressividade tumoral, essa classificação não se mostra um parâmetro independente na determinação prognóstica. Entretanto, a medida histopatológica da profundidade de invasão tumoral parece ser importante. Enquanto lesões *in situ* e superficiais apresentam uma alta taxa de cura, lesões infiltrativas na profundidade possuem um pior prognóstico. Entretanto, como não é possível determinar essa informação no tempo pré-cirúrgico, o uso dessa informação para a decisão terapêutica fica inviabilizada (SHAH e GIL 2009).

Quadro 1 - Classificação clínica do CEB

Sítio	Classificação	Critério
Tumor Primário	Tx	Tumor primário não pode ser avaliado
	T0	Sem evidência de tumor primário
	Tis	Carcinoma <i>in situ</i>
	T1	Tumor com 2 centímetros ou menos em seu maior diâmetro
	T2	Tumor com mais de 2 centímetros, mas não mais que 4 centímetros no seu maior diâmetro
	T3	Tumor com mais de 4 centímetros em seu maior diâmetro
	T4a	Doença moderadamente avançada - Tumor invade apenas estruturas adjacentes
	T4b	Doença muito avançada - Tumor invade espaço mastigatório, lâmina pterigoide, base de crânio e/ou recobre a artéria carótida
Linfonodos regionais	Nx	Linfonodos regionais não podem ser avaliados
	N0	Ausência de metástase linfonodal
	N1	Metástase em um único linfonodo ipsilateral, com 3 centímetros ou menos no maior diâmetro
	N2	Metástase em um único linfonodo ipsilateral, com mais de 3 centímetros mas não mais que 6 centímetros no maior diâmetro; ou metástase bilateral ou no linfonodo contralateral, sempre menor que 6 centímetros no maior diâmetro
	N2a	Metástase em um único linfonodo ipsilateral, com mais de 3 centímetros mas não mais que 6 centímetros no maior diâmetro
	N2b	Metástase em múltiplos linfonodos ipsilaterais, nenhum maior que 6 centímetros no maior diâmetro
	N2c	Metástase em linfonodo bilateral ou contralateral, nenhum maior que 6 centímetros de diâmetro
N3	Metástase em um linfonodo com mais de 6 centímetros de diâmetro	
Metástase à distância	M0	Ausência de metástase à distância
	M1	Presença de metástase à distância

Quadro 2 - Estadiamento clínico do CEB baseado no TNM

Estádio	T	N	M
0	Tis	N0	M0
I	T1	N0	M0
II	T2	N0	M0
III	T3	N0	M0
	T1	N1	M0
	T2	N1	M0
IVA	T3	N1	M0
	T4a	N0	M0
	T4a	N1	M0
	T1	N2	M0
	T2	N2	M0
IVB	T3	N2	M0
	T4a	N2	M0
IVC	Qualquer T	N3	M0
	T4b	Qualquer N	M0
	Qualquer T	Qualquer N	M1

Temos então que tumores avançados, caracterizados pelo grande tamanho e pela presença de metástases nos linfonodos cervicais, são relacionados a um curso pior da doença. Assim, podemos sugerir que o diagnóstico precoce, levando ao tratamento de tumores menores facilmente tratáveis, e a detecção precisa da metástase nodal, favorecendo a adequada indicação do tratamento do tumor disseminado, podem favorecer sobremaneira a melhoria nos índices de cura desta doença.

1.7 METÁSTASE LINFONODAL

A principal via de disseminação metastática do CEB são as vias linfáticas que desembocam nos linfonodos cervicais (SANO e MYERS 2007). Como já discutido, a presença da metástase cervical representa um dos

principais fatores relacionados à diminuição da sobrevida, mas outros fatores, como o número de linfonodos envolvidos e a presença de extravasamento de células tumorais para a cápsula linfonodal também são importantes na determinação de um pior prognóstico. A identificação e remoção dos focos metastáticos é de fundamental importância para se alcançar o bloqueio da progressão da doença (SHRIME et al. 2009; LEEMANS et al. 2011; BELCHER et al. 2014; MONTERO et al. 2014).

Determinar quais tumores irão gerar disseminações cervicais ainda é difícil, mesmo que algumas características clinicopatológicas possam ser indicativas de maior risco. Sítios como a língua oral, assoalho de boca e gengiva inferior e a profundidade de invasão tumoral são fatores relacionados a um risco maior de desenvolvimento metastático para os linfonodos cervicais (SPIRO et al. 1986; PO WING YUEN et al. 2002; SHAH e GIL 2009). Considerando especificamente tumores da língua e assoalho bucal, PIMENTA AMARAL et al. (2004) observaram que embolização tumoral vascular, reação desmoplásica e perda de diferenciação celular também indicam maior propensão para disseminação metastática. Apesar desses achados, muitas dessas características não mostram acurácia e aplicabilidade necessárias para determinação prognóstica ou influenciar na escolha terapêutica (SPARANO et al. 2004). Dessa forma, apenas o tamanho e profundidade de invasão do tumor primário e a detecção de sinais de metástase cervical ao exame clínico do pescoço serão levados em conta na mensuração do risco de desenvolvimento metastático e conseqüentemente na decisão terapêutica (SHAH e GIL 2009).

O tratamento dos pacientes com doença metastática linfonodal inclui a remoção cirúrgica das estruturas linfáticas cervicais, denominado de esvaziamento cervical. Este pode ser classificado como terapêutico quando é aplicado em indivíduos com evidência clínica de metástase (cN+), ou eletivo, quando é realizado naqueles indivíduos sem evidências clínicas de metástase (cN0), mas que apresentam características relacionadas a um maior risco de desenvolvimento metastático, como profundidade de invasão tumoral avançada. Outra alternativa para pacientes cN0 é a política de seguimento rigoroso, onde o intervalo de acompanhamento é reduzido para que ao primeiro sinal de doença cervical, o tratamento seja implementado (SANO e MYERS 2007).

Em pacientes com CEB, a principal característica clínica que indica o esvaziamento cervical eletivo é a presença de tumores grandes (T3 e T4), mesmo sendo classificados como cN0. Dessa forma, ao se realizar o tratamento eletivo do pescoço diminui-se o risco de “perder” pacientes com micrometástases (focos metastáticos invisíveis clinicamente evidenciados apenas ao exame histopatológico). Já a presença de linfonodos palpáveis e fixos ao toque representa um forte indício de doença metastática, sendo o principal sinal que determina a abordagem cirúrgica terapêutica (SANO e MYERS 2007; SHAH e GIL 2009).

Entretanto, a decisão terapêutica em pacientes com tumores pequenos (T1 e T2) cN0 ainda é alvo de discussão na literatura. Isso se dá pela ausência de recursos para mensuração precisa do risco de desenvolvimento metastático nestes casos. Os métodos diagnósticos

disponíveis atualmente possuem baixa sensibilidade, sendo relatadas taxas de 30% a 40% de metástases linfonodais não diagnosticadas (LEUSINK et al. 2012).

A mensuração da profundidade de invasão tem sido usada como uma possível solução para determinação de risco para ocorrência de metástases nodais em CEBs iniciais (SPIRO et al. 1986; PO WING YUEN et al. 2002). PO WING YUEN et al. (2002) observaram que o aumento da profundidade de invasão tumoral se correlacionava diretamente ao aumento da prevalência de metástases nodais subclínicas. Entretanto, a determinação da profundidade de invasão é realizada a partir da avaliação microscópica da peça cirúrgica, inviabilizando a mensuração deste parâmetro previamente ao tratamento do tumor primário. Assim, na prática clínica, essa medida é realizada arbitrariamente, a partir da palpação do tumor primário, separando tumores invasivos, candidatos ao esvaziamento cervical eletivo, de tumores superficiais que irão se beneficiar do tratamento conservador (SHAH e GIL 2009).

Mesmo assim, muitos autores defendem o emprego abrangente da abordagem eletiva, já que ela provê tratamento caso o paciente apresente doença não detectável clinicamente, e também permite o diagnóstico acurado e estadiamento preciso, melhorando a determinação prognóstica (SANO e MYERS 2007). D'CRUZ et al. (2015) compararam a sobrevida de pacientes com tumores iniciais (T1, T2) cN0, submetidos ao esvaziamento cervical eletivo (n=245), com pacientes submetidos ao esvaziamento cervical terapêutico (n=255), ou seja, foram submetidos ao tratamento do pescoço

apenas após evidências clínicas de metástase. Foi observado que os pacientes submetidos ao esvaziamento terapêutico sofreram mais recorrências (69,5% contra 45,9%) e a sobrevida em 3 anos foi significativamente menor (67.5% contra 80%), sendo observado um risco relativo de morte para pacientes do grupo de tratamento eletivo significativamente menor. Embora o tratamento eletivo tenha levado a um maior número de eventos adversos, os autores sugerem que, mesmo para tumores pequenos, essa abordagem represente a melhor opção de tratamento.

Outro fator importante a favor da preconização da terapia eletiva é o fato de que os linfonodos cervicais representam o principal sítio de recidiva em pacientes não submetidos ao esvaziamento do pescoço durante o tratamento do tumor primário. Além disso, quando as recidivas cervicais ocorrem, estas irão apresentar um estágio de envolvimento linfonodal mais avançado e maior incidência de extravasamento extra-capsular, refletindo negativamente na sobrevida. Assim, sugere-se que apenas metodologias com valor preditivo positivo e negativo excepcional para diagnóstico das metástases nodais poderiam substituir a implementação do esvaziamento cervical eletivo (D'CRUZ et al. 2015).

O grande problema relacionado à intervenção cervical eletiva está no fato de que parte dos pacientes cN0 nunca irá desenvolver metástases nodais. Acredita-se que pacientes cN0 que irão revelar metástases ao exame histopatológico representem 40% dos casos. Isso indica que os demais 60% que foram submetidos ao tratamento eletivo receberam

tratamento agressivo, sendo expostos à maior morbidade desnecessariamente (LEUSINK et al. 2012). WOOLGAR (1999) comparou a prevalência de metástase linfonodal detectada histopatologicamente com a prevalência dessa alteração determinada a partir do exame clínico baseado em palpação e ressonância magnética. Este autor observou que 21% dos linfonodos considerados negativos ao exame clínico revelaram metástase histopatológica. Por outro lado, apenas 77% dos linfonodos considerados positivos ao exame clínico se revelaram metastáticos ao exame histopatológico. A ocorrência de exames falso positivos foi significativamente maior em tumores pequenos (T1-T2) do que em tumores grandes (T3-T4). Embora necrose e áreas císticas sejam sinais radiológicos importantes para o diagnóstico da metástase nodal através de exames de imagem, em metástases de tumores pequenos esses eventos são raramente observados. Isso se deve provavelmente ao padrão de crescimento metastático nestes casos, que se caracteriza pela formação de micrometástases na periferia do parênquima linfonodal. Essa característica também justifica a ausência do sinal clínico mais comum à metástases nodais mais desenvolvidas, que é a presença de linfonodos fixos, enfiados e clinicamente palpáveis.

A dissecação do pescoço, tanto no manejo eletivo como no terapêutico, pode comprometer a qualidade de vida dos pacientes. O quadro clínico observado em 60% a 80% dos pacientes submetidos ao esvaziamento cervical radical é caracterizado por sintomas álgicos, limitações funcionais dos ombros e comprometimento postural. Inicialmente, considerava-se que a exérese do nervo acessório com consequente

desnervação do terço apical do trapézio fosse a causa desse quadro clínico. Entretanto, o advento da modalidade modificada de esvaziamento cervical e o conceito de esvaziamento cervical seletivo permitiram a conservação do nervo acessório. Embora um menor número de pacientes apresentassem comprometimento funcional do ombro (18% a 77% para pacientes submetidos ao tratamento radical modificado e 29% a 39% para pacientes submetidos ao esvaziamento seletivo), esse quadro clínico ainda é observado entre aqueles submetidos a intervenções menos invasivas. Outros fatores que corroboram com um pior quadro clínico são o emprego de radioterapia, manipulação direta do nervo ou outros fatores não relacionadas ao comprometimento do nervo acessório como a capsulite adesiva. Esses dados mostram que tanto abordagens radicais como abordagens conservadoras podem estar relacionadas ao desenvolvimento de desordens funcionais. Assim, faz-se necessário que o emprego do tratamento cervical seja balizado nas necessidades reais de cada paciente, evitando um aumento da morbidade naqueles indivíduos sem risco de desenvolvimento metastático (BRADLEY et al. 2011).

O maior impedimento para a implementação da indicação personalizada do esvaziamento cervical em pacientes com CEB é a falta de métodos diagnósticos acurados para detecção de focos metastáticos. Associado ao exame clínico de palpação cervical, exames de imagem como o ultrassom, tomografia computadorizada, ressonância magnética e PET-CT também são empregados para a detecção das metástases. BREE et al. (2014) revisaram a literatura em busca avanços técnicos no diagnóstico de

metástases nodais ocultas em CECP. De acordo com os autores, nenhum dos exames de imagem atualmente disponíveis possuem acurácia necessária para substituir o esvaziamento cervical eletivo.

Uma alternativa a esses exames, é a realização da biópsia do linfonodo sentinela (ALKUREISHI et al. 2009). Os linfonodos sentinela são os primeiros linfonodos da cadeia linfática a receberem drenagem do sítio anatômico referente a localização do tumor primário. Este linfonodo apresenta maior probabilidade de exibir metástase caso o tumor primário possua este fenótipo, podendo indicar o status dos demais linfonodos desta cadeia. Dessa forma, a detecção da metástase neste linfonodo é indicativa da necessidade de esvaziamento cervical, enquanto que a ausência de foco metastático indica que o paciente não necessita de abordagem cirúrgica cervical (ALKUREISHI et al. 2009; LEUSINK et al. 2012; DE BREE et al. 2014). Em geral, a sensibilidade e especificidade da detecção de metástases cervicais em CEB, utilizando a biópsia do linfonodo sentinela, encontram-se acima dos dos 90%, além disso, esta abordagem causa menor morbidade quando comparada ao esvaziamento cervical eletivo e apresenta acurácia muito maior que os exames de imagem (LEUSINK et al. 2012; DE BREE et al. 2014; PEDERSEN et al. 2015).

Entretanto, esta técnica depende da administração de contrastes radioativos, realização de linfocintilografia, exame histopatológico minucioso do espécime, qualificação técnica do cirurgião e adição de mais uma etapa cirúrgica em casos em que se confirme a presença de metástases (LEUSINK et al. 2012). Além disso, para localizações como o assoalho de

boca, a detecção do linfonodo sentinela, assim como a sensibilidade da detecção metastática são menores do que para outros sítios da cavidade oral. Dessa forma, evoluções técnicas ainda são necessárias para uma ampla implementação desta abordagem. A alta qualificação técnica, custos e aumento do tempo de procedimento cirúrgico também inviabilizam a ampla implementação desta abordagem (DE BREE et al. 2014; PEDERSEN et al. 2015).

Os trabalhos discutidos acima evidenciam a precariedade de alguns métodos de diagnóstico de metástase nodal em pacientes com CEB, assim como as limitações existentes relacionadas à metodologias mais acuradas. Estes obstáculos corroboram com a preconização da implementação do esvaziamento cervical eletivo na maioria dos casos de CEB. Embora sejam claras as vantagens da implementação dessa terapia na sobrevida desses pacientes, deve-se ter em mente que para parte dos pacientes tem se promovido morbidade desnecessária, diminuição da qualidade de vida a longo prazo e aumento dos custos médicos. Dessa forma, desenvolver metodologias mais precisas e de fácil implementação para a detecção de linfonodos metastáticos, ou mesmo predizer tumores com potencial metastático faz-se mais que necessário na melhoria do tratamento de pacientes com CEB.

Considerando que as características fenotípicas tumorais são determinadas pelo seu padrão mutacional e de expressão gênica, é possível que a investigação das diferenças transcricionais entre tumores metastáticos e não metastáticos possa indicar marcadores de metástase nodal e

determinar vias biológicas responsáveis por este processo. Conseqüentemente, a descoberta de tais marcadores poderá permitir uma melhor indicação da intervenção cervical em pacientes com CEB.

1.8 ANÁLISE TRANSCRICIONAL DE CEB METASTÁTICOS

O conjunto de todos os transcritos produzidos por uma unidade biológica em um determinado período de tempo é chamado de transcriptoma. Dessa forma, o transcriptoma representa o início da determinação fenotípica dessa unidade biológica. Assim, ao se comparar duas entidades biológicas fenotipicamente diferentes, espera-se observar variações no padrão de transcrição gênica que expliquem suas diferenças (WANG et al. 2009).

Como a separação entre CEB metastáticos e não-metastáticos ainda é limitada a partir de tecnologias clínicas, acredita-se que a comparação do padrão transcricional desses grupos de tumores poderá identificar transcritos diferencialmente expressos e, conseqüentemente, envolvidos no processo metastático.

A ideia de que padrões de expressão gênica poderiam ser utilizados para auxiliar o diagnóstico de metástases nodais já foi abordada em alguns estudos utilizando amostras de CEB (NAGATA et al. 2003; IRIÉ et al. 2004; SCHMALBACH et al. 2004; O'DONNELL et al. 2005; ROEPMAN et al. 2005, 2006; KATO et al. 2006; ZHOU et al. 2006; MÉNDEZ et al. 2007; NGUYEN

et al. 2007; KASHIWAZAKI et al. 2008; HENSEN et al. 2008; KANG et al. 2009; VAN HOOFF et al. 2012) (**Quadro 3**).

MÉNDEZ et al. (2007) investigaram diferenças transcricionais entre tumores primários não-metastáticos com células neoplásicas obtidas de linfonodos de pacientes com tumores metastáticos. Eles encontraram um grupo de 160 genes diferencialmente expressos entre esses grupos. Ao compararem o padrão de expressão desses 160 genes entre os tumores primários metastáticos e não metastáticos, encontraram 63 genes que mantinham diferença de expressão entre os grupos (24 hiperexpressos e 39 hipoexpressos no grupo metastático). A classificação destes tumores de acordo com o padrão de expressão desses genes mostrou-se eficiente para separação entre tumores metastáticos e não metastáticos. Ressalta-se que os tumores primários metastáticos apresentavam padrão de expressão gênica muito semelhante às células metastáticas dos linfonodos, sugerindo a manutenção do padrão de expressão gênica entre as células metastáticas dos linfonodos e os tumores primários correspondentes.

Quadro 3 - Estudos que buscaram a determinação de transcritos relacionados à metástase nodal em CEB

Trabalhos	Sítio	Amostras	N+	N0	Abordagem	DEG
<i>Irié et al. (2009)</i>	Cavidade oral	11	4	7	Microdissecção a laser + Microarray (Human Cancer 1.2 Atlas cDNA array - Clontech)	*
<i>Schmalbach et al. (2004)</i>	Cavidade oral e orofaringe	20	13	7	Microarray (HG_U95Av2 - Affymetrix)	57
<i>Kato et al. (2006)</i>	Cavidade oral e orofaringe	15	5	10	Microarray (AceGene Human oligo chip 30K - Hitachi)	20
<i>Méndez et al. (2007)</i>	Cavidade oral e orofaringe	11	6	5	Microdissecção a laser + Microarray (Human Genome Focus Array - Affymetrix)	62
<i>Zhou et al. (2006)</i>	Língua oral	25	11	14	Microarray (Human Genome U133 Plus 2.0 GeneChip - Affymetrix)	33
<i>O'Donnell et al. (2005)</i>	Cavidade oral	18	11	7	Microarray (Human Genome U133 Plus 2.0 GeneChip - Affymetrix)	116
<i>Nagata et al. (2003)</i>	Cavidade oral	15	8	7	Microarray (IntelliGene Human Cancer CHIP v2.1- Takara)	19
<i>Roepman et al. (2005)</i>	Cavidade oral e orofaringe	82	37	45	Microarray (Human Array-Ready Oligo v2.0 - Qiagen)	102
<i>Kashiwazaki et al. (2008)</i>	Cavidade oral	41	9	32	Microrray (in-house cDNA array of 1289 genes - GeneticLab)	39
<i>Nguyen et al (2007)</i>	Cavidade oral	30	17	13	Microdissecção a laser + Microarray (Human Genome U133 Plus 2.0 - Affymetrix)	85
<i>Roepman et al. (2006)</i>	Cavidade oral e orofaringe	66	28	38	Microarray (Human Array-Ready Oligo v2.0 - Qiagen)	825

N+ - Tumores com metástase nodal; N0 - tumores sem metástase nodal; DEG - Genes diferencialmente expressos;

* *Irié et al. (2009)* não indica o número total de genes diferencialmente expressos identificados.

Dados semelhantes foram obtidos por O'DONNELL et al. (2005) que também observaram padrão de expressão gênica semelhantes entre tumores primários e as respectivas metástases. Esses autores destacam que apesar de terem encontrado um grupo de 116 genes diferencialmente expressos entre tumores metastáticos e não metastáticos, nenhum desses marcadores havia sido relacionado à disseminação metastática em CEB em outros trabalhos.

WARNER et al. (2004) estudaram especificamente tumores de língua e assoalho bucal e, com base no padrão transcricional, conseguiram separar tumores em grupos clínicos distintos. O primeiro grupo foi formado principalmente por tumores de indivíduos do gênero masculino, grandes (T3 e T4) e com metástases nodais. O segundo grupo foi formado principalmente por mulheres e com menor taxa metástase linfonodal. Esses resultados demonstram que os padrões de expressão gênica podem variar significativamente entre grupos aparentemente homogêneos. Isso pode ser um indicador de que abordagens de análise transcricional de baixa cobertura podem não ser suficientes para estabelecimento de padrões transcricionais relativos a uma determinada característica clínica.

O maior estudo, em número de pacientes, relacionando o padrão de expressão gênica e o status dos linfonodos cervicais foi realizado por ROEPMAN et al. (2005). Esses autores analisaram 82 tumores primários da cavidade oral e orofaringe, sendo que 45 destes revelaram metástase nodal ao exame anatomopatológico. Em uma validação com amostras independentes (utilizando apenas 22 amostras) os autores atingiram 100% de acurácia para diagnóstico de pacientes livre de metástase nodal e de 77% para diagnóstico de indivíduos com acometimento linfonodal.

A assinatura gênica da metástase nodal proposta por ROEPMAN et al. (2005) e seu aprimoramento (ROEPMAN et al. 2006) foram desafiados em um estudo multicêntrico holandês envolvendo 316 pacientes com carcinoma epidermóide de boca e de orofaringe (VAN HOOFF et al. 2012). Além da confirmação da eficiência diagnóstica da assinatura gênica, os

autores propuseram que uma combinação de critérios clínicos e de expressão gênica fossem utilizados na determinação terapêutica do pescoço de indivíduos com câncer de boca. Os autores acreditam que a combinação destes critérios iria evitar que cerca de 50% de cirurgias desnecessárias de esvaziamento cervical fossem realizadas.

Segundo van VAN HOOFF et al. (2012) o uso de um grande número de genes para determinação da assinatura gênica da metástase é sempre preferível do que a determinação de um grupo restrito de genes. Isso evitaria que variações entre amostras ou variações técnicas invalidassem tal assinatura.

Apesar do sucesso obtido por alguns trabalhos, SCHMALBACH et al. (2004) detectou 101 genes diferencialmente expressos entre tumores metastáticos e não metastáticos, mas foi o único incapaz de separar estes grupos a partir do padrão de expressão gênica global. Entretanto, tumores não metastáticos pareciam ser mais semelhantes entre si do que os metastáticos. Esse resultado talvez indique que o processo metastático possa decorrer de alterações desordenadas no padrão de expressão gênica e não a partir da ocorrência de processos específicos, ressaltando a importância da investigação do padrão transcricional a partir de metodologias de ampla detecção de transcritos.

Apesar do sucesso obtido na validação dos resultados gerados por VAN HOOFF et al. (2012), esses resultados refletem o padrão observado a partir de um pequeno grupo de pacientes (ainda que seja o maior). Embora diferentes sub-sítios de CECP exibirem variações no comportamento clínico

e grande variação genética, a maioria desses estudos utiliza tumores de mais de um sub-sítio, principalmente tumores de cavidade oral e orofaringe, que podem apresentar etiologia e patogênese distintas (HUANG et al. 2002; FREIER et al. 2003). CHUNG et al. (2004) estabeleceram um grupo de genes relacionados à metástase nodal em CECP e verificaram que aqueles localizados na cavidade oral foram os únicos a não se enquadrar no padrão de expressão gênica esperado para tumores metastáticos. Isso indica que o padrão transcricional da metástase nodal em CEB se diferencia significativamente do padrão de outros CECP.

Adicionalmente, ao comparar a lista de genes diferencialmente expressos relacionados à metástase em diferentes trabalhos, verificamos que há uma baixa complementariedade entre as listas de diferentes estudos. Considerando os transcritos de acordo com o sentido da alteração de expressão para o grupo metastático (aumento ou diminuição da expressão), verificamos que nenhum transcrito foi encontrado alterado em todos os estudos (**Quadros 4 e 5**).

Quadro 4 - Transcritos diferencialmente hiperexpressos em tumores metastáticos encontrados em diferentes estudos

Trabalhos	Número de transcritos diferencialmente expressos	Número de transcritos em comum	Transcritos
<i>Roepman et al. (2005)</i>	102		
<i>Roepman et al. (2006)</i>	825	1	<i>POSTN</i>
<i>Schmalbach et al. (2004)</i>	57		
<i>Nagata et al. (2003)</i>	19		
<i>Roepman et al. (2005)</i>	102	1	<i>PLAU</i>
<i>Roepman et al. (2006)</i>	825		
<i>Roepman et al. (2005)</i>	102		<i>ADAM12, COL5A3, FAP, NCOR2, CTDSP1, MAN1B1, SERPINE1,</i>
<i>Roepman et al. (2006)</i>	825	14	<i>FTH1, IER3, MMD, TPM2, COL5A1, GREM1, TNFAIP3</i>
<i>Kashiwazaki et al. (2008)</i>	39	1	<i>VTN</i>
<i>Roepman et al. (2006)</i>	825		
<i>Schmalbach et al. (2004)</i>	57	6	<i>PCOLCE, GPX7, RARRES2, VCAN, PHC2, COL11A1</i>
<i>Roepman et al. (2006)</i>	825		
<i>O'Donnell et al. (2005)</i>	116	2	<i>BASP1, CA6</i>
<i>Roepman et al. (2006)</i>	825		
<i>Nagata et al. (2003)</i>	19	7	<i>MMP2, TNC, ITGB4, FN1, CCND1, MMP1, IL6</i>
<i>Roepman et al. (2006)</i>	825		
<i>Méndez et al. (2007)</i>	62	1	<i>IRF1</i>
<i>Nguyen et al. (2007)</i>	85		
<i>Kashiwazaki et al. (2008)</i>	39	1	<i>MMP3</i>
<i>Nagata et al. (2003)</i>	19		

Quadro 5 - Transcritos diferencialmente hipoexpressos em tumores metastáticos encontrados em diferentes estudos

Trabalhos	Número de transcritos em comum	Transcritos
O'Donnel et al. (2005) Roepman et al. (2005) Roepman et al. (2006) Schmalbach et al. (2004)	1	TGM3
Méndez et al. (2007) Roepman et al. (2005) Roepman et al. (2006)	1	PI3
Roepman et al. (2005) Roepman et al. (2006)	39	AZIN2, PPL, PDZK1IP1, PGLYRP2, STK24-AS1, ZDHHC18, SMC5, OSBP2, SSH2, LRRN2, PLK2, MMP25, DMKN, LYPD3, SLPI, S100A7, SPINK7, HPCAL1, LLGL2, JMJD7-PLA2G4B, COPG1, E2F5, IDI1, PARVB, CD40, MALL, IVL, IFFO2, SRP19, IL22RA1, FBXL19, MAL2, KRT23, DSG3, ALOX12B, HOPX, S100A9, SPINK5, RAD17
Roepman et al. (2005) Schmalbach et al. (2004)	2	PPT2, ECM1
Nguyen et al. (2007) Roepman et al. (2006)	1	RAB11FIP1
Kashiwazaki et al. (2008) Roepman et al. (2006)	1	JAK1
Schmalbach et al. (2004) Roepman et al. (2006)	5	SERPINB2, EXPH5, LCN2, ABLIM1, FGFBP1
Méndez et al. (2007) Roepman et al. (2006)	3	FOS, KRT16, PPIF
Roepman et al. (2006) Zhou et al. (2006)	1	MTUS1
O'Donnel et al. (2005) Roepman et al. (2006)	1	NTRK2
Nagata et al. (2003) Nguyen et al. (2007)	1	IGFBP2
Schmalbach et al. (2004) Zhou et al. (2006)	1	PLA2G4A

O resultado dessa comparação pode ser decorrente da análise de amostras de diferentes sub-sítios (orofaringe p.ex.), inclusão de cortes pequenas e utilização de tecnologias de detecção de transcritos com baixa resolução, já que todos os estudos basearam suas descobertas em tecnologias de *microarray* (LALLEMANT et al. 2010) (**Quadro 3**).

Por muitos anos a investigação do transcriptoma ficou limitada à capacidade técnica de *microarrays* e ensaios de PCR que funcionam baseados no uso de sondas ou iniciadores complementares à sequências

conhecidas dos transcritos. Embora essas metodologias possuam um grande poder de determinação de mudanças transcricionais, ela se limita a identificar apenas transcritos previamente conhecidos. O advento do sequenciamento massivo em paralelo do RNA (RNASeq), permitiu que não só transcritos conhecidos sejam detectados, mas também a identificação de novos RNAs e, até mesmo, alcançar a detecção da totalidade das moléculas transcritas em um tecido (REIS et al. 2005; TUCKER et al. 2009; WANG et al. 2009).

Apesar das várias abordagens relacionadas ao padrão transcricional da metástase em CEB, acreditamos que o uso de plataformas de sequenciamento em larga escala poderá prover um perfil de expressão gênica CEB metastático mais completo permitindo a identificação de transcritos com maior poder preditivo. Além disso, uma abordagem de sequenciamento em larga escala em busca de biomarcadores de metástase nunca foi realizada em pacientes com CEB da população brasileira, detentora de uma das maiores incidências desta doença no mundo.

2 OBJETIVOS

Os objetivos deste estudo foram:

- Determinar, a partir de sequenciamento em larga escala diferenças no perfil transcricional de carcinomas epidermóides de boca metastáticos e não-metastáticos de modo a identificar transcritos com valor para detecção da metástase linfonodal.
- Validar os achados do primeiro objetivo em amostras independentes, com o intuito de determinar um grupo de marcadores com provável importância clínica.

3 MATERIAIS E MÉTODOS

Este trabalho foi dividido em duas partes, sendo a primeira correspondente à fase de descoberta de biomarcadores de metástase nodal no CEB e a segunda refere-se à validação de prováveis biomarcadores. A primeira etapa foi baseada na tecnologia de sequenciamento de RNAs (RNASeq) utilizando um número restrito de amostras. Na etapa de validação lançamos mão de tecnologias de validação *in silico* e PCR em tempo real.

3.1 ASPECTOS ÉTICOS

Este estudo é um dos subprojetos originais do projeto temático “Fatores ambientais, clínicos, histopatológicos e moleculares associados ao desenvolvimento e ao prognóstico de carcinomas epidermóides de cabeça e pescoço” financiado pela Fundação de Amparo à Pesquisa do Estado de São Paulo (FAPESP) (2010/51168-0), sob a coordenação geral da Dra. Eloiza Tajara, da Faculdade de Medicina de São José do Rio Preto, SP. Por sua vez, este é uma continuidade do projeto temático anterior “Busca de marcadores de agressividade em tumores de cabeça e pescoço” também financiado pela FAPESP e sob a mesma coordenação. Ambos projetos foram submetidos aos comitês de ética de instituições participantes e também ao CONEP, tendo sido aprovados. O projeto atual foi aprovado pelos comitês de ética da Faculdade de Odontologia da UNESP (Campus de

SJ dos Campos - reunião de 12 de abril de 2011, parecer 04/2011) e da Faculdade de Medicina de São José do Rio Preto (parecer 099/2011, protocolo 2670/2011, aprovado em 18/04/2011). O subprojeto específico, ao qual este projeto de doutoramento refere-se, foi aprovado pelo Comitê de Ética em Pesquisas da Fundação Antônio Prudente (protocolo 1621/11, aprovado em 03/11/2011). (Anexo 1)

3.2 COLETA DE AMOSTRAS

Todos os tecidos utilizados neste estudo foram coletados de pacientes operados em alguma das instituições envolvidas no projeto temático, incluindo também casos do Departamento de Cirurgia de Cabeça e Pescoço do A.C.Camargo Cancer Center, São Paulo, SP. Sendo assim, as amostras estudadas provêm tanto do banco de amostras do projeto temático como também do banco de tumores do A.C.Camargo Cancer Center. O protocolo de coleta de amostras seguiu critérios padronizados baseados no congelamento imediato do fragmento tumoral em nitrogênio líquido imediatamente após a sua excisão cirúrgica.

3.3 CRITÉRIOS DE INCLUSÃO E EXCLUSÃO DE AMOSTRAS PARA SEQUENCIAMENTO

Neste estudo selecionamos casos de carcinoma epidermóide de boca localizados em língua ou assoalho bucal (CID10 - C02 e C04). Visando

determinar as diferenças no padrão de expressão transcricional entre CEBs metastático e não metastáticos selecionamos dois grupos polares de tumores que apresentavam comportamento biológico claramente distinto. Assim, incluímos amostras de 20 pacientes que se enquadravam em um dos dois grupos de estudo:

- Grupo de tumores grandes, mas não metastáticos: Pacientes com tumores pT3 ou pT4 submetidos ao esvaziamento cervical, sem doença metastática linfonodal comprovada ao exame anatomopatológico (pN0) acompanhados por no mínimo 3 anos após o tratamento.
- Grupo de tumores pequenos e metastáticos: Pacientes com tumores pT1 submetidos ao esvaziamento cervical, com doença metastática comprovada ao exame anatomopatológico (pN+).

Pacientes submetidos a qualquer tipo de tratamento neoadjuvante foram excluídos. Todos os indivíduos participantes assinaram o termo de consentimento livre e esclarecido, previamente aprovado pelos comitês de ética em pesquisa das instituições envolvidas.

3.4 EXTRAÇÃO DE ÁCIDOS NUCLÉICOS

Porções de até 30mg de tecido congelado foram submetidas à criosecção para mensuração da porcentagem do componente tumoral. Foram utilizadas apenas amostras que revelassem porcentagem mínima de 80% de tecido tumoral. Quando possível, os fragmentos foram

macrodissecados para eliminar o máximo de contaminações com tecido não neoplásico.

O RNA total foi extraído utilizando o kit AllPrep DNA/RNA/Protein (Qiagen), que permite a recuperação de DNA, RNA e proteínas. A qualidade dos RNAs obtidos foi verificada no aparelho Bioanalyzer (Agilent, Santa Clara, CA) e a quantidade dos mesmos foi mensurada através da plataforma Qubit 2.0 (Invitrogen). Apenas amostras que apresentassem padrão de qualidade mínimo ($RIN \geq 6$) foram utilizadas nas próximas etapas do estudo (**Figura 6a**).

3.5 PREPARO DE BIBLIOTECAS DE CDNA

Para o sequenciamento do transcriptoma tumoral, foram construídas bibliotecas de cDNA utilizando o Total RNA Seq kit (Life Technologies). Resumidamente, eventuais contaminações com DNA genômico foram removidas por tratamento com DNase (Turbo DNA-free Kit, Ambion) e os RNAs ribossomais (rRNA) foram depletados utilizando o sistema RiboMinus (Life Technologies, USA). Após este passo, as amostras foram analisadas na plataforma 2100 Bioanalyzer (Agilent Technologies) utilizando os kits RNA 6000 Nano, que permite a mensuração da concentração e do tamanho de ácidos nucleicos. Para determinadas amostras que não apresentaram depleção adequada dos rRNAs, este procedimento foi repetido (**Figuras 6b e c**).

Para o início da construção de bibliotecas foram utilizados 300 ng de RNA (livre de DNA e rRNA), previamente mensurado na plataforma Qubit 2.0 RNA HS Assay Kit (Life Technologies).

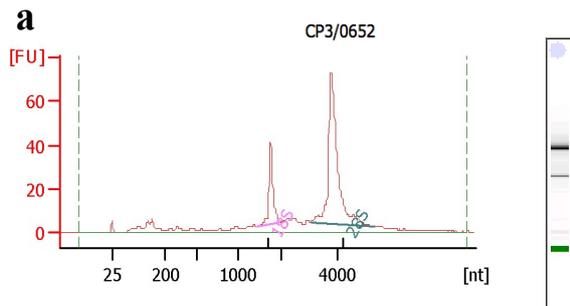
Todos os reagentes descritos a seguir fazem parte do Total RNA Seq Kit (Ambion). Inicialmente, o RNA foi submetido à fragmentação enzimática utilizando RNase III, seguida de purificação utilizando o kit RiboMinus Concentration Module (Invitrogen). A eficiência da fragmentação enzimática foi mensurada utilizando o chip RNA 6000 Pico para Bioanalyzer (Agilent Technologies) e a concentração do RNA fragmentado foi aferida utilizando o RNA HS Assay Kit (Invitrogen) para a plataforma Qubit 2.0. Observamos após a fragmentação, RNAs com tamanho médio variando de 125 a 200 nucleotídeos (nt), como esperado (**Figura 6d**).

Cem nanogramas (ng) de RNA fragmentado foram ligados e hibridizados a adaptadores de sequência simples de oligonucleotídeos. Em seguida, realizou-se a transcrição reversa gerando moléculas de DNA complementar (cDNA).

As moléculas de cDNA foram submetidas a duas etapas de purificação e seleção de tamanho utilizando esferas magnéticas (Agencourt AMPure XP Beads - Beckman Coulter, Inc), com o objetivo de eliminar fragmentos de nucleotídeos de tamanho fora do esperado (abaixo de 200 nt).

O cDNA purificado foi amplificado por PCR, quando foi feita a inserção de fragmentos de nucleotídeos de sequências conhecidas, denominados de *barcodes*, ao cDNA. Este passo permitiu a inclusão de

múltiplas amostras em uma mesma solução nas etapas finais de sequenciamento. Este produto foi novamente purificado utilizando esferas magnéticas (Agencourt AMPure XP Beads - Beckman Coulter, Inc). A eficiência da PCR e da purificação foi mensurada utilizando o chip High Sensitivity DNA Kit (Agilent Technologies) na plataforma Bioanalyzer (Agilent Technologies). Nesta etapa foi avaliada a quantidade de fragmentos com tamanho acima de 150 nt (**Figura 6e**). Amostras com alta concentração (mais de 50%) de fragmentos com tamanho entre 25 e 200 nt, foram submetidas a uma nova etapa de purificação. Essa seleção é importante já que estatísticas de mapeamento de dados de RNA-Seq melhoram quando os fragmentos de biblioteca apresentam tamanho superior a 200 nt. O tamanho médio das bibliotecas deste estudo foi de 260 nt.

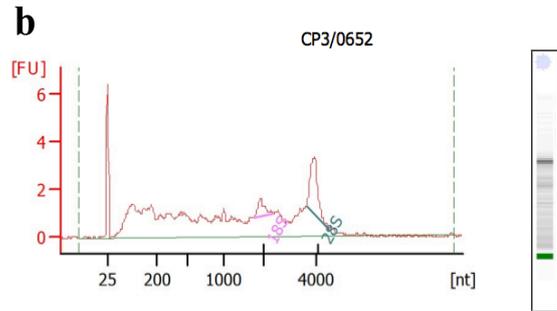


Overall Results for sample 11 : CP3/0652

RNA Area: 454.2
 RNA Concentration: 472 ng/μl
 rRNA Ratio [28s / 18s]: 3.6
 RNA Integrity Number (RIN): 9.3 (B.02.08)
 Result Flagging Color:
 Result Flagging Label: RIN: 9.30

Fragment table for sample 11 : CP3/0652

Name	Start Size [nt]	End Size [nt]	Area	% of total Area
18S	1,467	1,989	50.0	11.0
28S	3,057	5,300	177.8	39.2

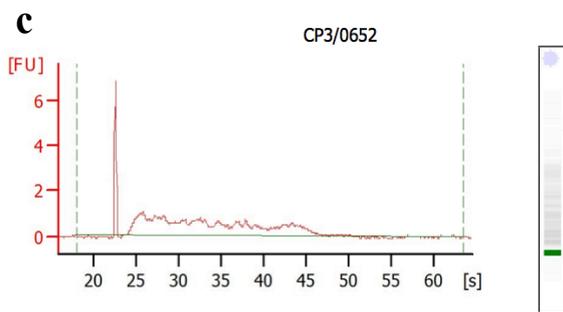


Overall Results for sample 11 : CP3/0652

RNA Area: 73.4
 RNA Concentration: 46 ng/μl
 rRNA Ratio [28s / 18s]: 2.9
 RNA Integrity Number (RIN): 4.4 (B.02.08)
 Result Flagging Color:
 Result Flagging Label: RIN: 4.40

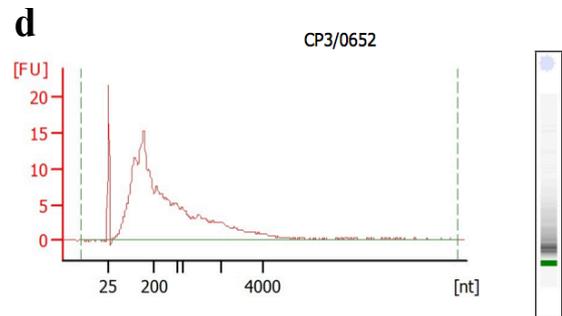
Fragment table for sample 11 : CP3/0652

Name	Start Size [nt]	End Size [nt]	Area	% of total Area
18S	1,758	2,345	1.9	2.6
28S	3,652	4,514	5.5	7.5



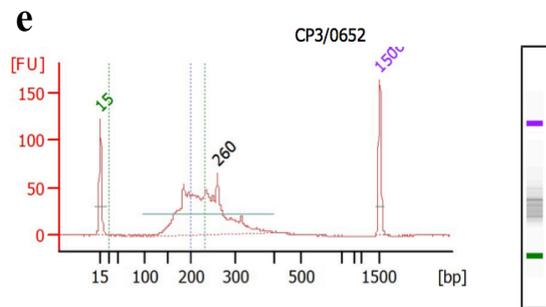
Overall Results for sample 2 : CP3/0652

RNA Area: 37.3
 RNA Concentration: 24 ng/μl
 rRNA Ratio [28s / 18s]: 0.0
 RNA Integrity Number (RIN): 1.6 (B.02.08)
 Result Flagging Color:
 Result Flagging Label: RIN: 1.60



Overall Results for sample 9 : CP3/0652

RNA Area: 285.2
 RNA Concentration: 755 pg/μl
 rRNA Ratio [28s / 18s]: 0.0
 RNA Integrity Number (RIN): 2.4 (B.02.08)
 Result Flagging Color:
 Result Flagging Label: RIN: 2.40



Overall Results for sample 4 : CP3/0652

Number of peaks found: 1
 Area 1: 176.9
 Area 2: 329.6

Region table for sample 4 : CP3/0652

From [bp]	To [bp]	Area	% of Total	Average Size [bp]	Size distribution in CV [%]	Conc. in [ng/μl]	Color
25	200	176.9	27	178	9.0	5.70	
25	233	329.6	50	196	11.9	10.17	

Figura 6 - Eletroferogramas demonstrando o processamento de uma amostra de RNA durante a construção de bibliotecas. a) RNA total revelando ótima relação entre os picos de rRNA 18s e 28s (RIN 9,3); b) RNA após primeira etapa de depleção do rRNA revelando eliminação parcial do produto; c) Segunda etapa de depleção de rRNA revelando eliminação adequada do produto; d) Fragmentação do RNA gerando produtos com pico de tamanho entre 125 e 200 nucleotídeos; e) Produto final da biblioteca de cDNA revelando pico médio de 260 nucleotídeos e baixa porcentagem (9%) de produtos com tamanho menor que 200 nucleotídeos.

Ao final do preparo da biblioteca espera-se ter fragmentos que contenham em suas extremidades adaptadores universais denominados de P1 e P2, onde a extremidade com o adaptador P1 será importante no anelamento de *primers* durante a próxima etapa de amplificação (**Figura 7**).



Figura 7 - Estrutura final da biblioteca de cDNA construída a partir do *Total RNA Seq Kit*. As bibliotecas de cDNA foram sequenciadas baseadas na metodologia “*paired-end multiplex*” onde são sequenciados 75 nucleotídeos no sentido senso e 35 nucleotídeos no sentido anti-senso, além de 5 a 10 nucleotídeos referentes ao código de barras.

3.6 PCR EM EMULSÃO E SEQUENCIAMENTO

Após construção das bibliotecas de cDNA, as amostras foram submetidas a um processo de amplificação denominado PCR em emulsão (ePCR). Neste processo, uma solução aquosa, contendo esferas recobertas por oligonucleotídeos de sequência conhecida (*beads* P1), complementar a um dos adaptadores presentes nas bibliotecas, é misturada a uma solução oleosa. Essa mistura é agitada de forma a gerar gotículas denominadas de micro reatores. Cada micro reator irá conter uma única *bead* P1, um único fragmento de biblioteca, *primers* P1 (complementar à sequência de nucleotídeos presente na superfície da *bead* P1) e P2 e polimerase. Essa solução é então colocada no EZ Bead Amplifier (Life Technologies) que atua como termociclador, permitindo a ocorrência de milhares de PCRs nos

inúmeros micro reatores, gerando produtos que irão recobrir a superfície das *beads* P1. Espera-se, dessa forma, que a amplificação da biblioteca ocorra de forma monoclonal. Neste estudo, utilizamos o kit SOLiD EZ Bead E120 (Life Technologies) que contém 120 bilhões de esferas, permitindo a identificação de 120 bilhões de diferentes sequências da biblioteca de cDNA. Como utilizamos *barcodes* adaptados à biblioteca, foi possível agrupar bibliotecas de 5 diferentes amostras em uma mesma ePCR. Para isso, utilizamos 50 pg/ μ L de cada biblioteca.

Em seguida, o produto da ePCR é purificado utilizando a plataforma EZ Bead Enricher. Esse dispositivo utiliza a complementariedade com a extremidade P2 dos fragmentos da biblioteca para eliminar beads sem fragmentos após a ePCR. Em seguida, a biblioteca ligada à *bead* sofrerá uma modificação na extremidade 3' para permitir que esta se ligue covalentemente a uma lâmina de vidro que servirá de base para o sequenciamento (**Figura 8**). Após deposição desse produto nas lâminas de vidro, as mesmas foram instaladas na plataforma de sequenciamento SOLiD 5500XL.

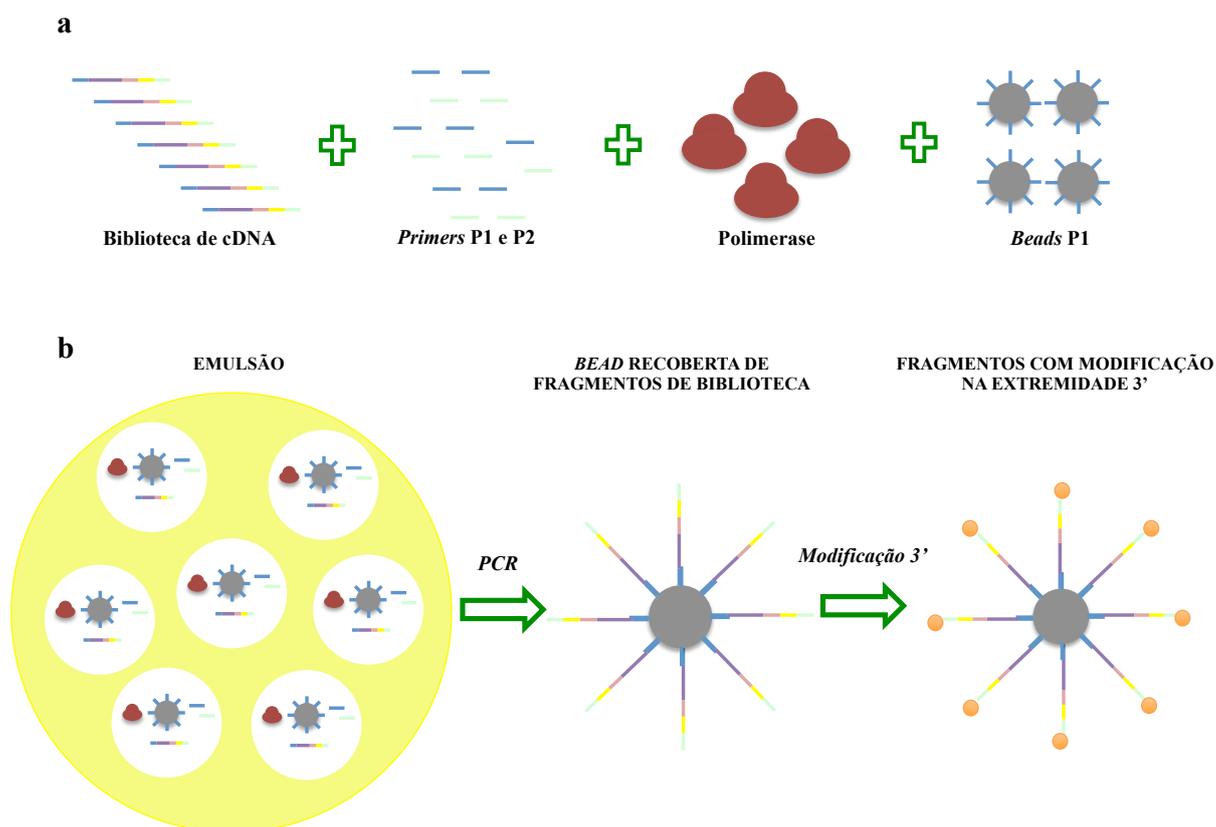
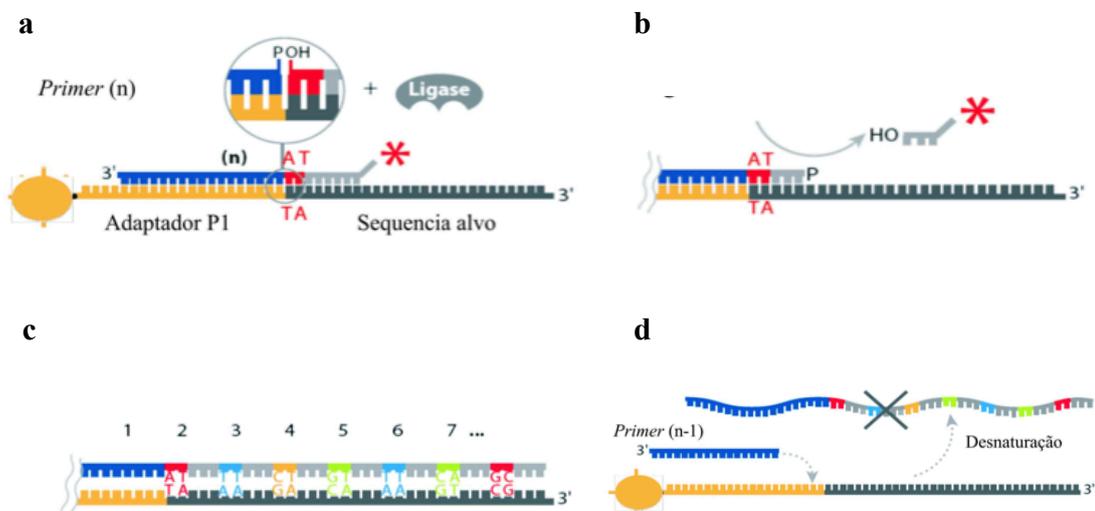


Figura 8 - Representação esquemática da ePCR. a) Para a ePCR são combinados as bibliotecas de cDNA, primers P1 e P2 complementares às extremidades dos fragmentos de biblioteca, polimerase e *beads* com fragmentos P1; b) São gerados micro reatores contendo um único fragmento de biblioteca. Após a PCR cada bead P1 será recoberta por cópias monoclonais de um fragmento de biblioteca. Uma modificação na extremidade 3' dos fragmentos ligados à *bead* permitirá a aderência das mesmas à superfície de sequenciamento.

Na abordagem de sequenciamento escolhida para este trabalho, denominada de *paired-end*, são sequenciados até 75 nucleotídeos no sentido “à frente” (região P1 em direção ao adaptador interno), até 35 nucleotídeos no sentido reverso (região do adaptador interno em direção à região P1), além da região do código de barras que identificará a origem de cada fragmento (**Figura 7**). O sequenciamento nesta plataforma é baseado

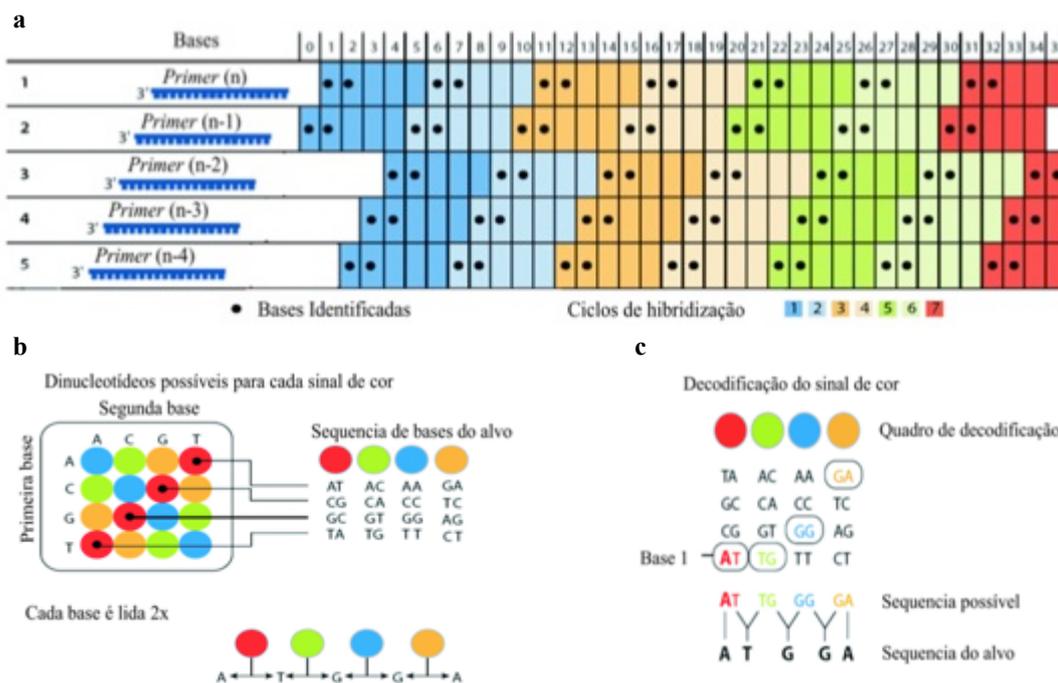
na utilização de *primers*, sondas fluorescentes e enzimas ligase. Os *primers* possuem sequências complementares à regiões que flanqueiam as regiões de interesse (cDNA e código de barras), como a região P1 e o adaptador interno. O processo de sequenciamento se inicia com o anelamento de um primer universal, de tamanho “n”, à extremidade do adaptador P1, por exemplo, imediatamente antes do início da sequência referente à biblioteca. Em seguida, a sonda que for complementar aos primeiros nucleotídeos da biblioteca se hibridará e será ligada ao primer pela ação de uma ligase (**Figura 9a**). Ao se hibridar, haverá a clivagem do fluoróforo, e esta fluorescência será detectada pelo aparelho. Além disso, a extremidade 5’ da sonda ficará exposta permitindo a ligação à sonda que será liberada pelo sequenciador no próximo ciclo (**Figura 9b**). Inicia-se então ciclos sucessivos de ligação de sondas e detecção de fluorescência até que todo o fragmento seja coberto. Entretanto, como apenas dois nucleotídeos de cada sonda são seletivos (as demais bases são degeneradas), uma nova etapa de hibridação de sondas é necessária para cobrir os “*gaps*” deixados pelas bases degeneradas (**Figura 8c**). Dessa forma, após a cobertura de todo o alvo, o fragmento gerado pelas sondas é desnaturado e inicia-se um novo ciclo com um novo primer (**Figura 9d**). Este *primer* terá uma base a menos do que o *primer* da ligação anterior (tamanho do primer é “n-1”) e se anelará à região P1, por exemplo, expondo a última base desta sequência (em decorrência da ausência de um nucleotídeo). O próximo ciclo de ligação de sondas se iniciará por esta base, permitindo a cobertura de regiões que estavam previamente hibridadas com as bases degeneradas. Esse ciclo se

repete com novos *primers* que apresentam tamanho gradativamente menor, perdendo sempre o último nucleotídeo da sequência (n-1, n-2, n-3 e n-4) (**Figura 10a**). Isso permitirá que a totalidade dos nucleotídeos do alvo sejam lidos duas vezes. Esse processo é essencial já que por se usar apenas 4 fluóroforos, uma leitura dupla de cada base é necessária, tendo como base o conhecimento da sequência flanqueadora (P1 ou adaptador interno) para se decifrar a sequência de bases do alvo. (**Figura 10b e c**).



Fonte: Adaptada de CARVALHO e SILVA (2010)

Figura 9 - Química de sequenciamento da plataforma de sequenciamento SOLiD.



Fonte: Adaptada de CARVALHO e SILVA (2010)

Figura 10 - Representação do esquema de sequenciamento na plataforma SOLiD. a) Esquema dos ciclos de anelamento de *primers* e hibridação de sondas de sequenciamento. As bases identificadas em cada ciclo estão marcadas com um círculo, enquanto as demais bases estão sendo ocupadas por nucleotídeos degenerados da sonda de sequenciamento. Note que cada base é lida duas vezes. b) Mapa de fluorescência usado no sistema SOLiD. Cada fluoróforo pode representar 4 diferentes pares de nucleotídeos. A identificação de cada base depende de uma segunda leitura da base adjacente. c) Exemplo de decodificação de uma região sequenciada baseada no mapa de fluorescência.

3.7 ANÁLISE DE BIOINFORMÁTICA

Os dados gerados no sequenciamento foram alinhados utilizando o *software* Lifescope (versão 2.5.1, Life Technologies). Este *software* filtra seqüências de RNA ribossomal e demais contaminações, além de alinhar as seqüências obtidas ao genoma de referência utilizado (versão GRCh37 do genoma humano, *Ensembl*) e exportar esses alinhamentos para arquivos no

formato BAM. Os arquivos BAM foram processados com o *software* HTSeq (versão 0.6.0, opção “*union*”) (ANDERS et al. 2014), para a obtenção do número de *counts* para cada gene presente na versão GRCh37 do genoma humano (*Ensembl*). Foram então excluídos das análises os genes que apresentavam um total de *reads* menor que 5, uma média de *reads* por amostra menor que 0.2 e *counts* por milhão de reads mapeados (cpm) menor que 1.

Para determinação dos genes diferencialmente expressos, foi utilizado o algoritmo DESeq2 (LOVE et al. 2014). Foram aplicados os parâmetros padrões do *software* para a normalização dos dados de expressão e para a estimação dos parâmetros do modelo estatístico empregado. Foram considerados como diferencialmente expressos aqueles genes cuja análise estatística dos níveis de expressão, comparados entre os grupos analisados, apresentou p-valor $< 0,01$ e fdr (*false discovery ratio*) $< 0,1$.

3.8 VALIDAÇÃO *IN SILICO*

Com o objetivo de desafiar nossos resultados em uma nova coorte de casos, utilizamos uma abordagem de validação *in silico* baseado em dados de sequenciamento gerados pelo *The Cancer Genome Atlas* (TCGA) (LAWRENCE et al. 2015). Para isto, realizamos 3 seleções de grupos de pacientes utilizando diferentes critérios de inclusão. No primeiro grupo (TCGA_1), incluímos apenas dados de pacientes que apresentavam os

seguintes critérios: indivíduos com 40 anos ou mais, tabagistas e que se enquadravam nos grupos pT1/pN+ ou pT3-4/pN0. Nesta seleção obtivemos 3 casos de tumores metastáticos e 25 não metastáticos. No segundo grupo (TCGA_2), afrouxamos o critério tabagismo, incluindo também não-fumantes. Nesta seleção obtivemos 6 casos de tumores metastáticos e 35 não metastáticos. No terceiro grupo (TCGA_3), além de incluirmos não fumantes, incluímos também pacientes com tumores T2/N+ ao grupo metastático. Nesta seleção obtivemos dados de 31 tumores metastáticos e de 35 tumores não metastáticos.

Em seguida, foram realizadas análises de expressão diferencial entre os grupos metastáticos e não metastáticos de cada uma das três seleções do TCGA. Para tal análise, foram obtidas tabelas de “*counts*” de sequências para genes presentes na anotação RefSeq do genoma humano a partir do website do TCGA (<http://cancergenome.nih.gov/>). A filtragem de genes do TCGA assim como a seleção dos genes diferencialmente expressos foi realizada como descrito anteriormente para os casos sequenciados neste estudo (algoritmo DESeq2). Por fim, os resultados obtidos nesta análise foram cruzados com os resultados do nosso estudo, permitindo a identificação dos genes cuja expressão diferencial foi confirmada em ambas análises.

Considerando a possibilidade de que a seleção descrita a cima possa estar favorecendo seleção de genes que apresentam expressão diferencial em decorrência do tamanho tumoral, realizamos uma nova seleção de casos do TCGA ignorando esta variável (tamanho do tumor). Dessa forma, dois

novos grupos do TCGA foram selecionados para comparação: o primeiro foi constituído da totalidade de casos de carcinoma epidermóide de cavidade oral (C02, C03, C04, C05 e C06), já o segundo foi restrito aos casos de língua (C02) e assoalho bucal (C04).

O grupo formado por tumores de língua e assoalho bucal de todos os tamanhos, por representar um grupo mais homogêneo e eliminar um possível fator de confusão, foi escolhido como base para a determinação dos genes candidatos.

Os transcritos foram classificados como validados se apresentassem p-valor menor que 0,01 e fdr menor que 0,1 em ambos os grupos amostrais (amostras deste estudo e amostras do TCGA) e se o sentido da diferença de expressão (aumento ou diminuição da expressão) também fosse o mesmo.

3.9 DETERMINAÇÃO DE GENES CANDIDATOS

A partir do grupo de genes validados através dos dados do TCGA, realizamos um teste para verificação daqueles transcritos com maior potencial de discriminação entre tumores metastáticos e não metastáticos. Isso foi feito a partir da determinação da sensibilidade e especificidade do uso de cada marcador e todas as suas possíveis combinações a partir da realização de curvas ROC usando os dados de sequenciamento gerados neste trabalho. Para análises univariadas, considerando um único transcrito, foi utilizado o software GraphPad Prism versão 6.0 para MAC OS X. Para análises multivariadas, testando o efeito da combinação de diferentes

transcritos, foi utilizada a ferramenta online *ROCET: ROC Curve Explorer & Tester* (www.rocet.ca) (XIA et al. 2013). Os transcritos com melhores valores de área abaixo da curva (AUC) foram selecionados para as etapas seguintes de validação.

3.10 PCR EM TEMPO REAL QUANTITATIVO

Baseados nos resultados da curva ROC, escolhemos ensaios de PCR em tempo real quantitativo (qRT-PCR) baseados na tecnologia *Taqman*, para os 3 genes com maior potencial de representar biomarcadores de metástase nodal. Para correção dos valores de expressão gênica utilizamos um ensaio *Taqman* para o gene GAPDH como controle endógeno. Todas as reações foram realizadas em duplicata a partir de uma abordagem “*multiplex*”, permitindo uma aferição mais precisa do nível de expressão dos genes de interesse. Os ensaios selecionados e suas características estão descritas na **Quadro 6**.

Quadro 6 - Ensaios *Taqman* utilizados para mensuração da expressão dos genes candidatos a biomarcadores e do controle endógeno

Transcrito alvo	Ensembl Id	Identificação do ensaio	Tamanho do amplicon (nt)	Fluoróforo	Região de hibridação da sonda
<i>NR3C2</i>	ENSG00000151623	Hs01031809_m1	68	FAM	Entre éxons
<i>FADD</i>	ENSG00000168040	Hs04187499_m1	118	FAM	Entre éxons
<i>CCND1</i>	ENSG00000110092	Hs00765553_m1	57	FAM	Entre éxons
<i>GAPDH</i>	ENSG00000111640	4326317E	122	VIC	Éxon 3

Para o cálculo dos níveis de expressão gênica a partir dos ensaios de qPCR foram calculados os Cts de cada amostra a partir da média de cada duplicata. Os cálculos para a determinação do *Fold change* (FC) estão descritos na **Figura 11**. O ensaio foi repetido todas as vezes que a diferença de Ct entre as duplicatas fosse maior que 1,0.

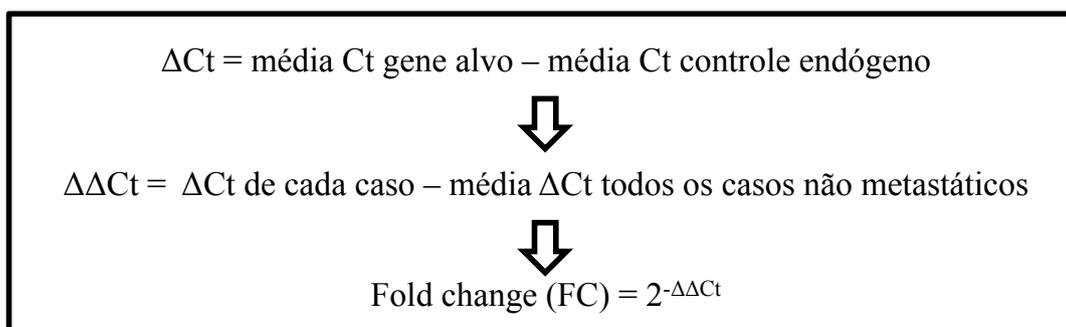


Figura 11 - Fórmulas utilizadas para cálculo dos níveis de expressão gênica a partir dos ensaios de qPCR.

3.11 ANÁLISE ESTATÍSTICA

Foi utilizado o teste não paramétrico de Mann-Whitney para verificar associações entre duas variáveis contínuas, ou teste de Kruskal-Wallis para verificar associações entre 3 ou mais variáveis contínuas. O teste do qui-quadrado ou teste exato de Fisher foi utilizado para verificar associações entre variáveis categóricas. Associações que revelassem p-valor < 0,05 foram consideradas estatisticamente significativas.

4 RESULTADOS E DISCUSSÃO

4.1 CARACTERÍSTICAS CLÍNICO-PATOLÓGICAS DAS AMOSTRAS SEQUENCIADAS

As amostras utilizadas neste estudo, suas características clínicas e o respectivo agrupamento estão descritos no **Quadro 7**. Foram incluídas 10 amostras no grupo não metastático (N0) e 10 amostras no grupo metastático (N+). A idade média foi de 57,2 anos para o grupo N0 e 58,4 anos para o grupo N+, sendo esta diferença não significativa ($p=0,69$). A distribuição dos gêneros foi idêntica, sendo de 8 homens e 2 mulheres em cada grupo. Considerando a localização do tumor primário, 5 tumores do grupo N+ localizavam-se no assoalho de boca (C04) e 5 na língua oral (C02). Todos os casos do grupo N0 se desenvolveram na língua oral. A distribuição de tumores estágio III e IV foi idêntica entre os grupos, sendo 4 e 6 casos, respectivamente, em cada grupo. Embora nossos grupos se diferenciem em relação à localização do sítio primário, sabe-se que tumores da língua e assoalho bucal possuem comportamento bastante semelhante (SHAH e GIL 2009). Além disso, no caso de tumores grandes, a determinação do sítio primário de desenvolvimento tumoral (língua ou assoalho bucal) é extremamente limitada pelo frequente envolvimento dessas duas estruturas anatômicas durante o crescimento tumoral. Assim, apesar dessa diferença,

nossos grupos de estudo são bastante homogêneos, reduzindo o risco de introdução de vieses clínicos na análise.

As características histopatológicas dos tumores estudados (**Quadro 8**) foram obtidas retrospectivamente a partir do laudo anatomopatológico (casos obtidos no ACCCC) e a partir do banco de dados do projeto temático GENCAPO. Apesar dessa análise se basear em critérios subjetivos, mensurados por diferentes indivíduos e sujeita à variações na metodologia de avaliação, não foram observadas diferenças significativas entre os grupos de estudo para as variáveis histopatológicas. É importante destacar que não observamos diferenças significativas na distribuição de tumores entre os grupos considerando a gradação histopatológica tumoral. Observamos o mesmo número de tumores bem diferenciados nos dois grupos de estudo, indicando a limitada utilidade do fenótipo histopatológico na mensuração de risco para desenvolvimento metastático.

A mensuração da profundidade de invasão mostrou-se também um critério limitado para predição da ocorrência de disseminação metastática, já que mesmo um tumor micro-invasivo (CP3/0012) gerou metástase nodal. Características importantes como presença de reação desmoplásica e intensidade de infiltrado inflamatório peritumoral não puderam ser adequadamente avaliadas devido ao caráter retrospectivo da análise do caráter morfológico destes tumores. Essas informações nem sempre são contempladas nos laudos anatomopatológicos.

Além disso, a maioria dos critérios descritos na **Quadro 8** são extremamente dependentes da extensão da avaliação histopatológica do

tumor, como por exemplo, presença de invasão vascular, linfática e perineural. Dessa forma, apesar da aparente homogeneidade, a variabilidade tumoral dificilmente será contemplada fidedignamente neste quadro.

Quadro 7 - Características clínicas dos pacientes cujas amostras foram avaliadas neste estudo

Amostra	Idade ao diagnóstico	Gênero	Localização	pT	pN	M	Estadiamento	Grupo	Origem da amostra
CP1/0151	47	M	C02	3	0	0	III	N0	GENCAPO
CP2/1022	44	M	C02	3	0	0	III	N0	GENCAPO
CP1/0291	55	M	C02	4	0	0	IV	N0	GENCAPO
CP3/0486	56	M	C02	3	0	X	III	N0	GENCAPO
CP103T	75	M	C02	3	0	X	III	N0	ACC
CO102T	39	F	C02	4	0	X	IV	N0	ACC
CO106T	59	M	C02	4	0	X	IV	N0	ACC
CO100T	80	F	C02	4	0	X	IV	N0	ACC
CO40T	69	M	C02	4	0	X	IV	N0	ACC
CP1/0040	48	M	C02	4	0	0	IV	N0	GENCAPO
CP3/0652	63	M	C04	1	1	X	III	N+	GENCAPO
CP3/0012	50	M	C04	1	1	0	III	N+	GENCAPO
CO21T	66	M	C02	1	1	X	III	N+	ACC
CO16T	51	M	C02	1	1	X	III	N+	ACC
CP3/0280	52	M	C04	1	2b	X	IV	N+	GENCAPO
CO30T	40	M	C04	1	2b	X	IV	N+	ACC
CO85T	75	F	C02	1	2b	X	IV	N+	ACC
CO31T	72	F	C02	1	2b	X	IV	N+	ACC
CP3/0613	56	M	C02	1	2b	X	IV	N+	GENCAPO
CP1/0397	59	M	C04	1	3	0	IV	N+	GENCAPO

M - masculino; F - feminino; ACC - AC Camargo Cancer Center

Quadro 8 - Características patológicas dos tumores utilizados no sequenciamento

Caso	Gradação histopatológica	Profundidade de invasão	Reação desmoplásica	Invasão sanguínea	Invasão linfática	Invasão perineural	IIP	NLD	NLA	Localização	EC
CP1/0040	Bem diferenciado	Plano muscular	NI	Não detectada	Não detectada	Presente	Moderado	37	0	-	-
CP1/0151	Moderadamente diferenciado	Plano muscular	NI	Não detectada	Não detectada	Presente	Moderado	21	0	-	-
CP1/0291	Bem diferenciado	Plano muscular	NI	Não detectada	Não detectada	Não detectada	Escasso	36	0	-	-
CP2/1022	Bem diferenciado	Plano muscular	NI	Não detectada	Presente	Não detectada	Intenso	4	0	-	-
CP3/0486	Bem diferenciado	NI	NI	Não detectada	Não detectada	Presente	Escasso	41	0	-	-
CP103T	Bem diferenciado	Plano muscular	Moderada	Não detectada	Não detectada	Presente	Intenso	41	0	-	-
CO102T	Bem diferenciado	Plano muscular	NI	Não detectada	Não detectada	Não detectada	NI	73	0	-	-
CO106T	Moderadamente diferenciado	Plano muscular	Moderada	Não detectada	Não detectada	Não detectada	Escasso	35	0	-	-
CO100T	Moderadamente diferenciado	Plano muscular	Moderada	Não detectada	Não detectada	Presente	Moderado	10	0	-	-
CO40T	Moderadamente diferenciado	Plano muscular	Moderada	Não detectada	Não detectada	Presente	Moderado	49	0	-	-
CP3/0012	Bem diferenciado	Micro invasivo	NI	Não detectada	Não detectada	Não detectada	Ausente	89	1	Nível I	Não
CP3/0652	Moderadamente diferenciado	Lâmina própria	NI	Presente	Presente	NI	Moderado	42	1	Nível II	Não
CO21T	Bem diferenciado	Plano muscular	NI	Não detectada	Não detectada	Não detectada	NI	10	1	Nível I	Não
CO16T	Moderadamente diferenciado	Plano muscular	NI	Não detectada	Não detectada	Não detectada	NI	55	1	Nível II	Não
CP3/0613	Bem diferenciado	Plano muscular	NI	Não detectada	Não detectada	Presente	Moderado	21	2	Nível II	Sim
CO30T	Moderadamente diferenciado	Lâmina própria	NI	Não detectada	Não detectada	Não detectada	NI	35	2	Nível IB e IV	Não
CO31T	Moderadamente diferenciado	Lâmina própria	NI	Não detectada	Presente	Não detectada	NI	20	2	Nível I e II	Sim
CP3/0280	Moderadamente diferenciado	Plano muscular	NI	Não detectada	Não detectada	Presente	NI	46	3	Nível I e II	Não
CO85T	Bem diferenciado	Plano muscular	NI	Não detectada	Não detectada	Não detectada	NI	30	5	Nível II	Sim
CP1/0397	Pouco diferenciado	Lâmina própria	NI	Não detectada	Presente	Presente	Escasso	25	6	Níveis II, III e IV	Não

IIP - Infiltrado inflamatório peritumoral; NLD: número de linfonodos dissecados; NLA: número de linfonodos acometidos; EC: extravasamento capsular; NI: não informado no exame histopatológico

4.2 RESULTADOS DO SEQUENCIAMENTO DO TRANSCRIPTOMA

O planejamento de sequenciamento do transcriptoma das 20 amostras selecionadas foi baseado em duas corridas *multiplex*, cada uma envolvendo 10 amostras (5 casos N+ e 5 casos N0). Entretanto, enquanto o primeiro grupo de amostras sequenciadas apresentou critérios de qualidade adequados após análise preliminar, durante o sequenciamento do segundo grupo de amostras tivemos problemas técnicos no sequenciador que comprometeram a qualidade dos resultados. Este problema ocasionou a geração de sequências com tamanho significativamente menor que o esperado (25-50 nucleotídeos). Dessa forma, foi necessária a realização de um terceiro sequenciamento do segundo grupo de amostras. Embora este terceiro experimento tenha gerado sequências com tamanho maior (25-69 nucleotídeos), obtivemos queda expressiva na qualidade das leituras. Todos os problemas ocorridos durante esses experimentos foram decorrentes de problemas técnicos na plataforma SOLiD 5500XL.

Com o intuito de melhorar os aspectos matemáticos relativos à análise de dados, decidimos manter ambas corridas referentes ao segundo grupo de sequenciamento, somando-as já que tratam-se de réplicas técnicas. Os achados destes sequenciamentos estão descritos na **Quadro 9**.

Geramos ao todo cerca de 3 bilhões de sequências, representando uma média de 153 milhões de fragmentos por amostra. O menor número de sequências foi de 99 milhões de fragmentos para a amostra CO85T e o

maior foi 296 milhões para a amostra CO21T. Esses valores estão dentro da extensão de cobertura recomendada para análises de expressão diferencial e descoberta de novos transcritos, que de acordo com o *Standards, Guidelines and Best Practices for RNA-Seq V1.0* (de junho de 2011) (https://genome.ucsc.edu/ENCODE/protocols/dataStandards/ENCODE_RNAseq_Standards_V1.0.pdf) do *The ENCODE Consortium*, deve variar entre 100 e 200 milhões de fragmentos por amostra.

É importante destacar os resultados de TARAZONA et al. (2011) que ressaltam que o aumento da profundidade de sequenciamento irá favorecer apenas a detecção de transcritos não codificadores de proteínas e transcritos de baixa expressão com provável função regulatória.

A distribuição do número de sequências de acordo com os grupos de estudo esta exposta na **Figura 12**. Observamos um maior número de sequências no grupo de tumores metastático, representando 61% do total de sequências, em relação ao grupo de tumores não-metastáticos, representando 39% do total de sequências. A diferença no número de sequências entre os dois grupos foi significativa ($p=0,01$).

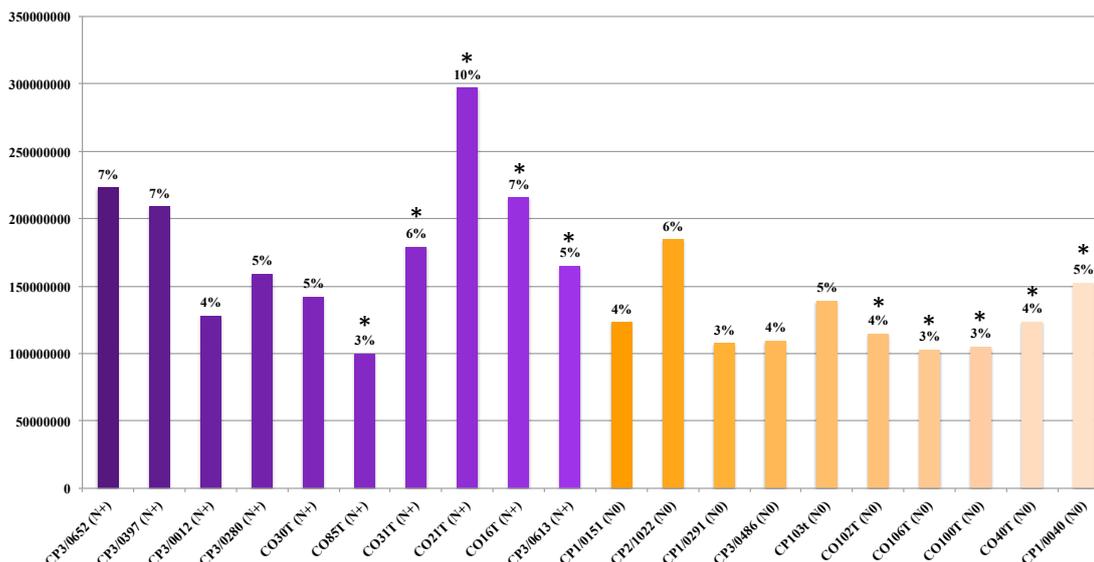


Figura 12 - Distribuição do número de fragmentos sequenciados por amostra de acordo com o grupo de estudo. *Amostras sequenciadas duas vezes. Os valores referem-se a soma das duas corridas.

Na contabilidade geral para todas as amostras, foram sequenciados 18815 transcritos únicos, sendo classificados, de acordo com o sistema de anotação do *Ensembl*, nos seguintes grupos:

- Codificadores de proteína - *Protein coding* - (PC);
- Pseudogenes - *Pseudogene* - (PsG);
- Não codificadores de proteína longos - *Long noncoding* - (LNC);
- Não codificadores de proteína curtos - *Short noncoding* - (SNC).

As subdivisões de cada um desses grupos, assim como os critérios de classificação estão descritos na **Quadro 10**. A distribuição dos transcritos nestes grupos e respectivos subgrupos está representada a seguir nas **Figuras 13 e 14**, respectivamente.

Quadro 9 - Dados globais das corridas de sequenciamento

Amostra	Grupo	Total de sequências por corrida	Corrida	Total de sequências por amostra	Qualidade média	Tamanho das sequências
CP3/0652	N+	223099890	1	223099890	31,1	25-75
CP3/0397	N+	208867300	1	208867300	30,1	25-75
CP1/0151	N0	123026366	1	123026366	29,7	25-75
CP3/0012	N+	127654066	1	127654066	30,7	25-75
CP2/1022	N0	184486268	1	184486268	29,9	25-75
CP3/0280	N+	158983990	1	158983990	31,0	25-75
CP1/0291	N0	107596918	1	107596918	29,7	25-75
CP3/0486	N0	109264726	1	109264726	30,3	25-75
CO30T	N+	141798754	1	141798754	30,6	25-75
CP103T	N0	138997596	1	138997596	29,5	25-75
CO102T	N0	34019407	2	114467039	35,4	25-50
		80447632	3		25,4	25-69
CO106T	N0	30861554	2	102689648	34,4	25-50
		71828094	3		24,3	25-69
CO100T	N0	37857572	2	104987174	30,2	25-50
		67129602	3		24,9	25-68
CO40T	N0	48813428	2	123206158	28,8	25-50
		74392730	3		24,4	25-68
CP1/0040	N0	46113543	2	152103569	35,5	25-50
		105990026	3		25,8	25-69
CO85T	N+	33715931	2	99836305	35,4	25-50
		66120374	3		25,7	25-69
CO31T	N+	87044268	2	179037238	30,4	25-50
		91992970	3		24,8	25-68
CO21T	N+	147851644	2	296969618	29,6	25-50
		149117974	3		25,9	25-68
CO16T	N+	72701650	2	215400724	35,9	25-50
		142699074	3		26,6	25-69
CP3/0613	N+	74039890	2	164683416	31,3	25-50
		90643526	3		25,1	25-68

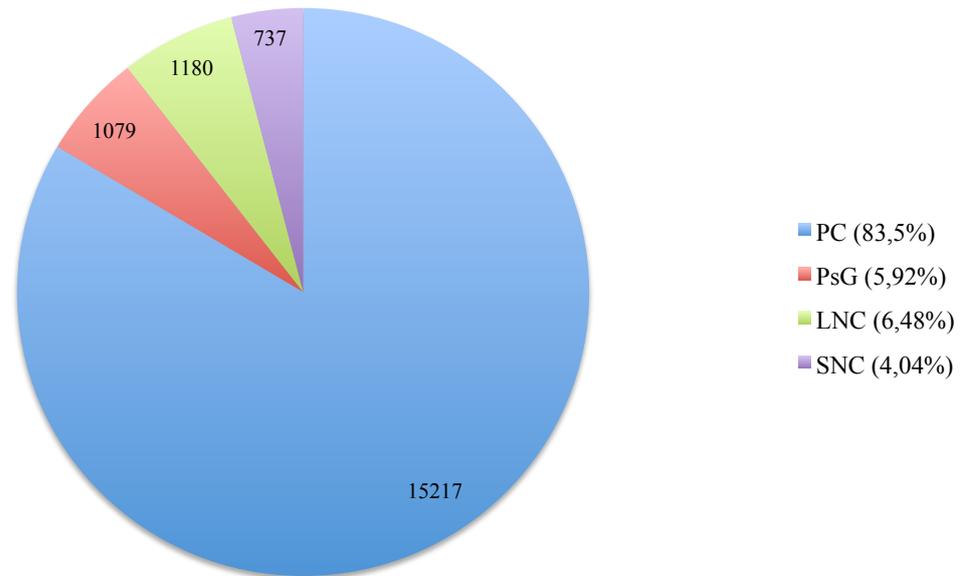


Figura 13 - Classificação em grupos dos 18.815 transcritos identificados no sequenciamento. PC - Codificadores de proteína; PsG - Pseudogenes; LNC - não codificadores longos; SNC - não codificadores curtos.

Quadro 10 - Classificação dos transcritos sequenciados neste estudo

Biótipos de transcritos (Ensembl)	Subtipos	Classificação
Codificadores de proteína	<i>IGC gene</i> <i>IGJ gene</i> <i>IGV gene</i> <i>TRC gene</i> <i>TRV gene</i>	Transcritos referentes à cadeia variável das imunoglobulinas e receptores de células T
	<i>Protein coding</i>	Transcritos que possuem uma <i>Open Reading Frame</i> (ORF)
Pseudogenes	<i>IGC pseudogene</i> <i>IGJ pseudogene</i> <i>IGV pseudogene</i>	Transcritos inativados de imunoglobulinas
	<i>Pseudogene</i>	Possuem similaridade com proteínas, mas geralmente apresentam sequência codificadora corrompida, sendo que um gene homólogo ativo pode ser encontrado em outro locus. Outras vezes, estes podem apresentar uma sequência codificante intacta, mas a ORF é truncada.
	<i>Polymorphic pseudogene</i>	Pseudogene que possui um polimorfismo (SNP ou INDEL)
Não codificadores longos	<i>lincRNA</i>	RNA intergênico longo não codificante
	<i>Sense intronic</i>	RNA longo não codificante transcrito a partir de íntrons de um gene codificante e que não sobrepõe um éxon
	<i>Sense overlapping</i>	Transcrito não codificante longo que contém um gene codificante no seu íntron na mesma fita
	<i>Processed transcript</i>	Transcrito que não possui uma ORF
	<i>Antisense</i>	Transcrito que cobre uma região do genoma referente a um locus gênico, mas na fita oposta do DNA
Não codificadores curtos	<i>miRNA</i>	Transcrito com 21 a 25 nt que atua na regulação da expressão gênica
	<i>miscRNA</i>	Outros RNAs pequenos
	<i>Mt rRNA</i>	RNA ribossomal mitocondrial
	<i>Mt tRNA</i>	RNA transportador transcrito a partir do genoma mitocondrial
	<i>snoRNA</i>	Pequeno RNA nucleolar - guiam modificações durante maturação de outros RNAs
	<i>snRNA</i>	Pequeno RNA nuclear - Presente no núcleo, estão envolvidos com splicing, regulação de fatores de transcrição e manutenção de telômeros
	<i>rRNA</i>	RNA ribossomal

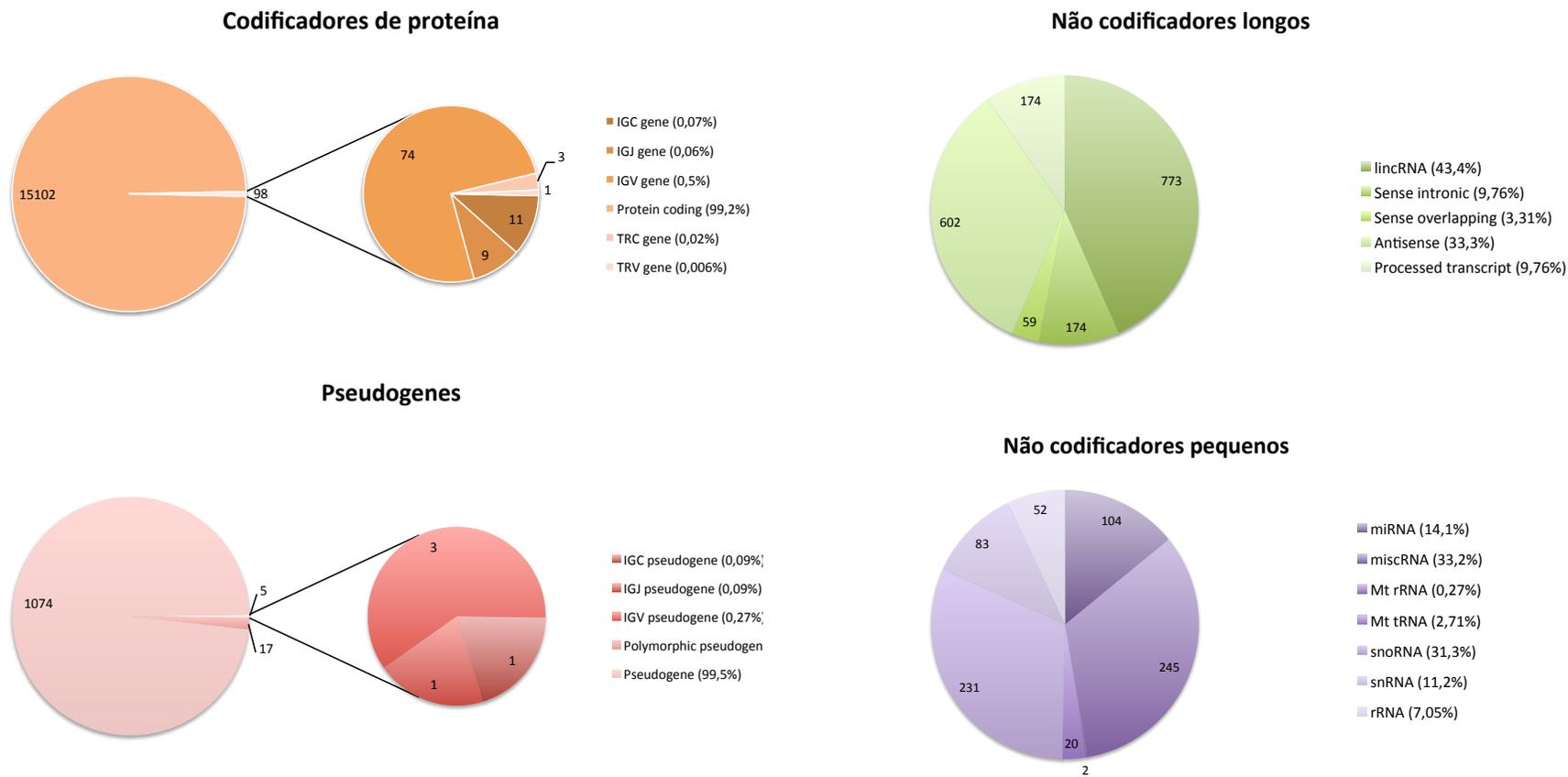


Figura 14 - Classificação em subgrupos dos 18815 transcritos identificados no estudo de acordo com a classificação do *Ensembl*.

De acordo com PALAZZO e LEE (2015) RNAs transportadores (tRNA) e rRNA são os transcritos mais abundantes (em número de moléculas) em células de mamíferos. Na nossa amostra observamos que 7% dos transcritos identificados representavam rRNA. Essa medida representa o número de transcritos únicos identificados e não a quantidade de rRNA detectado. O grau de contaminação com rRNA foi alto, mesmo após a repetição do processo de eliminação destes RNAs com o kit ribominus (*Life Technologies*), indicando uma baixa eficiência deste sistema, o que reduz a eficácia na análise dos transcritos de maior interesse.

Dentre os 100 transcritos mais expressos (em número de fragmentos sequenciados) observamos um grande predomínio de transcritos SNC (cerca de 206 milhões de fragmentos) distribuídos em 37 transcritos únicos. Estes são seguidos de transcritos PC (cerca de 85 milhões de fragmentos) distribuídos em 52 transcritos únicos, e transcritos LNC (cerca de 69 milhões de fragmentos) distribuídos em 8 transcritos únicos (**Quadro 11**). Esses resultados referentes aos RNAs não codificantes corroboram com Palazzo e Lee 2015, que revisaram a literatura, determinando que transcritos SNC são mais abundantes que transcritos LNC.

Quadro 11 - Classes de transcritos mais e menos representadas nos 20 casos de CEB

	Biótipo	Número de transcritos únicos	Total de sequências identificadas
Cem transcritos únicos mais abundantes	<i>Mt rRNA (SNC)</i>	2	107847598
	<i>Protein coding (PC)</i>	52	85947195
	<i>snoRNA (SNC)</i>	33	85779094
	<i>Processed transcript (LNC)</i>	3	61313483
	<i>lincRNA (LNC)</i>	3	27007982
	<i>miscRNA (SNC)</i>	1	11811016
	<i>Antisense (LNC)</i>	3	7692646
	<i>Sense intronic (LNC)</i>	2	2573417
	<i>snRNA (SNC)</i>	1	874726
Cem transcritos únicos menos abundantes	<i>Pseudogene (PsG)</i>	27	6666
	<i>lincRNA (LNC)</i>	28	6506
	<i>Antisense (LNC)</i>	15	4027
	<i>Protein coding (PC)</i>	11	2938
	<i>rRNA (SNC)</i>	7	1763
	<i>Sense intronic (LNC)</i>	4	1048
	<i>snoRNA (SNC)</i>	3	906
	<i>miRNA (SNC)</i>	3	728
	<i>Sense overlapping (LNC)</i>	1	302
	<i>Processed transcript (LNC)</i>	1	298

4.3 DETERMINAÇÃO DOS GENES DIFERENCIALMENTE EXPRESSOS

A partir da análise quantitativa encontramos 150 genes com sugestão de expressão diferencial (DE) entre os grupos N+ e N0. A expressão relativa de cada um destes 150 genes em cada um dos 20 tumores sequenciados está representada na **Figura 15**. É possível observar que a partir do padrão

de expressão destes 150 genes há uma clara separação dos grupos de tumores de acordo com o status linfonodal dos pacientes.

Exceção se faz aos casos CO31T e CO85T. Esses tumores, embora apresentem metástases linfonodais confirmadas ao exame anatomopatológico, se assemelham mais ao padrão de expressão de tumores não metastáticos. Interessantemente, ao revisarmos as características clínico-patológicas referentes a estes pacientes, destaca-se o fato de ambos serem do sexo feminino, negarem histórico de tabagismo ou etilismo durante anamnese e serem diagnosticadas a partir da oitava década de vida. Estes aspectos compreendem um grupo clínico específico de pacientes com CEB formado por mulheres em idade avançada que nunca se expuseram a fatores de risco conhecidos para o desenvolvimento do CEB. Essas pacientes usualmente possuem tumores na língua oral e acredita-se que apresentem uma menor sobrevida doença-específica. Sugere-se que esses tumores apresentem características moleculares específicas, o que explicaria o seu desenvolvimento em um grupo tão peculiar (KOO et al. 2013). Nossos dados suportam essa teoria já que estes dois casos não se encaixam nos padrões de expressão gênica esperados para tumores metastáticos de acordo com os dados deste estudo. Além disso, esses dados favorecem a ideia de que os mecanismos moleculares envolvidos no processo metastático podem variados, mesmo para tumores de uma mesma região anatômica.

De qualquer forma, ressalta-se a especificidade desses genes candidatos para a determinação da doença metastática nos demais

pacientes que relatam exposição a fatores de risco. Mesmo os dois outros pacientes do gênero feminino (CO100T e CO102T) que revelaram histórico de tabagismo à anamnese, são categorizados ao se avaliar o padrão de expressão dos 150 genes DE. Demonstrando que as diferenças observadas para as pacientes CO31T e CO85T são decorrentes de um provável subtipo clínico de CEB e não devido ao gênero.

As classes e subclasses de transcritos diferencialmente expressos estão representadas nas **Figuras 16 e 17**, respectivamente. Assim como destacado para a expressão gênica global, observamos um maior número de genes DE entre o grupo de transcritos PC, sendo seguidos de longe pelos transcritos LNC.

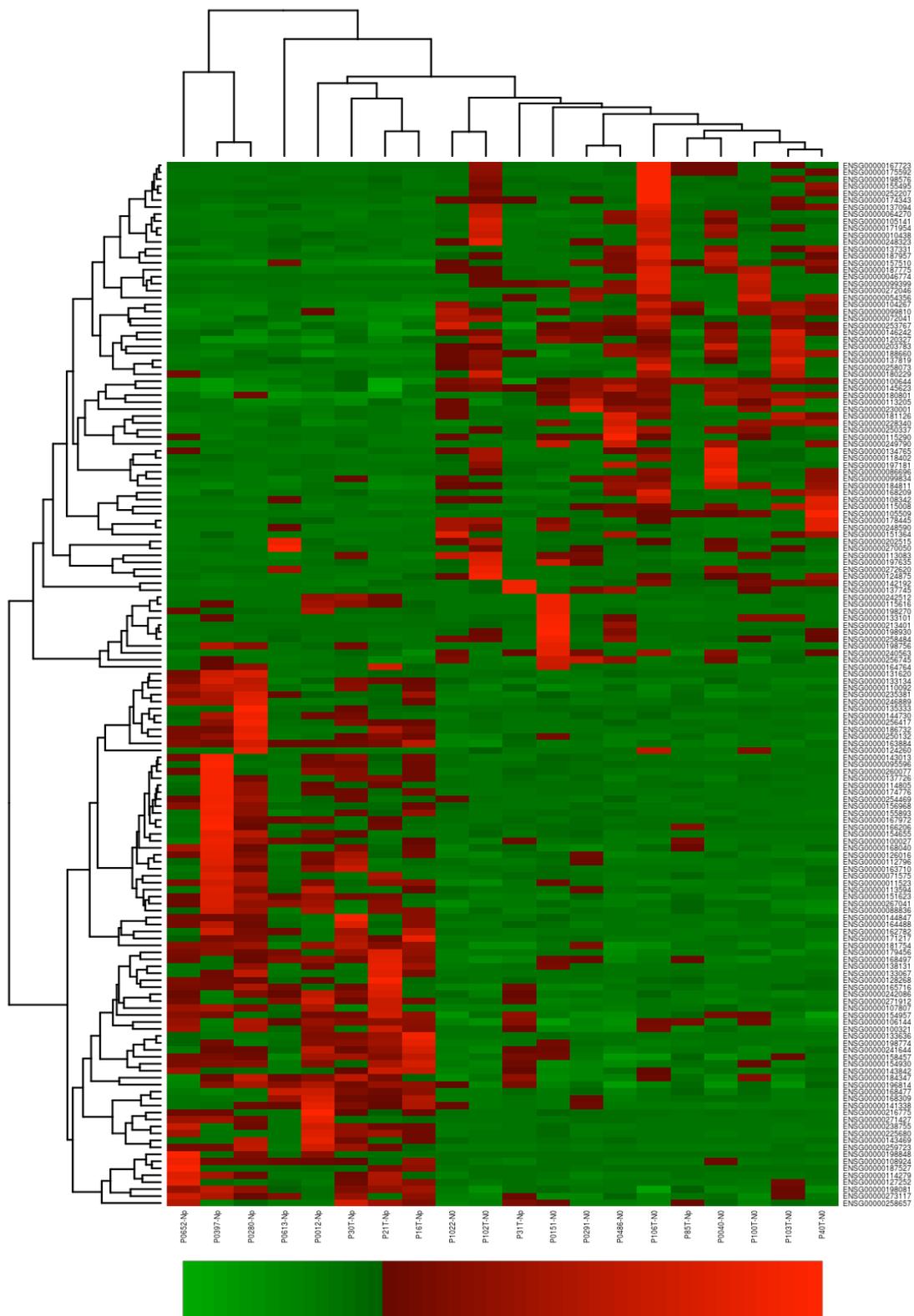


Figura 15 - Heatmap representando a expressão relativa dos 150 genes diferencialmente expressos de acordo com as amostras sequenciadas.

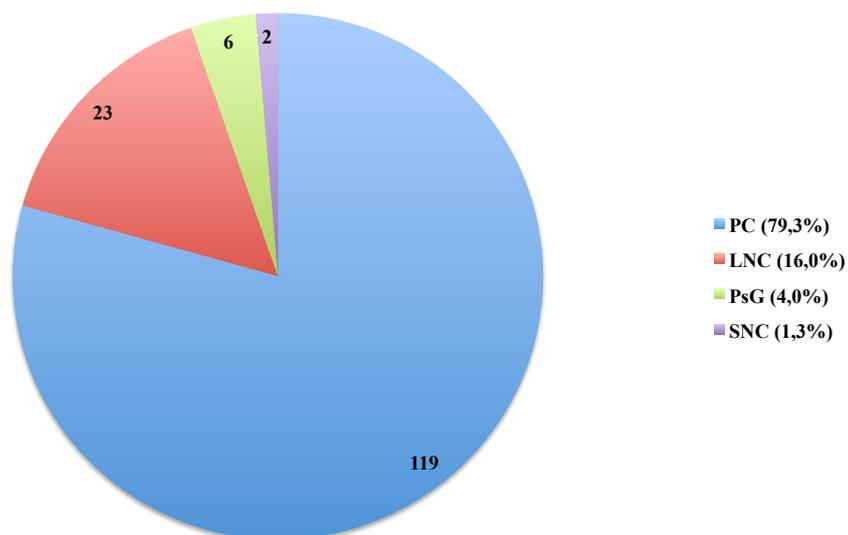


Figura 16 - Classificação em grupos dos 150 transcritos diferencialmente expressos entre tumores metastáticos e tumores não metastáticos. PC - Codificadores de proteína; PsG - Pseudogenes; LNC - não codificadores longos; SNC - não codificadores curtos.

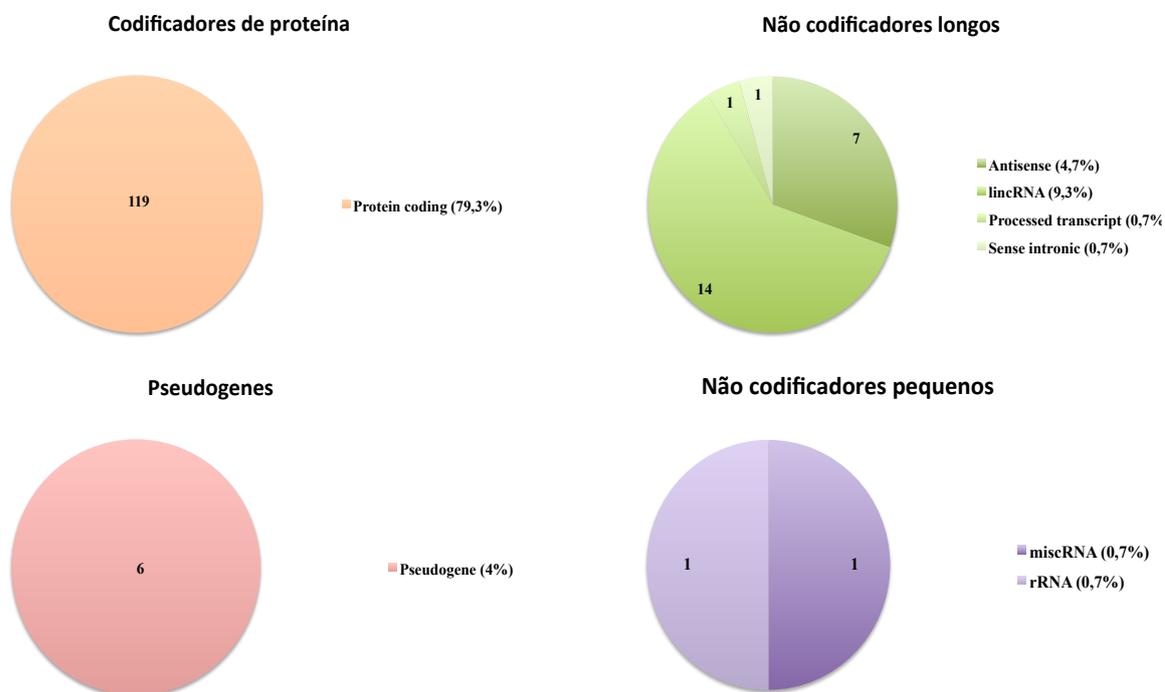


Figura 17 - Classificação em subgrupos dos 150 genes diferencialmente expressos entre os grupos de estudo.

A distribuição dos genes DE em relação ao cromossomo (Chr) de origem está representada na **Figura 18**. Observa-se que entre os tumores metastáticos há um maior número de genes hiperexpressos localizados nos Chr 3, Chr 1 e Chr 6. Entretanto, ao corrigir o número de genes DE pelo número estimado de genes por Chr, temos que os Chr com maior porcentagem relativa de genes hiperexpressos são o Chr3 (1,4%), o Chr18 (1,0%) e o Chr22 (0,8%). Já os Chr com o maior número de genes hipoexpressos são o Chr 5, Chr X, Chr 4. Corrigindo esses valores pelo número de genes em cada Chr, temos que os Chr 21 (1,0%), Chr 5 (1,0%) e Chr X (0,8%) apresentam a maior porcentagem relativa de genes hipoexpressos em tumores metastáticos.

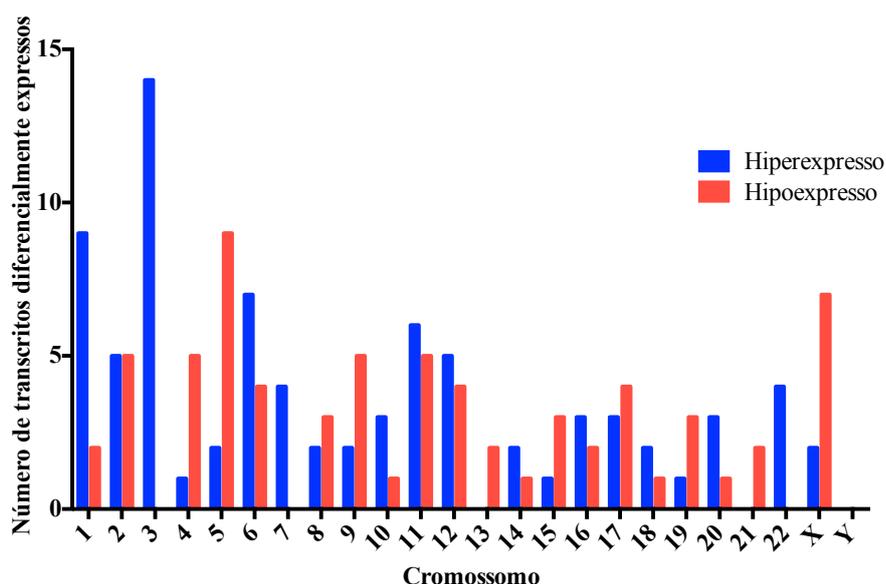


Figura 18 - Distribuição dos 150 transcritos diferencialmente expressos de acordo com o cromossomo de origem

Em seguida, verificamos a localização do gene de cada transcrito DE em relação à sua posição cromossômica (**Figura 19**). Observamos que as regiões a seguir possuem pelo menos dois genes diferencialmente expressos, podendo ser indicativo de um evento de amplificação ou deleção de parte do cromossomo: 1q32.1 (*LGR6*, *SOX13*), 3q23 (*PCOLCE2*, *PXYLP1*), 3q29 (*ATP13A5*, *HRASLS*, *LINC00909*, *FGF12*), 5p13.1 (*LIFR*, *OSMR*), 5q31.3 (*PCDHGA8*, *PCDHB3*, *PCDHB14*, *VTRNA1-3*), 6q14.1 (*TPBG*, *ELOVL4*), 6q25.3 (*CLDN20*, *RP11477D19.2*), 8q24.3 (*RP11-661A12.14*, *ARC*), 9p13.3 (*PRSS3*, *DNAJB5*), 11q13.3 (*AP000487.5*, *CCND1*, *ANO1*, *RP11-849H4.2*, *FADD*, *FOSL1*), 12q21.31 (*NTS*, *RASSF9*, *SLC6A15*, *RP11-1079J22.1*), 15q23 (*PAQR5*, *SPESP1*), 18p11.31 (*L3MBTL4*, *ZBTB14*), 19p13.12 (*ZNF850*, *CASP14*, *CYP4F22*), 22q13.1 (*MGAT3*, *SYNGR1*), Xq27.2 (*MAGEC1*, *MAGEC2*) e Xq28 (*CSAG1*, *MAGEA10*, *MAGEA12*).

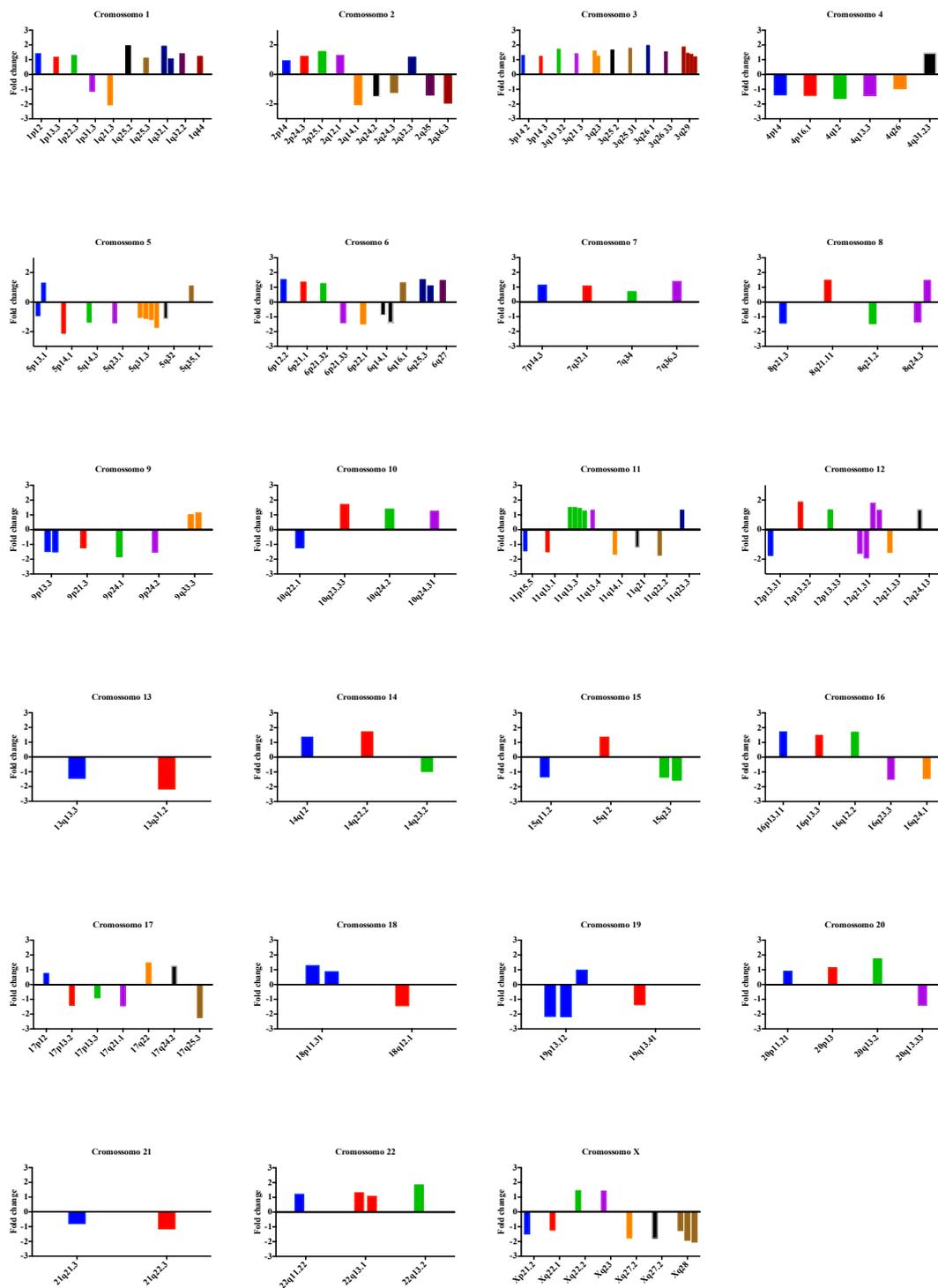


Figura 19 - Distribuição dos 150 genes diferencialmente expressos por região cromossômica de acordo com o *fold change*.

Assim, buscamos na literatura por análises que relacionassem alterações cromossômicas e a ocorrência de metástase nodal em CEB. SUGAHARA et al. (2011) verificaram que ganhos da região 11q13 estavam relacionados à tumores que apresentavam metástases nodais. Esta foi a região que apresentou o maior número de genes DE no nosso estudo, sendo 5 genes hiperexpressos no grupo N+ (*AP000487.5*, *CCND1*, *ANO1*, *RP11-849H4.2*, *FADD*) e apenas um hipoexpresso (*FOSL1*). Esses autores ainda verificaram que os transcritos *CCND1* e *FADD*, representam importantes marcadores prognósticos para metástase nodal. Vale destacar que o gene *CCND1* também se apresentou hiperexpresso em tumores metastáticos em outros dois estudos de análise transcricional (NAGATA et al. 2003; ROEPMAN et al. 2006), como será discutido adiante. DA SILVA et al. (2015) observaram ganhos cromossômicos na região 3q21.3 em CEB metastáticos. Em nossos achados, detectamos um gene hiperexpresso na mesma região cromossômica, o *KLF15*. UCHIDA et al. (2011) ao compararem 25 CEBs metastáticos com 25 não metastáticos observaram perdas cromossômicas nas regiões 11p15 e 20q13.33. Em nossos resultados observamos transcritos hipoexpressos em tumores metastáticos nas mesmas regiões: *CDHR5* e *MIR646HG*, respectivamente. Esses achados da literatura, de certa forma, dão embasamento para os nossos resultados já que, embora as metodologias comparadas sejam distintas, provêm informações que se complementam.

Para verificação de possíveis vias biológicas enriquecidas entre os genes diferencialmente expressos, submetemos os dados para análise na

ferramenta *online* KOBAS 2.0 (XIE et al. 2011). Os genes foram separados em hiperexpressos e hipoexpressos, sendo cada grupo analisado individualmente. Foram considerados significativos os resultados com p-valor corrigido $<0,05$ e com pelo menos 2 transcritos alocados na via encontrada. Apenas a via “Collagen formation” (REACT_120729) foi representada significativamente (p-valor corrigido = 0,01) contendo 4 dos genes DE (*COLGALT2*, *MMP13*, *PCOLCE2*, *LOXL4*) identificados neste trabalho, sendo todos hiperexpressos em tumores metastáticos.

O remodelamento da matriz extracelular têm sido reconhecido como um processo importante no desenvolvimento metastático. O aumento da expressão de genes relacionados a este processo foi relacionado a maior mortalidade em alguns tumores humanos (GILKES et al. 2014). De fato, a presença de um estroma denso associado ao tumor, denominado de reação desmoplásica, está relacionada à maior ocorrência de metástase linfonodal oculta em tumores da língua oral (PIMENTA AMARAL et al. 2004).

A formação dos colágenos fibrilares (I, II e III) é dividida em duas etapas, sendo uma intracelular e outra extracelular. Na etapa intracelular ocorre a formação dos polipetídeos de cadeia alfa, que serão processados no retículo endoplasmático (RE). No RE as cadeias alfa sofrerão modificações permitindo a sua interação em triplas-hélices, sendo então exportadas para o meio extracelular. No meio extracelular, proteinases irão clivar as regiões C e N terminais da tripla hélice, permitindo que estas interajam entre si formando fibrilas de colágeno. O processo final de formação do colágeno fibrilar é determinado a partir de enzimas da família

LOX que irão permitir a interação molecular entre essas fibrilas estabelecendo estruturas mais densas denominadas de fibras colágenas (GILKES et al. 2014).

Considerando os genes DE alocados na via “*Collagen formation*” (REACT_120729), observamos que 3 genes possuem papel fundamental em todas as etapas da formação da estrutura da fibra colágena.

O gene *COLGALT2*, ou *GLT25D2*, é uma beta galactosil-transferase que participa da glicosilação de resíduos de hidroxilisina nas cadeias alfa no RE (SCHEGG et al. 2009). Esse processo é essencial na biogênese do colágeno, permitindo uma estabilidade térmica da tripla hélice do colágeno. Cadeias alfa de pró-colágeno que não são adequadamente hidroxiladas estão mais propícias à degradação proteolítica e diminuição da deposição de colágeno. A interação entre fibras de colágeno hidroxiladas com resíduos de lisina têm mais estabilidade, o que leva à formação de um tecido mais rígido, podendo estar relacionado à uma facilitação da disseminação tumoral (GILKES et al. 2014).

Após a exportação das triplas hélices para o meio extracelular, os pró-peptídeos C e N terminais devem ser clivados para permitir que fibrilas de colágeno sejam formadas. Os pró-peptídeos da região C terminal são clivados pela ação da BMP-1 e por algumas metaloproteinases. Sabe-se que a atividade de BMP-1 pode ser aumentada pela ação de alguns catalizadores como a PCOLCE2, que age como um aditivo para a ocorrência da reação de quebra da região C terminal da tripla hélice (STEIGLITZ et al. 2002).

Após a formação das fibrilas de colágeno, a interação entre as mesmas é iniciada por uma família de enzimas denominada LOX. Estas enzimas estão diretamente relacionadas ao aumento da fibrose, sendo que o aumento de sua expressão relaciona-se ao remodelamento do colágeno tanto no sítio tumoral primário quanto no sítio metastático. Acredita-se que a indução da remodelação do colágeno em linfonodos por enzimas da família LOX facilitaria a colonização deste sítio por células tumorais. O aumento da expressão de enzimas LOX foi associada a ocorrência de metástase à distância em CECP (ERLER et al. 2006). Especificamente a LOXL4, encontrada diferencialmente expressa neste trabalho, foi associada ao desenvolvimento de metástases linfonodais em CECP, assim como sua expressão foi intensa em células tumorais metastáticas dos linfonodos cervicais (WEISE et al. 2008).

Dessa forma, podemos verificar que os genes DE que se alocaram neste via podem, de fato, ter papel de extrema importância na determinação do fenótipo metastático.

Além disso, o remodelamento da matriz extracelular não é só determinado por moléculas que promovem a formação de fibras colágenas, sendo que a degradação dessas fibras também é de extrema importância. A metaloproteinase 13 (MMP13) é uma metaloproteinase da família das collagenases que possui capacidade de degradação do colágeno fibrilar I, II e III, além de degradar também colágeno tipo IV. Dessa forma, a MMP13 possui um papel muito importante na remodelação da matriz extracelular interagindo diretamente com as vias descritas acima. A expressão de MMP13

em CEB têm sido relacionada à ocorrência de metástase nodal (VINCENT-CHONG et al. 2014). Mostrou-se que sua expressão é muito maior no tecido tumoral do que na mucosa normal do mesmo paciente, tendo uma expressão intermediária em lesões potencialmente malignas. Isso indica um papel importante da MMP13 na progressão da doença (CHIANG et al. 2006). É importante destacar que a expressão de MMP13 é significativamente maior no fronte de invasão tumoral do CEB (VINCENT-CHONG et al. 2014).

A inibição da expressão de MMP13 em células de CEB *in vitro*, possui consequências importantes na capacidade metastática dessas células. Além de causar inibição da sua capacidade invasiva (YU et al. 2011), observou-se também inibição da capacidade de implantação dessas células quando injetadas em um modelo animal, impedindo assim a consolidação do sítio metastático (LEIVONEN et al. 2006).

Para verificar a complementariedade da lista de genes DE gerada neste trabalho, buscamos nossos transcritos em listas de genes DE geradas em outros 9 trabalhos que também buscaram transcritos marcadores de metástase nodal em CEB (NAGATA et al. 2003; SCHMALBACH et al. 2004; O'DONNELL et al. 2005; ROEPMAN et al. 2005, 2006; ZHOU et al. 2006; NGUYEN et al. 2007; MÉNDEZ et al. 2007; KASHIWAZAKI et al. 2008). As comparações foram realizadas considerando separadamente os transcritos hiper e hipoexpressos. Apesar de peculiaridades metodológicas e aspectos distintos de separação dos grupos amostrais, dentre os 9 trabalhos que identificamos, apenas 4 apresentaram genes DE encontrados também no

nosso estudo, sendo estes representados por apenas 4 transcritos (**Quadro 12**).

Quadro 12 - Genes DE deste estudo também observados em outros trabalhos

Interseção	Total de genes	Gene	Expressão em CEB metastáticos
Genes DE + Nagata et al. (2003) + Roepman et al. (2006)	1	<i>CCND1</i>	Aumentada
Genes DE + Nagata et al. (2003)	1	<i>MMP13</i>	Aumentada
Genes DE + Kashiwazaki et al. (2008)	1	<i>CYP26A1</i>	Aumentada
Genes DE + O'Donnell et al. (2005)	1	<i>GABRB3</i>	Aumentada

Essa grande variabilidade de transcritos entre diferentes estudos pode ser explicada a partir dos resultados de ROEPMAN et al. (2006). Os autores verificaram que a partir de um grande número de genes diferencialmente expressos, diferentes combinações de assinaturas gênicas relacionadas à metástase podem ser geradas, obtendo-se sempre alta acurácia para cada combinação na predição do desfecho. De acordo com o autores, não há um único grupo de genes que irá portar uma acurácia preditiva perfeita. Dessa forma, embora haja baixa complementariedade de transcritos DE os trabalhos, isso não invalida nossos achados nem diminui a importância dos mesmos para diagnóstico da metástase nodal. Vale ressaltar que este é o primeiro estudo a utilizar uma tecnologia de sequenciamento em larga escala para determinar marcadores de metástase nodal em CEB. Todos os demais trabalhos citados nesta análise utilizaram *microarrays* que são tecnologias de resolução limitada.

Outro fator importante é que nosso estudo é o único que comparou tumores com comportamento metastático absolutamente polar. Incluímos

tumores que desenvolveram metástases nodais em estágio de desenvolvimento ainda precoce e comparamos com tumores de longo desenvolvimento (em decorrência do tamanho grande) de pacientes que foram acompanhados por pelo menos 3 anos, garantindo assim o caráter não metastático dos mesmos. Essa seleção cuidadosa dos casos ressalta a importância dos dados gerados neste estudo.

Finalmente, para confirmar a efetividade do grupo de transcritos gerados neste trabalho como DE, realizamos uma Análise de Componentes Principais (PCA) (**Figura 20**) para verificar a distribuição das amostras de acordo com o padrão de expressão desses genes. Nesta análise, verificamos uma adequada separação de todas as 20 amostras sequenciadas.

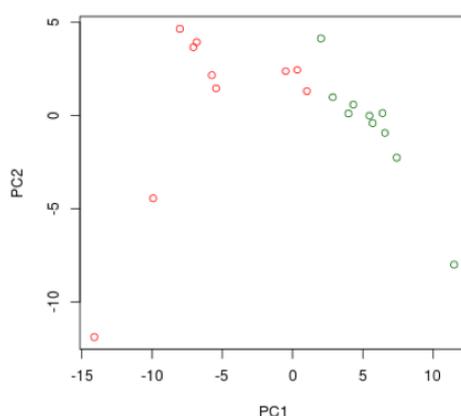


Figura 20 - PCA utilizando o padrão de expressão dos 150 transcritos DE. A figura mostra o potencial dos 150 genes DE na separação de tumores metastáticos (círculos vermelhos) de tumores não metastáticos (círculos verdes)

- **Resultados da validação *in silico***

Embora o número de trabalhos utilizando a tecnologia de RNASeq tenha crescido nos últimos anos, não existem diretrizes bem definidas para determinação do número de fragmentos necessários para se cobrir adequadamente os transcriptomas dos diferentes tecidos humanos. Isso se deve ao fato de que esse critério é dependente das características biológicas que estão sendo investigadas, como tipo de tecido ou seu grau de desenvolvimento, por exemplo. Dessa forma, mais importante que o número de fragmentos sequenciados e a cobertura do transcriptoma de um determinado tecido seria o número de réplicas biológicas em um determinado experimento (SIMS et al. 2014). Tendo isso em vista, nossa abordagem metodológica incluiu um grupo de amostras para validação *in silico* que aumenta significativamente o universo de tumores analisados dando suporte matemático importante aos nossos achados.

Utilizamos nesta validação com grupo amostral independente, os dados de sequenciamento gerados pelo *The Cancer Genome Atlas* (TCGA). Para isto, foram realizadas 5 seleções de pacientes utilizando diferentes critérios de inclusão. A descrição destes critérios, assim como o número de pacientes selecionados em cada grupo estão descritos na **Quadro 13**.

Quadro 13 - Critérios de seleção para formação dos grupos de validação *in silico*

Grupo TCGA	Tamanho grupo pN+				Tamanho grupo pN0				Sítio			Tabagismo		Número de pacientes	
	pT1	pT2	pT3	pT4	pT1	pT2	pT3	pT4	C02	C04	C02-C06	Sim	Não	N0	N+
TCGA_1	✓	-	-	-	-	-	✓	✓	✓	✓	-	✓	-	25	3
TCGA_2	✓	-	-	-	-	-	✓	✓	✓	✓	-	✓	✓	35	6
TCGA_3	✓	✓	-	-	-	-	✓	✓	✓	✓	-	✓	✓	35	31
TCGA _{n1}	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	-	✓	✓	43	66
TCGA _{n2}	✓	✓	✓	✓	✓	✓	✓	✓	-	-	✓	✓	✓	77	105

Inicialmente, os genes DE obtidos a partir dos dados do TCGA dos grupos de tumores pequenos (TCGA_1, TCGA_2 e TCGA_3) foram comparados aos genes DE obtidos a partir dos dados gerados neste estudo. Não foi possível identificar nenhum gene DE em comum entre a seleção TCGA_1 e TCGA_2 e os nosso casos. Já ao se comparar com a seleção TCGA_3, encontramos 17 genes diferencialmente expressos em ambas análises (**Figura 21**).

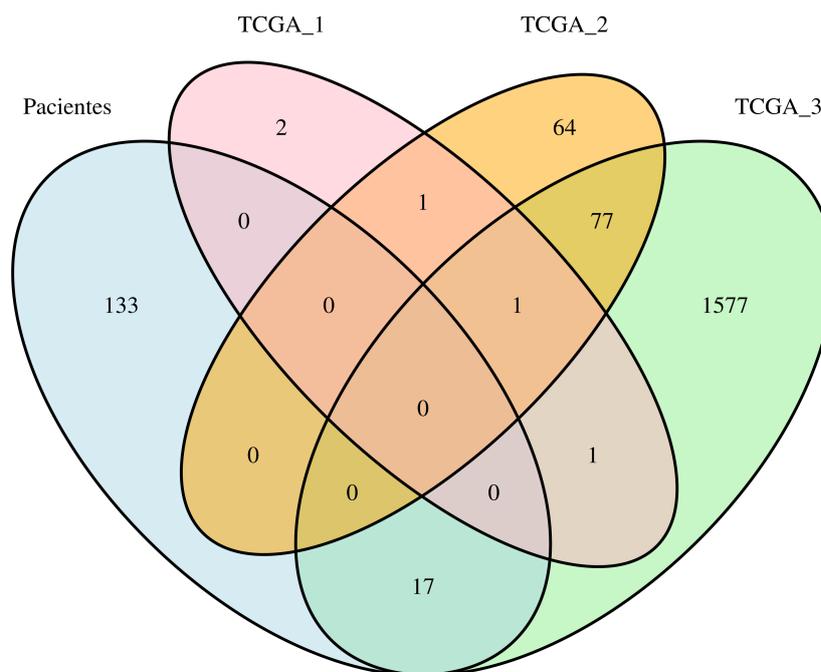


Figura 21 - Distribuição de genes DE deste estudo em relação às seleções TCGA_1, TCGA_2 e TCGA_3. O Diagrama de Venn representa genes diferencialmente expressos encontrados em cada seleção de casos e suas interseções.

Entretanto, 3 (*MMP13*, *PTPRN*, *SOX13*) dos 17 genes apresentaram sentido da alteração de expressão inversa (p.ex., hiperexpressão de um gene em tumores metastático nos dados do TCGA e hipoexpressão em tumores metastáticos nos dados gerados neste estudo, ou vice-versa). Dessa forma, apenas 14 genes foram confirmados como diferencialmente expressos entre tumores metastáticos e não metastáticos nesta seleção de casos do TCGA. Esses genes 14 genes validados, assim como as respectivas variações de expressão (*fold-change*) entre os grupos de estudo, estão descritas na **Tabela 1**.

Tabela 1 - Genes diferencialmente expressos validados na seleção TCGA_3

	Genes	Pacientes (T1N+ vs. T3-4N0)				TCGA_3 (T1-2N+ vs. T3-4N0)			
		logFC	FC	p	FDR	logFC	FC	p	FDR
ENSG00000171954	<i>CYP4F22</i>	-2,170	0,22	3,8E-08	2,7E-04	-0,96	0,52	>0,001	0,06
ENSG00000151623	<i>NR3C2</i>	1,433	2,70	2,1E-06	3,0E-03	0,94	1,92	>0,001	0,03
ENSG00000124260	<i>MAGEA10</i>	-1,910	0,27	6,4E-06	5,8E-03	-0,78	0,58	0,01	0,10
ENSG00000178445	<i>GLDC</i>	-1,834	0,28	6,5E-06	5,8E-03	-1,91	0,27	>0,001	>0,001
ENSG00000126016	<i>AMOT</i>	1,415	2,67	4,1E-05	1,6E-02	1,23	2,34	>0,001	0,01
ENSG00000198848	<i>CES1</i>	1,688	3,22	5,5E-05	2,0E-02	0,85	1,81	>0,001	0,06
ENSG00000100027	<i>YPEL1</i>	1,200	2,30	7,1E-05	2,5E-02	0,56	1,48	>0,001	0,05
ENSG00000171217	<i>CLDN20</i>	1,519	2,87	1,4E-04	4,0E-02	1,17	2,24	>0,001	0,03
ENSG00000197181	<i>PIWIL2</i>	-1,416	0,37	1,9E-04	4,5E-02	-1,30	0,41	>0,001	0,01
ENSG00000086696	<i>HSD17B2</i>	-1,495	0,35	1,9E-04	4,5E-02	-1,39	0,38	>0,001	>0,001
ENSG00000168477	<i>TNXB</i>	1,240	2,36	3,5E-04	6,3E-02	0,90	1,86	>0,001	0,06
ENSG00000137726	<i>FXYP6</i>	1,312	2,48	3,9E-04	6,8E-02	1,05	2,07	>0,001	0,02
ENSG00000010438	<i>PRSS3</i>	-1,474	0,36	4,9E-04	8,0E-02	-0,95	0,52	>0,001	0,05
ENSG00000198756	<i>COLGALT2</i>	1,090	2,13	7,1E-04	9,3E-02	1,08	2,12	>0,001	0,03

Considerando a natureza da comparação que realizamos (tumores pequenos metastáticos vs. tumores grandes não metastáticos) inserimos um viés em nossa análise relacionado ao tamanho tumoral. Se por um lado esta comparação permitiria identificar genes indicativos de comportamento mais agressivo já expressos nos tumores pequenos, ela também poderia representar um viés de expressão relacionado ao tamanho tumoral e não ao seu potencial metastático. A fim de verificar se a diferença de expressão observada estava relacionada exclusivamente ao fenótipo metastático e não à diferença de tamanho entre os grupos de tumores, realizamos uma comparação com outros dois grupos de casos selecionados do TCGA (TCGAn1 e TCGAn2). Nesses dois grupos eliminamos o fator tamanho como um critério de seleção. Além disso, testamos se o grupo de genes DE

determinados a partir de tumores de língua e assoalho bucal também seria validado em uma seleção de tumores que incluísse todos os sítios da mucosa oral (TCGAn2) (**Quadro 13**).

A partir dessa análise, obtivemos uma lista de 6 genes diferencialmente expressos em todos os quatro grupos de tumores analisados. Entretanto, com o intuito de eliminar possíveis variações biológicas decorrentes de tumores localizados em diferente sítios da cavidade oral, decidimos prosseguir o estudo considerando apenas os 7 genes DE no nosso estudo e na seleção TCGAn1 (**Figura 22**). Os genes DE assim como as respectivas variações na expressão de cada gene estão descritas na **Tabela 2**.

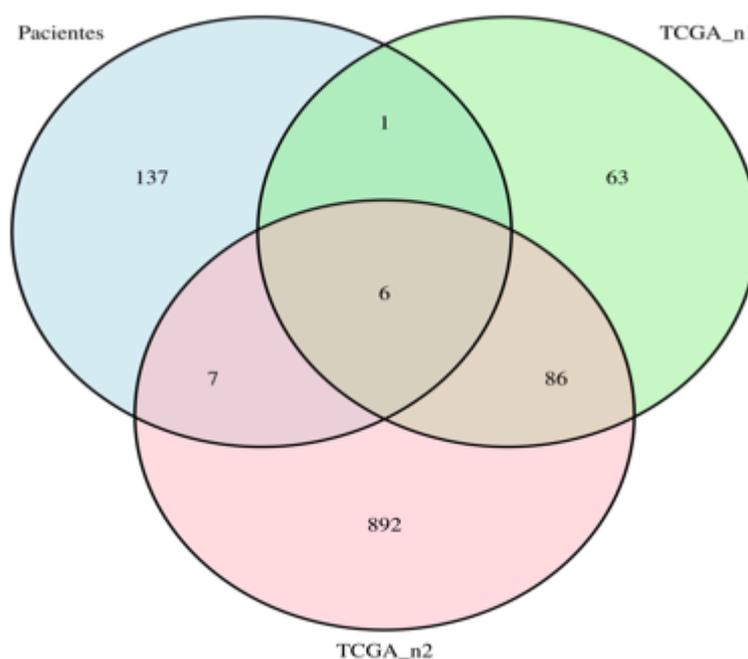


Figura 22 - Distribuição de genes DE deste estudo em relação às seleções TCGAn1 e TCGAn2. O Diagrama de Venn representa genes diferencialmente expressos encontrados em cada seleção de casos e suas interseções.

Tabela 2 - Genes diferencialmente expressos validados na seleção TCGAn1

	Genes	Pacientes (T1N+ vs. T3-4N0)				TCGAn1 (língua e assoalho bucal)			
		logFC	FC	p	FDR	logFC	FC	p	FDR
ENSG00000151623	<i>NR3C2</i>	1,433	2,701	<0,001	0,003	0,774	1,710	0,001	0,089
ENSG00000178445	<i>GLDC</i>	-1,834	0,281	<0,001	0,006	-1,472	0,361	<0,001	<0,001
ENSG00000198848	<i>CES1</i>	1,688	3,221	<0,001	0,020	1,402	2,642	<0,001	<0,001
ENSG00000010438	<i>PRSS3</i>	-1,474	0,360	<0,001	0,080	-0,840	0,559	0,001	0,087
ENSG00000198756	<i>COLGALT2</i>	1,090	2,129	0,001	0,093	0,969	1,957	<0,001	0,031
ENSG00000110092	<i>CCND1</i>	1,494	2,816	<0,001	0,006	0,697	1,621	<0,001	0,067
ENSG00000168040	<i>FADD</i>	1,249	2,377	<0,001	0,030	0,750	1,682	<0,001	0,063

As diferenças de expressão dos respectivos genes entre tumores metastáticos e não metastáticos nos pacientes deste estudo e no grupo TCGAn1 estão descritas em mais detalhes na **Figura 23**. Observamos uma clara diferença de expressão destes genes entre os tumores analisados.

Em seguida, realizamos nova análise de PCA considerando as amostras sequenciadas separando-as de acordo com o padrão de expressão dos 7 genes validados *in silico* (**Figura 24**). Mais uma vez verificamos que o padrão de expressão desses genes é adequado para separar tumores metastáticos e não metastáticos.

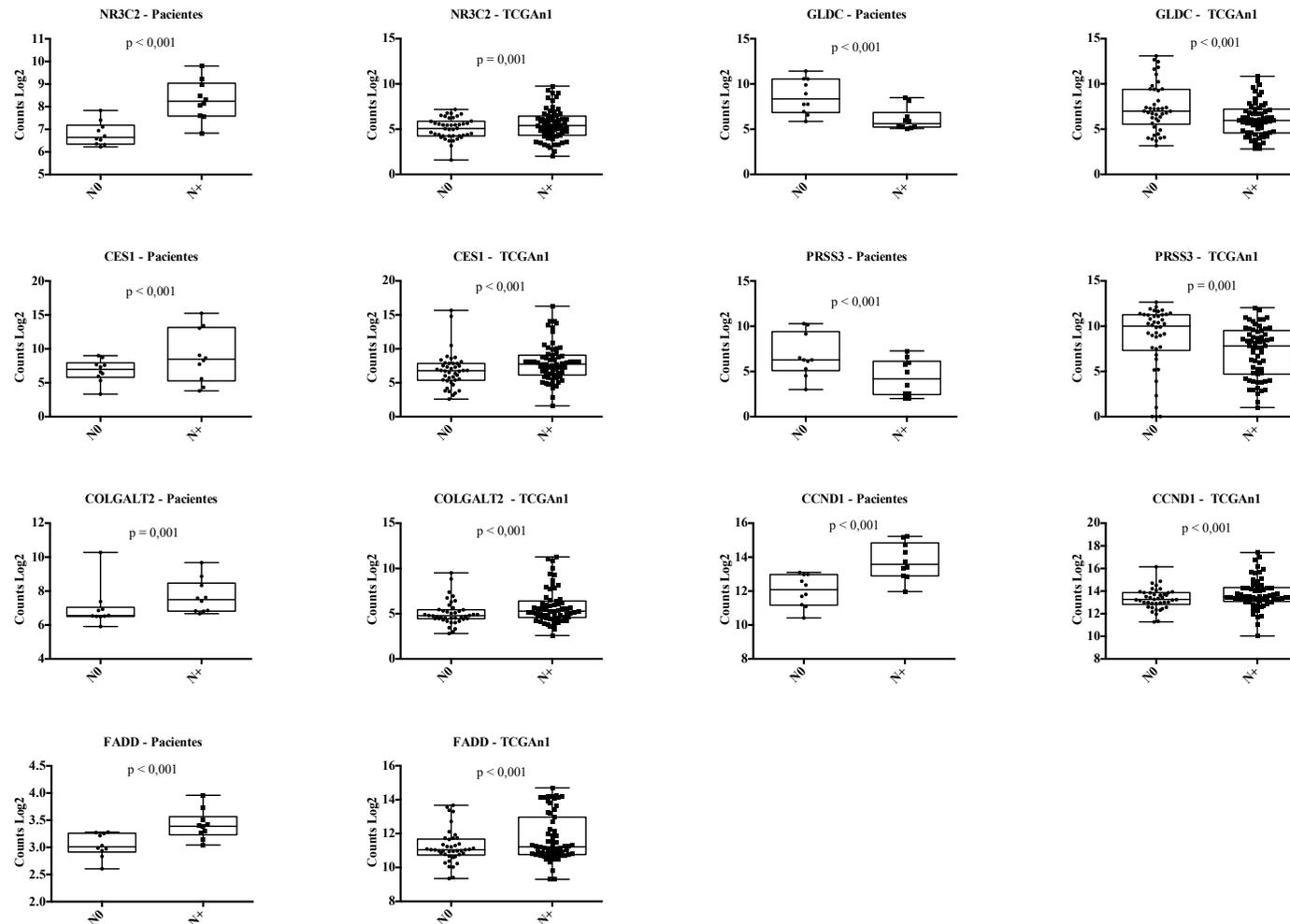


Figura 23 - *Boxplots* referentes aos genes validados na seleção TCGAn1. Valores de expressão em Log2 representando os níveis de expressão dos casos deste estudo e dos casos da seleção TCGAn1

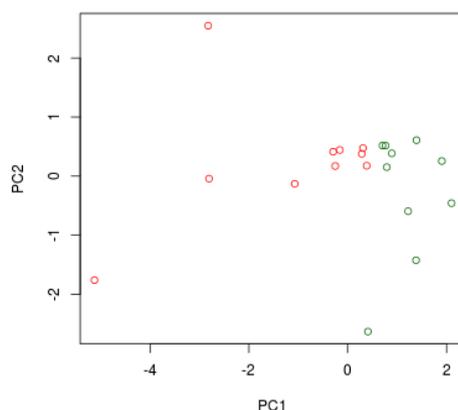


Figura 24 - PCA utilizando o padrão de expressão dos 7 genes validados *in silico*. A figura demonstra o potencial dos 7 genes em separar os tumores metastáticos (círculos vermelhos) de tumores não metastáticos (círculos verdes).

4.4 CURVA ROC DOS GENES CANDIDATOS

Considerando que o principal objetivo deste estudo foi determinar possíveis biomarcadores de metástase nodal, utilizamos a ferramenta estatística curva ROC, destinada a determinar quantitativamente o desempenho de um teste diagnóstico. A curva ROC se baseia no cálculo da especificidade e sensibilidade de um teste diagnóstico. A sensibilidade é a probabilidade de um teste em fornecer resultado positivo para um indivíduo doente. Já a especificidade é a medida de probabilidade do teste diagnóstico em fornecer o resultado negativo para um indivíduo não doente. A partir do cálculo da especificidade e sensibilidade de um determinado teste calcula-se uma medida do seu desempenho denominada de “área sob a curva” ou AUC. A AUC determina a probabilidade do indivíduo portador da doença ter um resultado de maior magnitude que um não doente. Valores de AUC iguais a 0,5 indicam que o teste não tem capacidade de discriminar entre

doentes e não doentes. Quanto mais o valor de AUC se aproximar de 1,0 maior a capacidade do teste diagnóstico em separar doentes de não doentes (MARTINEZ et al. 2003).

Dessa forma, verificamos através dessa ferramenta a capacidade dos 7 diferentes transcritos validados pelos dados do TCGA em discriminar pacientes com e sem metástase nodal. Os resultados obtidos nesta análise estão descritos na **Figura 25**. Considerando que os testes com valor de AUC que mais se aproximam de 1,0 são aqueles com maior potencial de diagnóstico da doença ou desfecho de interesse, concluímos que os transcritos NR3C2, FADD e CCND1 possuem, na análise univariada, maior potencial para se tornarem biomarcadores de metástase nodal.

Em seguida, verificamos, através de análise multivariada, os valores de AUC gerados por diferentes combinações dos 7 transcritos. Todas as possíveis combinações de transcritos foram testadas, mas apenas aquelas combinações com maior AUC estão expostas na **Tabela 3**.

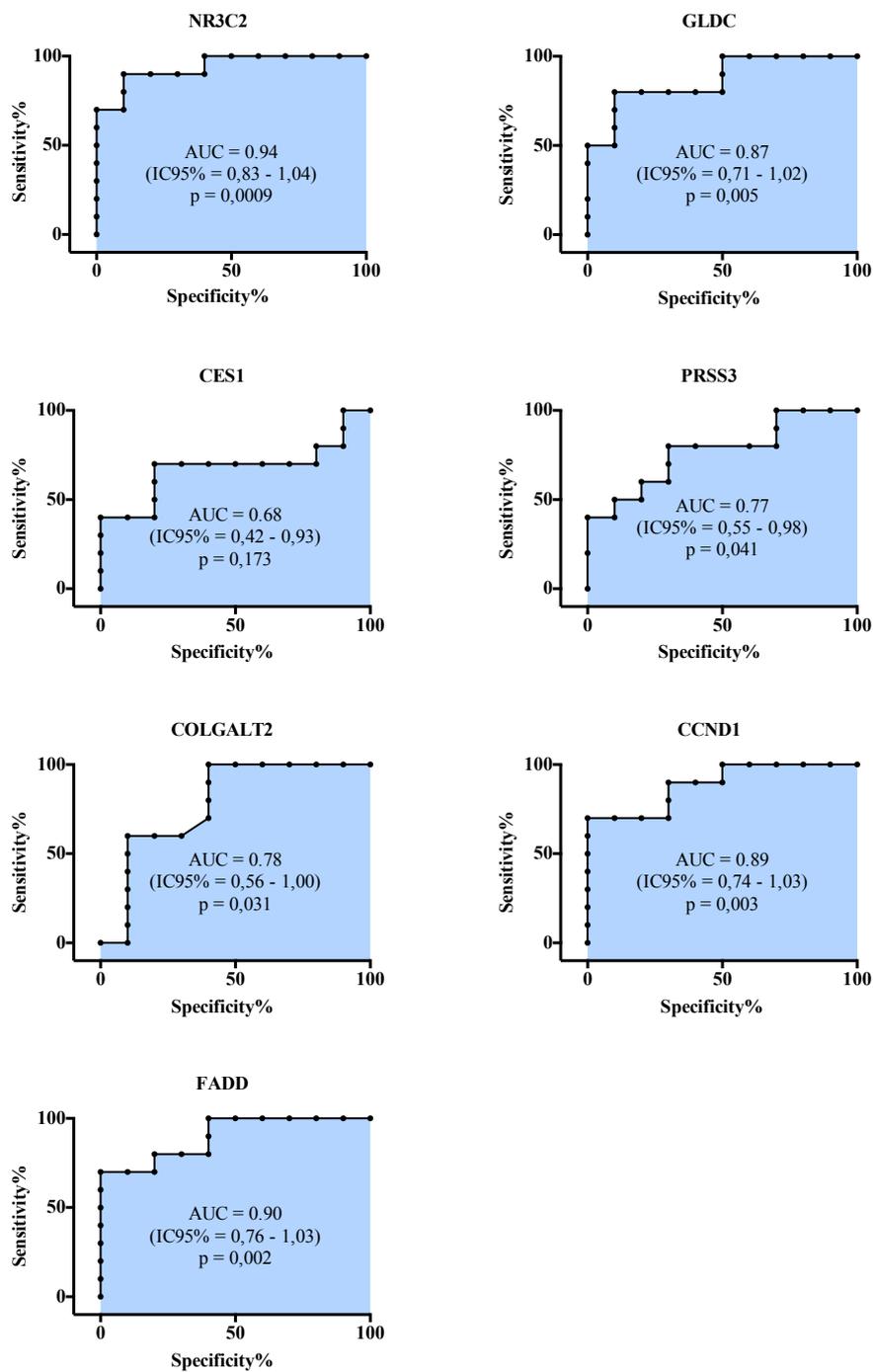


Figura 25 - Curvas ROC utilizando padrão de expressão dos 7 genes validados *in silico*. Dados expressão referem-se ao casos sequenciados neste estudo.

Tabela 3 - Análise ROC multivariada baseada nos 7 transcritos validados *in silico*

Combinação de transcritos	AUC	IC 95%
<i>NR3C2 + FADD + CCND1 + GLDC + COLGALT2 + PRSS3 + CES1</i>	0,93	0,55 - 1,00
<i>NR3C2 + FADD + CCND1 + GLDC + COLGALT2 + PRSS3</i>	0,93	0,66 - 1,00
<i>NR3C2 + FADD + CCND1 + GLDC + COLGALT2</i>	0,92	0,60 - 1,00
<i>NR3C2 + FADD + CCND1 + GLDC</i>	0,96	0,83 - 1,00
<i>NR3C2 + FADD + CCND1</i>	0,97	0,83 - 1,00
<i>NR3C2 + GLDC</i>	0,94	0,66 - 1,00
<i>NR3C2 + FADD</i>	0,95	0,77 - 1,00
<i>NR3C2 + CCND1</i>	0,98	0,88 - 1,00
<i>FADD + GLDC</i>	0,93	0,66 - 1,00
<i>FADD + CCND1</i>	0,90	0,66 - 1,00

Considerando que um teste ideal envolveria o menor número possível de biomarcadores para se testar, selecionamos aqueles transcritos que foram mais frequentemente relacionados a maiores valores de AUC nas análises multivariadas para serem avaliados nas etapas seguintes de validação. Dessa forma, temos como melhores candidatos a biomarcadores os transcritos *NR3C2*, *CCND1* e *FADD*.

Com o intuito de demonstrar o potencial de separação dos 3 genes com maior valor de AUC, realizamos outra análise de PCA (**Figura 26**). Verificamos que mesmo diminuindo significativamente o número de transcritos avaliados, ainda temos uma separação adequada das amostras considerando o status linfonodal de cada uma.

esta fase do estudo, além de limitações técnicas da abordagem de validação (qRT-PCR) que é menos adequada para capturar menores variações de expressão que são prontamente identificadas por abordagens mais robustas tais como o sequenciamento em larga escala do transcriptoma.

Quadro 14 - Amostras utilizadas na validação técnica

Amostra	Idade ao diagnóstico	Gênero	Localização	pT	pN	M	Estadiamento	Grupo	Origem da amostra
CP1/0151	47	M	C02	3	0	0	III	N0	GENCAPO
CP2/1022	44	M	C02	3	0	0	III	N0	GENCAPO
CP1/0291	55	M	C02	4	0	0	IV	N0	GENCAPO
CP3/0486	56	M	C02	3	0	X	III	N0	GENCAPO
CP1/0040	48	M	C02	4	0	0	IV	N0	GENCAPO
CP3/0652	63	M	C04	1	1	X	III	N+	GENCAPO
CP3/0012	50	M	C04	1	1	0	III	N+	GENCAPO
CP3/0280	52	M	C04	1	2b	X	IV	N+	GENCAPO
CO30T	40	M	C04	1	2b	X	IV	N+	ACC
CP3/0613	56	M	C02	1	2b	X	IV	N+	GENCAPO
CP1/0397	59	M	C04	1	3	0	IV	N+	GENCAPO

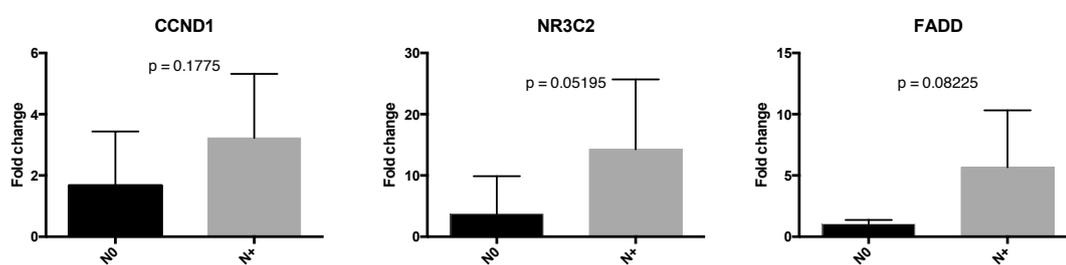


Figura 27 - Padrão de expressão dos genes com maior AUC nos grupos N+ e N0.

Com o intuito de verificar se o padrão de expressão verificado na etapa de validação técnica condiz com os resultados gerados durante o sequenciamento, realizamos análises de correlação de *Spearman*, onde observamos uma correlação significativa entre os dados (**Figura 28**). Para estas análises foram utilizados dados de delta Ct e “counts” normalizados obtidos no sequenciamento para cada um dos genes de interesse.

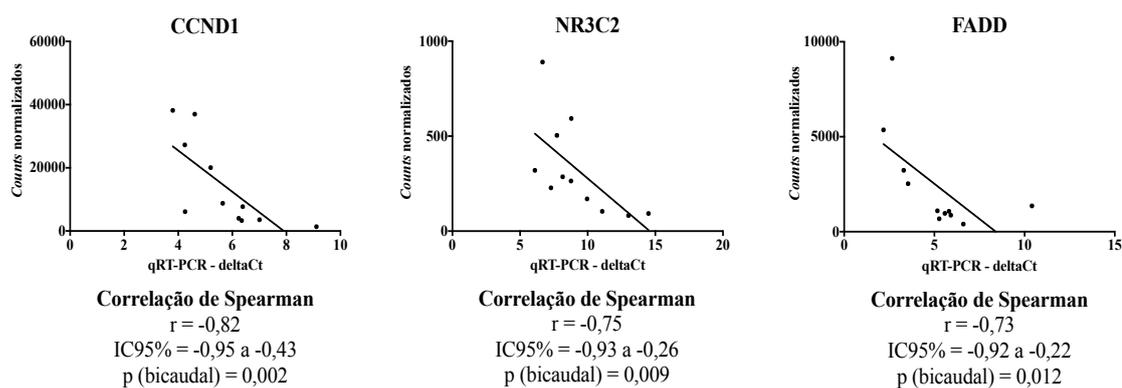


Figura 28 - Correlação entre valores de expressão gênica obtidos no sequenciamento e no qRT-PCR.

A partir dos dados gerados nesta análise é possível dizer que há uma tendência de comprovação dos resultados gerados no sequenciamento, embora a inclusão do restante das amostras seja importante para esta etapa do trabalho.

4.6 VALIDAÇÃO BIOLÓGICA

A validação dos nossos resultados em um grupo amostral independente é de fundamental importância para confirmação destes transcritos como potenciais biomarcadores de metástase nodal em CEB.

Estes biomarcadores teriam impacto clínico significativo, principalmente, no grupo de tumores cT1 e cT2 que não apresentam evidências de metástase linfonodal ao exame clínico (cN0). O tratamento deste grupo de pacientes ainda é contraditório, já que não é possível saber se estes indivíduos já apresentam metástases nodais subclínicas ou se irão apresentar no futuro. A decisão de se intervir cirurgicamente pode favorecer o diagnóstico precoce das metástases nodais. Entretanto, aqueles pacientes que não apresentam doença metastática nodal vão estar sendo submetidos a um tratamento desnecessário que causa significativa morbidade.

A intenção inicial deste trabalho foi de realizar uma validação biológica neste subgrupo de pacientes. Considerando a disponibilidade de amostras parafinadas de um grupo de pacientes com tais características, iniciamos uma abordagem de validação baseada na extração de RNA de tais tumores. Embora tenhamos tido sucesso na extração do RNA, os ensaios de qRT-PCR ainda não geraram resultados satisfatórios para exposição neste documento. Testes estão sendo feitos para padronização de tais reações, permitindo que no futuro possamos determinar a validade de tais marcadores no subgrupo clínico que mais se beneficiaria dos resultados deste trabalho.

Dessa forma, iniciamos uma nova abordagem de validação biológica utilizando amostras congeladas disponíveis. Para esta etapa selecionamos amostras de CEB de língua oral (CO2) e de assoalho bucal (C04) com características semelhantes às das amostras utilizadas na fase de descoberta de marcadores. A descrição das características clínicas de tais amostras está exposta na **Quadro 15**. Mais uma vez, separamos amostras com e sem metástase linfonodal confirmada por exame anatomopatológico. Foram utilizadas 18 amostras nesta fase do estudo, sendo que 7 amostras pertenciam ao grupo N0 e 11 ao grupo N+.

As reações de qRT-PCR foram realizadas utilizando o mesmo protocolo descrito anteriormente para a fase de validação técnica. Os resultados referentes ao padrão de expressão gênica dos 3 transcritos candidatos estão descritos na **Figura 29**.

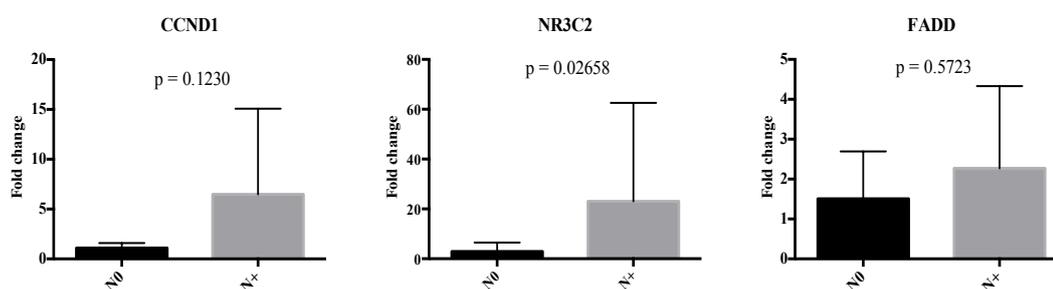


Figura 29 - Padrão de expressão dos transcritos candidatos em uma amostra independente de CEB.

Quadro 15 - Características clínicas das amostras utilizadas na validação biológica

Amostra	Idade ao diagnóstico	Gênero	Localização	pT	pN	M	Estadiamento	Grupo	Origem da amostra
CP3/0033	51	M	C04	3	0	0	III	N0	GENCAPO
CP2/0109	63	M	C04	3	0	0	III	N0	GENCAPO
CP1/0418	64	M	C02	4	0	0	IV	N0	GENCAPO
CP1/0340	51	M	C04	4	0	0	IV	N0	GENCAPO
CO95T	70	M	C02	1	0	0	I	N0	ACC
4566	53	F	C02	2	0	0	II	N0	ACC
4542	52	M	C04	3	0	0	III	N0	ACC
CP3/0605T	66	M	C02	1	2b	X	IV	N+	GENCAPO
CP2/0195	47	M	C04	1	1	0	III	N+	GENCAPO
CP1/0258	46	M	C04	1	2b	0	IV	N+	GENCAPO
13T	59	M	C02	1	1	0	III	N+	ACC
17188	66	M	C02	1	1	0	III	N+	ACC
CP3/0677	58	M	C04	1	2b	0	IV	N+	GENCAPO
CP3/0441	54	M	C02	1	1	0	III	N+	GENCAPO
CP2/1642	80	M	C02	1	1	0	III	N+	GENCAPO
CP2/1381	70	M	C04	1	2b	0	IV	N+	GENCAPO
CP2/1231	56	M	C04	1	2b	0	IV	N+	GENCAPO
CP1/0374	56	M	C04	1	2b	0	IV	N+	GENCAPO

Os resultados da análise de expressão em cada grupo (FC) foi expresso na forma de média \pm desvio padrão. Os resultados observados para cada grupo foram os seguintes:

- *CCND1*: N+ = 6,45 \pm 8,60 / N0 = 1,09 \pm 0,49;
- *NR3C2*: N+ = 23,01 \pm 39,59 / N0 = 2,88 \pm 3,59;
- *FADD*: N+ = 2,26 \pm 2,06 / N0 = 1,50 \pm 1,88.

Dentre os 3 transcritos, apenas o *NR3C2* se mostrou diferencialmente expresso entre os grupos de estudo. A direção da diferença de expressão entre os grupos foi a mesma observada no sequenciamento e na validação biológica, sendo maior a presença desse transcrito entre tumores metastáticos. Embora não tenhamos observado diferença significativa entre os grupos para a expressão de *CCND1*, houve uma clara tendência de maior

presença desse transcrito nos tumores metastáticos, assim como verificado nos experimentos anteriores.

Em seguida, verificamos a capacidade dos 3 transcritos em identificar corretamente os tumores de acordo com os respectivos grupos através de análise de curva ROC. Os resultados observados estão expostos na **Figura 30**.

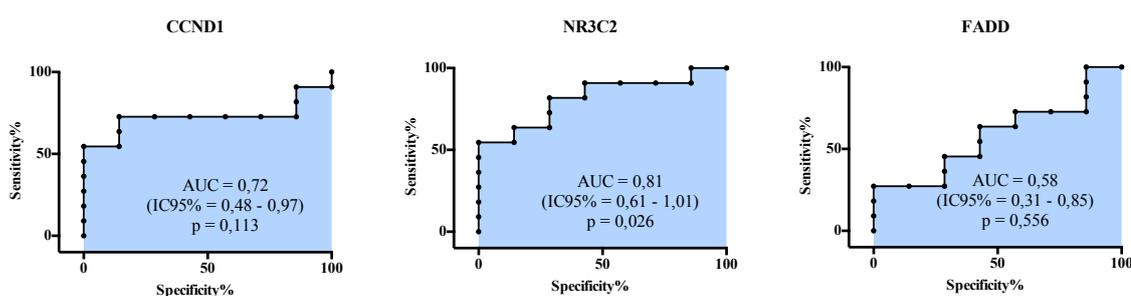


Figura 30 - Curvas ROC utilizando o padrão de expressão dos 3 genes candidatos em amostras independentes.

A partir desses resultados observamos que apenas o transcrito NR3C2 mostrou valores de AUC estatisticamente significativos e promissores para determinação do status metastático nas amostras investigadas. Os transcritos *CCND1* e *FADD* não apresentaram performance adequada como biomarcadores para este desfecho de acordo com as curvas ROC geradas.

A **Tabela 4** mostra os diferentes valores de sensibilidade e especificidade do teste na classificação dos CEB analisados de acordo com diferentes pontos de corte referentes à expressão (*fold change*) do transcrito NR3C2. Observa-se que a melhor performance deste marcador ocorre no

ponto de corte de 2,825, onde observamos uma sensibilidade de 81,82% e especificidade de 71,43%.

Em seguida, testamos o efeito da combinação de diferentes transcritos na curva ROC, onde todas as combinações possíveis de marcadores foram desafiadas. Como observa-se na **Tabela 5**, nenhuma dessas combinações mostrou melhores valores de AUC do que os observados nas análises univariadas. Assim, concluímos que análise do transcrito *NR3C2*, unicamente, possui a maior eficácia para determinação do status linfonodal nas amostras testadas nesta fase do estudo.

Tabela 4 - Valores de sensibilidade e especificidade do teste utilizando o transcrito *NR3C2* de acordo com diferentes pontos de corte de expressão gênica medida em *fold change*

Ponto de corte (Fold change)	Sensibilidade (%)	Intervalo de confiança (95%)	Especificidade (%)	Intervalo de confiança (95%)
0,07225	100	71,51 - 100	14,29	0,36 - 57,87
0,1929	90,91	58,72 - 99,77	14,29	0,36 - 57,87
0,4913	90,91	58,72 - 99,77	28,57	3,66 - 70,96
0,8418	90,91	58,72 - 99,77	42,86	9,89 - 81,59
1,338	90,91	58,72 - 99,77	57,14	18,41 - 90,10
1,963	81,82	48,22 - 97,72	57,14	18,41 - 90,10
2,825	81,82	48,22 - 97,72	71,43	29,04 - 96,33
3,709	72,73	39,03 - 93,98	71,43	29,04 - 96,33
5,445	63,64	30,79 - 89,07	71,43	29,04 - 96,33
7,926	63,64	30,79 - 89,07	85,71	42,13 - 99,64
9,000	54,55	23,38 - 83,25	85,71	42,13 - 99,64
9,161	54,55	23,38 - 83,25	100	59,04 - 100

Tabela 5 - Análise ROC multivariada baseada nos 3 transcritos candidatos nas amostras de validação biológica

Combinação de transcritos	AUC	IC 95%
<i>FADD + CCND1 + NR3C2</i>	0,65	0,12 - 1,00
<i>NR3C2 + CCND1</i>	0,60	0,00 - 1,00
<i>NR3C2 + FADD</i>	0,45	0,00 - 0,83
<i>FADD + CCND1</i>	0,43	0,08 - 0,83

Considerando este importante achado referente à relação entre a expressão de *NR3C2* no tumor e status dos linfonodos cervicais, decidimos verificar se o padrão de transcrição desse gene possuía alguma relação com a sobrevida de pacientes com CEB. Para isso, utilizamos dados clínicos obtidos da seleção de pacientes TCGAn1, composta por 105 pacientes com dados de acompanhamento disponíveis. Para esta análise, consideramos o tempo em meses, contando a partir do diagnóstico histopatológico até o último segmento, sendo o óbito o desfecho de interesse. As análises de sobrevida foram realizadas no *Software Prism 6 for Mac OS X Version 6.0f*, utilizando o teste de *Log-rank (Mantel-Cox)* para determinação da diferença entre as curvas de sobrevida, assim como para a determinação do risco relativo (HR). Inicialmente, geramos duas curvas de sobrevida, sendo uma separando os pacientes de acordo com o status linfonodal (determinado a partir do pN) e outra a partir da expressão do transcrito *NR3C2*. Para dicotomização dos pacientes de acordo com a expressão do *NR3C2*, determinou-se como ponto de corte o valor referente à média de expressão do grupo N0. Casos com valores iguais ou inferiores à média foram classificados como negativos (*NR3C2-*) e valores acima da média foram

classificados como positivos (NR3C2+). As curvas de sobrevida referentes a esta análise estão descritas na **Figura 31**. Não observamos diferença significativa entre as curvas de sobrevida de pacientes dos grupos N+ (sobrevida média de 36,4 meses) e N0 (sobrevida média de 56,6 meses) ($p = 0,25$). Entretanto, quando os pacientes foram categorizados de acordo com o padrão de expressão do NR3C2, observamos uma tendência de melhor separação das curvas de sobrevida, embora a diferença entre as curvas do grupo NR3C2+ (sobrevida média de 30,4 meses) e NR3C2- (sobrevida média de 54,6 meses) não tenha sido significativa ($p=0,09$). Dessa forma, podemos dizer que o padrão de expressão de NR3C2 tende a ser um fator prognóstico importante.

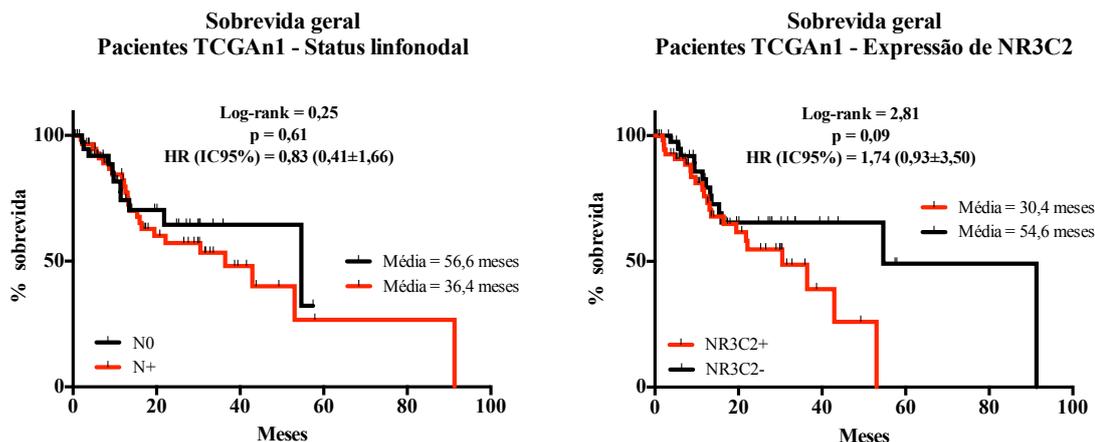


Figura 31 - Curvas de Kaplan-Meier geradas a partir da categorização dos pacientes do TCGAn1 de acordo com o status linfonodal e a expressão de NR3C2.

Adicionalmente, realizamos um novo agrupamento considerando o status linfonodal e a expressão de *NR3C2*, concomitantemente. Dessa forma, geramos quatro grupos: pacientes com metástase nodal e alta expressão de *NR3C2* (N+ NR3C2+), pacientes com metástase nodal e baixa expressão de *NR3C2* (N+ NR3C2-), pacientes sem metástase nodal e alta expressão de *NR3C2* (N0 NR3C2+) e pacientes sem metástase nodal e baixa expressão de *NR3C2* (N0 NR3C2-). A partir dessa separação geramos três curvas de sobrevida, sendo a primeira incluindo os quatro grupos, e as outras duas incluindo separadamente os pacientes do grupo N+ e do grupo N0 (**Figura 32**).

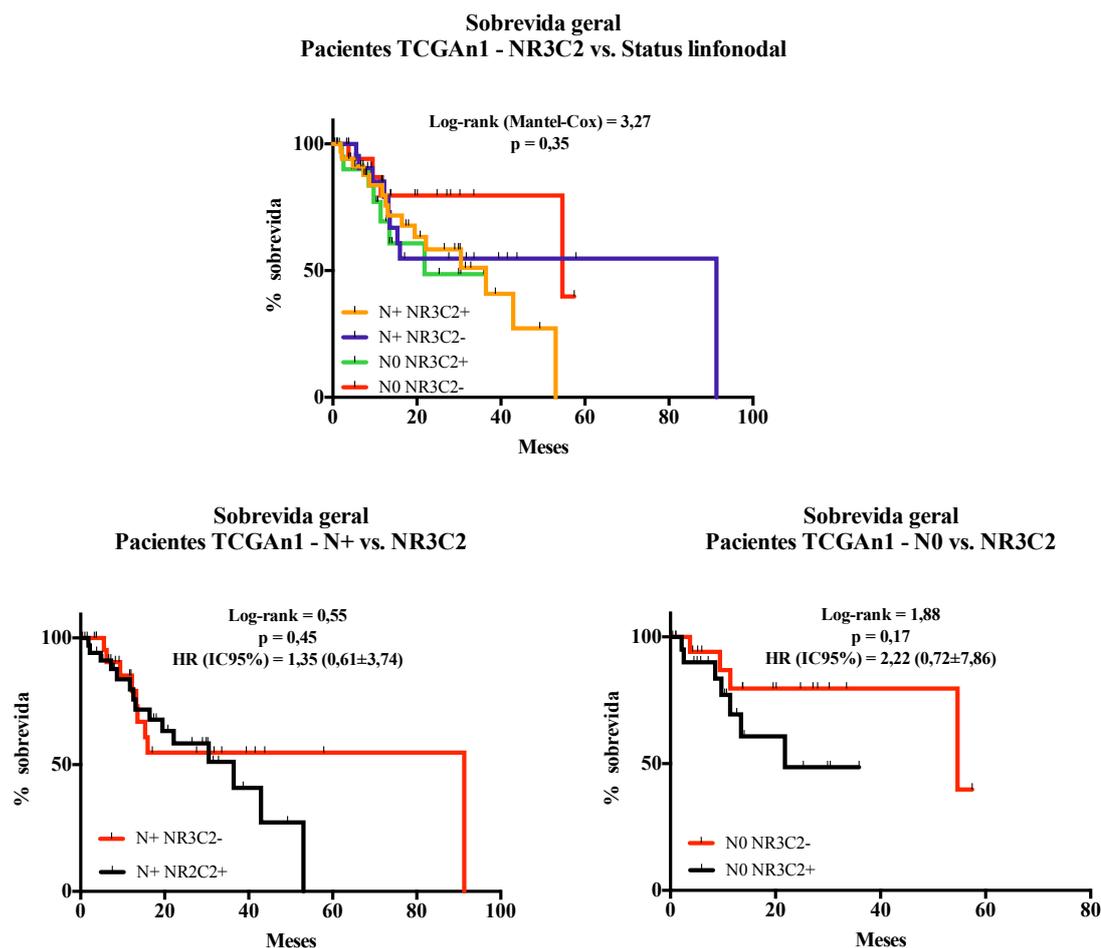


Figura 32 - Curvas de Kaplan-Meier geradas a partir da categorização dos pacientes da seleção TCGAn1 combinando o status linfonodal e a expressão de *NR3C2* e seus respectivos desmembramentos de acordo com o status linfonodal.

Embora não se observe diferenças significativas entre as curvas de sobrevida, verificamos que tanto no grupo de pacientes com tumores metastáticos quanto no grupo de tumores não metastáticos, a presença de alta expressão de *NR3C2* parece estar relacionada à menor sobrevida (média de 21,8 meses para N0 NR3C2+, 36,8 meses para N+ NR3C2+, 54,6 meses para N0 NR3C2- e 91,3 meses para N+ NR3C2-). Adicionalmente, podemos dizer que, considerando esses dados, o padrão de expressão do

NR3C2 parece ser um marcador de sobrevida mais promissor que o status dos linfonodos cervicais.

Tendo em vista os resultados promissores obtidos a partir das análises de expressão diferencial e curva ROC, assim como as tendências demonstradas nas curvas de Kaplan-Meier, podemos supor que o transcrito *NR3C2* provavelmente possui um papel importante no CEB. Entretanto, até o momento não há relatos na literatura sobre a importância deste gene nesta neoplasia.

O gene *NR3C2* (*nuclear receptor subfamily 3, group C, member 2*) está localizado na região 4q31.1 e seu transcrito irá traduzir um receptor de mesmo nome. O receptor *NR3C2* é mais conhecido como receptor mineralocorticoide (RM) e faz parte de uma superfamília de receptores nucleares que funcionam como fatores de transcrição dependentes de um ligante. Seu principal agonista é a aldosterona que induz a transcrição de sinais que irão culminar na manutenção da osmorregulação através do controle de canais de iônicos (ZENNARO et al. 2001; FULLER 2015). Outros ligantes desse receptor são o cortisol e a progesterona, sendo que a última faz papel de antagonista (FULLER 2015).

O mecanismo mais conhecido dos RM está na sua interação com a aldosterona nas células epiteliais renais, induzindo o aumento da atividade de canais iônicos levando à retenção de Na^+ e liberação de K^+ (STOCKAND 2002; DUPONT et al. 2014). Esse mecanismo ocorre pela ação do RM sobre os canais de Na^+ , tanto pelo aumento da abertura desses canais, como também ampliando o número de canais ativos (STOCKAND 2002).

A interação entre a aldosterona e o RM também altera a atividade de outros canais iônicos, como os canais de cálcio voltagem-dependentes e os canais de potássio ativados por cálcio. Dessa forma, o RM possui um amplo papel na regulação de canais iônicos controlando o fluxo de íons no meio intracelular (DUPONT et al. 2014).

SCHWAB et al. (2012) discutem que a aldosterona induz a transcrição de canais epiteliais de Na⁺ aumentando a capacidade de migração celular. O aumento dos canais epiteliais de Na⁺ causa despolarização de certas regiões da membrana celular causando a migração da célula através da indução da formação do citoesqueleto de actina. Tendo em vista que a modulação deste processo é diretamente dependente da ação do RM, acreditamos que nossos achados de que o aumento da transcrição de *NR3C2* em tumores metastáticos não é decorrente de um simples acaso, mas está diretamente relacionado a um processo biológico importante para o desenvolvimento metastático em CEB.

HUANG et al. (2011) investigaram o padrão de expressão gênica de tumores do cérvix uterino para predição de metástases linfonodais pélvicas através de análise de *microarray*. A assinatura gênica da metástase nodal nestes tumores incluía o *NR3C2* como um importante gene para diagnóstico de metástase, sendo observada uma tendência de aumento da sua expressão nos casos metastáticos, assim como observamos em nosso estudo.

Embora o transcrito *NR3C2* tenha sido pouco abordado na literatura, especialmente na sua relação com neoplasias malignas, acreditamos que

este possa representar um importante marcador tumoral, especialmente considerando o seu papel biológico como modulador de canais iônicos.

Têm-se demonstrado uma importante relação entre os canais iônicos e suas atividades com os processos movimentação e migração celular. Consequentemente, muita atenção têm sido dada a estes mecanismos na modulação do processo metastático (FISKE et al. 2006; DEL MÓNACO et al. 2009; PREVARSKAYA et al. 2011; BRACKENBURY 2012; SCHWAB et al. 2012).

O mecanismo básico de movimentação celular se baseia em 3 processos. O primeiro consiste no princípio de que a célula precisa sofrer uma polarização estrutural que ocorre a partir da geração de prolongamentos citoplasmáticos (lamelopódio) no sentido do movimento, e uma contração da membrana celular na região oposta. A ocorrência deste processo de forma repetitiva irá permitir o deslocamento celular. Associado a este movimento, deve haver a formação de contatos adesivos com a matriz extracelular. Estes contatos devem ser temporários permitindo que, após o deslocamento da célula, eles se quebrem e ocorram em novos sítios do substrato, caracterizando um deslocamento espacial. Essa movimentação será determinada pela presença de agentes estimuladores de movimento, denominados de agentes quimiotáticos. Os agentes quimiotáticos irão promover mudanças na superfície da membrana celular, alterando a constituição do citoesqueleto favorecendo tanto a formação do lamelopódio quanto a formação de interações com a matriz extracelular. Variações no gradiente do agente quimiotático irão determinar a direção da formação do

lamelopódio e conseqüentemente a direção do deslocamento celular (SCHWAB et al. 2012).

Curiosamente, todos os passos descritos acima, são modulados por proteínas envolvidas na regulação do transporte iônico celular. Essas proteínas, representadas principalmente pelos canais iônicos da membrana, irão mediar este processo através da estimulação de mudanças no potencial de membrana das células, através da regulação do volume celular, da concentração de Ca^{2+} intracitoplasmático e também modulando o pH do meio intra e extracelular (SCHWAB et al. 2012). Todas essas ações dos canais iônicos irão favorecer ou desfavorecer o processo de deslocamento da célula pelo substrato, e dessa forma podem ter importância no processo metastático. De fato, a relação entre os canais iônicos e o processo metastático têm sido bastante estudada, comprovando a relação entre mudanças no funcionamento desses canais e o favorecimento do processo de disseminação tumoral (FISKE et al. 2006; PREVARSKAYA et al. 2011; BRACKENBURY 2012).

SCHWAB et al. (2012) ressaltam algumas proteínas e canais iônicos com papel importante na modulação da migração e invasão celular como: canais de K^+ (canais de K^+ voltagem-dependentes, canais de K^+ sensíveis ao Ca^{2+} , canais de K^+ retificadores de influxo), canais de Na^+ voltagem dependentes, canais de Na^+ epiteliais, canais de Cl^- (canais aniônicos regulados por volume, canais TMEM16A/ANO1, receptores GABA), proteínas regulatórias do transporte de cálcio e aquaporinas.

O cálcio têm sido considerado um importante ator no processo metastático. Isso porque as concentrações de Ca^{2+} são importantes na regulação da migração celular (PREVARSKAYA et al. 2011). É importante ressaltar que entrada de Ca^{2+} é mediada por diversos canais, dentre eles encontram-se os canais de Ca^{2+} voltagem-dependentes que são controlados pelos RM (PREVARSKAYA et al. 2011; DUPONT et al. 2014).

Entretanto, a relação entre canais iônicos e o processo metastático ainda é pouco abordada em CEB (LI et al. 2014; MATSUO e KAWANO 2014). LI et al. (2014) demonstrou que o aumento da expressão de ANO1, que é uma proteína transmembrana que atua como um canal de Cl^- ativado por Ca^{2+} , esteve associada à ocorrência de metástase nodal em CEB. MATSUO e KAWANO (2014) investigaram o papel das aquaporinas em CEB. As aquaporinas são canais de membrana responsáveis pela regulação do volume de água intracelular. Os autores verificaram que a expressão desses canais está relacionada à diferenciação dos queratinócitos, e que a perda da sua expressão está diretamente ligada ao aumento da ocorrência de metástases nodais em CEB.

Dessa forma, buscamos nos nossos dados evidências de que alterações em vias relacionadas à modulação dos canais iônicos celulares estivessem ocorrendo de forma distinta entre tumores metastáticos e não metastáticos. Tendo como critério estatístico o p-valor corrigido, não observamos nenhum enriquecimento de vias relacionadas a estes processos. Entretanto, ao considerar apenas o p-valor, diversas vias biológicas relacionadas à modulação desses canais surgiram

diferencialmente expressas entre tumores metastáticos e não metastáticos

(Tabelas 6 e 7).

Tabela 6 - Vias biológicas hiperativadas em tumores metastáticos relacionadas ao transporte de íons pela membrana celular

Termo	Gene Ontology (ID)	Total de genes da via	p-valor	p-valor corrigido	Genes
<i>Monovalent inorganic cation homeostasis</i>	GO:0055067	109	0,007	0,47	NR3C2,SLC9A2,SLC4A11
<i>Sodium ion transmembrane transport</i>	GO:0035725	111	0,007	0,47	SLC9A2,SLC4A11,FGF12
<i>Inorganic anion transmembrane transport</i>	GO:0098661	122	0,009	0,47	ANO1,SLC4A11,GABRB3
<i>Inorganic anion transmembrane transporter activity</i>	GO:0015103	132	0,011	0,47	ANO1,SLC4A11,GABRB3
<i>Chloride channel complex</i>	GO:0034707	47	0,012	0,47	ANO1,GABRB3
<i>Active transmembrane transporter activity</i>	GO:0022804	272	0,016	0,47	ABCA8,ABCA3,SLC9A2,SLC4A11
<i>Hydrogen ion channel activity</i>	GO:0015252	5	0,020	0,47	SLC4A11
<i>Cellular monovalent inorganic cation homeostasis</i>	GO:0030004	63	0,021	0,47	NR3C2,SLC4A11
<i>Regulation of pH</i>	GO:0006885	65	0,023	0,47	SLC9A2,SLC4A11
<i>Chloride channel activity</i>	GO:0005254	67	0,024	0,47	ANO1,GABRB3
<i>Sodium ion transport</i>	GO:0006814	175	0,024	0,47	SLC9A2,SLC4A11,FGF12
<i>Transmitter-gated ion channel activity</i>	GO:0022824	7	0,027	0,47	GABRB3
<i>Transmitter-gated channel activity</i>	GO:0022835	7	0,027	0,47	GABRB3
<i>Transmembrane transporter activity</i>	GO:0022857	832	0,029	0,47	ANO1,ABCA8,GABRB3,ABCA3,SLC9A2,SLC4A11,FXYP6
<i>Anion channel activity</i>	GO:0005253	76	0,030	0,47	ANO1,GABRB3
<i>Sodium:proton antiporter activity</i>	GO:0015385	9	0,034	0,47	SLC9A2
<i>Hydrolase activity, acting on acid anhydrides, catalyzing transmembrane movement of substances</i>	GO:0016820	83	0,035	0,47	ABCA8,ABCA3
<i>Monovalent cation:proton antiporter activity</i>	GO:0005451	10	0,037	0,47	SLC9A2
<i>Inorganic anion transport</i>	GO:0015698	208	0,038	0,47	ANO1,SLC4A11,GABRB3
<i>Ion channel activity</i>	GO:0005216	363	0,040	0,47	FXYP6,ANO1,SLC4A11,GABRB3
<i>Intracellular calcium activated chloride channel activity</i>	GO:0005229	11	0,040	0,47	ANO1
<i>Chloride transport</i>	GO:0006821	91	0,041	0,47	ANO1,GABRB3
<i>Substrate-specific channel activity</i>	GO:0022838	372	0,043	0,47	FXYP6,ANO1,SLC4A11,GABRB3
<i>Cellular sodium ion homeostasis</i>	GO:0006883	13	0,047	0,47	NR3C2
<i>Inorganic anion exchanger activity</i>	GO:0005452	13	0,047	0,47	SLC4A11
<i>Bicarbonate transmembrane transporter activity</i>	GO:0015106	13	0,047	0,47	SLC4A11
<i>Channel activity</i>	GO:0015267	388	0,049	0,47	FXYP6,ANO1,SLC4A11,GABRB3
<i>Passive transmembrane transporter activity</i>	GO:0022803	388	0,049	0,47	FXYP6,ANO1,SLC4A11,GABRB3

Tabela 7 - Vias biológicas hipoativadas em tumores metastáticos relacionadas ao transporte de íons pela membrana celular

Termo	Gene Ontology (ID)	Total de genes da via	p-valor	p-valor corrigido	Genes
<i>Calcium ion transmembrane transporter activity</i>	GO:0015085	117	0,005	0,34	<i>CHRNA9, TRPV3, ATP2C2</i>
<i>Divalent inorganic cation transmembrane transporter activity</i>	GO:0072509	141	0,009	0,36	<i>CHRNA9, TRPV3, ATP2C2</i>
<i>Calcium ion binding</i>	GO:0005509	605	0,01	0,37	<i>CDHR5, DSC1, PCDHB3, PCDHGA8, PRSS3, DNER</i>
<i>Calcium ion transmembrane transport</i>	GO:0070588	165	0,01	0,37	<i>CHRNA9, TRPV3, ATP2C2</i>
<i>Transmitter-gated ion channel activity</i>	GO:0022824	7	0,02	0,37	<i>CHRNA9</i>
<i>Calcium-transporting atpase activity</i>	GO:0005388	9	0,02	0,37	<i>ATP2C2</i>
<i>Calcium channel activity</i>	GO:0005262	100	0,03	0,37	<i>CHRNA9, TRPV3</i>
<i>Cation:amino acid symporter activity</i>	GO:0005416	12	0,03	0,37	<i>SLC6A15</i>
<i>Cellular copper ion homeostasis</i>	GO:0006878	13	0,04	0,37	<i>APP</i>
<i>Cellular response to fluid shear stress</i>	GO:0071498	13	0,04	0,37	<i>CA2</i>
<i>Cellular cation homeostasis</i>	GO:0030003	437	0,04	0,37	<i>HIF1A, CHRNA9, APP, CA2</i>
<i>Copper ion homeostasis</i>	GO:0055070	15	0,04	0,37	<i>APP</i>
<i>Cellular ion homeostasis</i>	GO:0006873	449	0,04	0,37	<i>HIF1A, CHRNA9, APP, CA2</i>

Dessa forma, podemos dizer que processos relacionados ao controle de canais iônicos na células de CEB podem ter um papel de extrema importância na regulação do processo metastático, e que um dos transcritos relacionados a estes processos, o *NR3C2*, mostrou-se um importante biomarcador para diagnóstico da metástase nodal.

5 CONCLUSÃO

A partir dos resultados gerados neste trabalho podemos concluir que:

- O padrão de expressão gênica do carcinoma epidermóide de boca pode determinar quais tumores possuem fenótipo metastático;
- Identificamos 150 transcritos com potencial de diferenciação entre tumores metastáticos e não metastáticos;
- Através de etapas de validação em amostras independentes, *in silico* e através de qRT-PCR, identificamos um grupo de transcritos potenciais biomarcadores de metástase nodal em carcinoma epidermóide de boca, sendo o transcrito *NR3C2* o mais promissor.

6 PERSPECTIVAS FUTURAS

Como o intuito deste projeto é estabelecer biomarcadores de metástase nodal que possam ter aplicabilidade clínica para melhorar a acurácia do diagnóstico das metástases nodais em pacientes com carcinoma epidermóide de boca, diversas outras análises ainda serão necessárias.

Inicialmente, como o grupo clínico que mais se beneficiaria destes biomarcadores é representado por pacientes com tumores T1 e T2 sem evidências clínicas de metástase nodal (cN0), novas validações devem ser feitas exclusivamente nestes indivíduos. Temos disponíveis amostras de pacientes com estas características em nosso arquivo. Entretanto, como estas são compostas por tecidos fixados em formalina e embebidos em parafina, ainda estão sendo realizadas padronizações para obter eficiência nas reações de validação. Os resultados obtidos a partir destas novas análises irão determinar a aplicabilidade clínica dos marcadores encontrados neste estudo.

Além disso, será importante verificar a importância da combinação da expressão de diferentes transcritos na acurácia de diagnóstico da metástase nodal. Ensaios de expressão em amostras independentes incluindo todos os 150 genes diferencialmente expressos iriam favorecer a determinação da combinação de marcadores que proveriam maior especificidade e sensibilidade ao teste diagnóstico. Ensaios customizados de painéis *TaqMan*

Array (ThermoFisher Scientific) podem representar uma solução técnica para verificação de múltiplos transcritos em uma única amostra e determinar a melhor combinação de transcritos para o diagnóstico acurado da metástase nodal. Por fim, a super-expressão ou o bloqueio da expressão dos genes candidatos, seguidos de ensaios funcionais envolvendo migração celular podem reforçar a importância dos marcadores aqui apresentados.

7 DOUTORADO SANDUÍCHE

Durante o ano de 2014, realizei estágio doutorado sanduíche financiado pela FAPESP (Processo 2013/09142-2) na *School of Dentistry - University of Califórnia, Los Angeles*, no *Center for Oral/Head & Neck Oncology Research* chefiado pelo Dr. David T. W. Wong, para a realização do projeto “Estudo de marcadores salivares relacionados à metástase nodal em carcinoma epidermóide de boca”.

O objetivo deste estágio foi verificar possíveis marcadores transcricionais de metástase nodal na saliva de pacientes com carcinoma epidermóide de boca. Entretanto, devido à natureza prospectiva deste projeto, não foi possível coletar amostras de saliva de pacientes com CEB suficientes para a finalização deste estudo. Durante o período de um ano apenas 6 amostras foram coletadas.

Apesar desta limitação, o Dr. David Wong permitiu que eu dirigisse outros 4 projetos em seu laboratório que permitiriam abordar todas as possibilidades de pesquisa no uso da saliva como meio diagnóstico de doenças humanas. Dessa forma, desenvolvi estudos de descoberta e validação de biomarcadores na saliva, assim como um estudo de demonstração dos mecanismos biológicos envolvidos na liberação de biomarcadores na saliva.

Nesta tese será apresentado na forma de um artigo um desses projetos que submetido ao periódico Clinical Cancer Research. O artigo está em fase de correção após primeira avaliação pelos revisores.

Além deste projeto, foram realizados outros três trabalhos ainda não finalizados e, dessa forma, decidimos não incluí-los nesta tese.

A seguir, será apresentado o artigo referente à validação de biomarcadores salivares para detecção de CEB e diferenciação entre a neoplasia e lesões potencialmente malignas. Os resultados deste trabalho podem representar um grande avanço em políticas de prevenção ao câncer de boca através da detecção de lesões potencialmente malignas e de diagnóstico precoce a partir do diagnóstico de lesões neoplásicas.

Artigo submetido ao periódico *Clinical Cancer Research*

Salivary biomarkers for detection of oral squamous cell carcinoma in a Taiwanese population

Frederico Omar Gleber-Netto^{1,2}, Maha Yakob¹, Feng Li¹, Ziding Feng³, Jianliang Dai³, Huang-Kai Kao⁴, Yu-Liang Chang⁵, Kai-Ping Chang^{6,7}, David T.W. Wong¹

¹Center for Oral/Head & Neck Oncology Research, University of California, Los Angeles, 73-017 CHS, 10833 Le Conte Avenue, Los Angeles, CA 90095, U.S.A.

²Medical Genomic Laboratory, Centro Internacional de Pesquisa e Ensino - CIPE, AC Camargo Cancer Center, Rua Taguá 440, São Paulo, SP 01508-010, Brazil.

³Department of Biostatistics, The University of Texas, MD Anderson Cancer Center, Houston, Tx 77030, U.S.A.

⁴Department of Plastic & Reconstructive Surgery, Chang Gung Memorial Hospital, Tao-Yuan, Taiwan

⁵ Department of Oral and Maxillofacial Surgery, Chang Gung Memorial Hospital, Tao-Yuan, Taiwan

⁶Department of Otolaryngology - Head & Neck Surgery, Chang Gung Memorial Hospital, Tao-Yuan, Taiwan

⁷ College of Medicine, Chang Gung University, Tao-Yuan, Taiwan.

Abstract

Purpose: This study evaluated the discriminatory power of salivary transcriptomic and proteomic biomarkers in distinguishing oral cancer (OSCC) cases from controls and potentially malignant oral disorders (PMOD).

Experimental design: A total of 180 samples (60 OSCC patients, 60 controls and 60 PMOD patients) were used in the study. Seven transcriptomic markers (IL-8, IL-1 β , SAT1, OAZ1, DUSP1, S100P, H3F3A) were measured using quantitative real time PCR and two proteomic markers (IL-8 and IL-1 β) were evaluated by ELISA.

Results: Among 7 transcriptomic markers, the transcript level of DUSP1 was significantly lower in OSCC patients than in controls and PMOD patients. Between the proteomic markers, the protein concentration of IL-8 and IL-1 β was significantly higher in OSCC patients than controls and dysplasia patients. Univariate fractional polynomial models revealed that salivary IL-8 protein has the highest AUC value between OSCC patients and controls (0.74) and between OSCC and PMOD patients (0.72). Applying a 2-markers fractional polynomial model, salivary IL-8 protein combined with IL-1 β gave the best AUC value for discrimination between OSCC patients and controls, as well as the IL-8 protein combined with H3F3A mRNA gave the best AUC value for discrimination between OSCC and PMOD patients. Multivariate models analysis combining salivary analytes and risk factor exposure related to oral carcinogenesis formed the best combinatory variables for differentiation between OSCC vs PMOL (AUC=0.80), OSCC vs controls (AUC=0.87) and PMOD vs. controls (AUC=0.78).

Conclusions: Combination of transcriptomic and proteomic salivary markers is of great value for oral cancer detection and differentiation from PMOD patients and controls.

Introduction

Oral squamous cell carcinoma (OSCC) is the most common cancer of the head and neck region. High morbidity and mortality is associated with this disease, but little improvement has been observed in the five-years survival rate for patients with OSCC along the years (CARVALHO et al. 2005). One of the main prognostic factors for OSCC patients is advanced disease (American Cancer Society-ACS 2012; VAN HARTEN et al. 2014). Considering that, early diagnosis of oral cancer is an important approach to decrease morbidity and mortality rate.

A multi-step carcinogenesis process characterizes OSCC development. Cumulative mutational events occur in the mucosal epithelial stem cells, and the cellular proliferation promotes the expansion of a field of DNA altered cells in the epithelial lining (CALIFANO et al. 1996; BRAAKHUIS et al. 2003; LEEMANS et al. 2011). Clinical and histopathological signs of altered epithelium can, sometimes, be observed in a form of leukoplakia/erythroplakia (also called potentially malignant oral disorders - PMOD) and cellular dysplasia, respectively (NAPIER e SPEIGHT 2008). Under the effect of new mutational events, malignant transformation can occur in these areas, leading to development of an infiltrative disease (CALIFANO et al. 1996).

Although the rate of malignant transformation of clinically altered epithelium is low (about 0.13% to 17.5%) (NAPIER e SPEIGHT 2008), detection and close follow up of these lesions are the best approach to early diagnosis of oral cancer. Nowadays, the unique available method for detection of altered

epithelium and oral cancer is clinical examination, but it does not allow a reliable differentiation between PMOD and lesions with no risk to cancer progression (MOYER 2014). The rate of detection of early stage oral cancer is low. This can be explained by the asymptomatic characteristic of early stage disease and the lack of an adequate routine mucosal exam by health care practitioners (NEVILLE e DAY 2002; MCGURK et al. 2005). The development of a reliable detection method of OSCC and PMOD would be of importance to improve early diagnosis of oral cancer. Consequently, this would favor early treatment, changing survival rates and avoiding the devastating consequences of advanced tumors treatment.

Saliva has been considered an important source of biological information for detection of human diseases. Beyond the obvious relationship with the oral mucosa surface, several studies have demonstrated synergism between expression of molecular markers in saliva and systemic or distant sites diseases. Metabolites, proteins, coding and non-coding RNAs and DNA have been detected in saliva of diseased patients, showing important value in disease detection (LI et al. 2004; ARELLANO-GARCIA et al. 2008; PARK et al. 2009; NAGLER 2009; BERNARDES et al. 2011; ELASHOFF et al. 2012; KIM et al. 2013; WEI et al. 2014).

The role of saliva to detect oral cancer has been studied showing encouraging results (ARELLANO-GARCIA et al. 2008; NAGLER 2009; PARK et al. 2009; BRINKMANN et al. 2011; CHENG et al. 2011; ELASHOFF et al. 2012). In 2004 LI et al. it was determined the salivary transcriptome of oral cancer, showing that it contains a set of promising extracellular RNA

(exRNA) markers. These salivary exRNA markers were tested in different OSCC populations showing promising performance for disease detection with sensitivity and specificity higher than 80% (BRINKMANN et al. 2011; ELASHOFF et al. 2012). Inflammatory cytokines also have been investigated as potential biomarkers of oral cancer (HOFFMANN et al. 2007; ARELLANO-GARCIA et al. 2008; ELASHOFF et al. 2012; KOROSTOFF et al. 2011; PUNYANI e SATHAWANE 2013). HOFFMANN et al. (2007) verified that serum levels of interleukin 8 (IL-8) were significantly higher in patients with squamous cell carcinoma of the head and neck region when compared with individuals without cancer. ARELLANO-GARCIA et al. (2008) also studied the potential of cytokines to predict cancer, but using saliva as a source of biological information. They observed that IL-8 and IL-1 β were significantly more expressed in saliva of OSCC patients than in healthy controls.

ELASHOFF et al. (2012) demonstrated that the combination of such protein and RNA markers improved the power of predictability of salivary OSCC biomarkers. Although the efficacy of these RNA and protein markers of oral cancer has been demonstrated, race related variations could occur in biomarker discovery. Considering that an ideal biomarker should have a widespread efficacy regardless of ethnicity it is important to challenge these biomarkers in a new population.

Bearing in mind the significance of PMOD in the early diagnosis of OSCC, we decided, for the first time, to test the efficacy of these biomarkers to detection of precursor lesions. The primary objective of this analysis is to assess the discriminatory power of 7 salivary transcriptomic markers (IL-1 β ,

IL-8, SAT1, OAZ1, DUSP1, S100P, H3F3A) and 2 proteomic markers (IL-8, IL-1b) in distinguishing OSCC and PMOD patients from health individuals.

Materials and Methods

Patient selection

Saliva samples were collected, after approval by the Institutional Review Board (#104-2602C), from patients that agreed to sign an informed consent in the Linko Medical Center of Chang Gung Memorial Hospital at Tao-Yuan, Taiwan. A total of 180 samples were included in the dataset, including 60 cases of oral squamous cell carcinoma (OSCC), 60 cases of potentially malignant oral disorders (PMOD) with histopathological evidence of dysplasia and 60 individuals with no clinical sign of malignant or potentially malignant disease in the oral cavity.

Controls were gender and age-matched subjects enrolled during the same period when OSCC subjects were recruited. PMOD and controls received routine physical examination of head and neck regions, and patients with any head and neck disease (exception for PMOD) were excluded. All patients with a history of prior cancer, diabetes, autoimmune disorders, hepatitis, or HIV infection were excluded.

Demographical information was obtained by an IRB approved questionnaire. Clinical information of OSCC patients was obtained by pathological report generated after tumor resection.

Saliva collection and processing

Unstimulated saliva was collected and processed separately for RNA and protein according to our published protocol (LI et al. 2004). For saliva collection, the donors avoided eating, drinking, smoking, and using oral hygiene products for at least 1 hour before the procedure. The collected samples were centrifuged at $3000 \times g$ for 15 min at 4°C . The supernatants were immediately treated with a protease inhibitor mixture (Roche; Cat. No.: 11836145001) and RNase inhibitor (Invitrogen, 10777-019). The samples were aliquoted into smaller volumes and stored at -80°C refrigerator. To avoid protein degradation thawed saliva samples were used once. Saliva was collected at diagnosis for patients with OSCC and PMOD before any surgical procedure.

All laboratory measurements of salivary biomarkers were performed at School of Dentistry, Center for Oral/Head & Neck Oncology Research, University of California - Los Angeles, Los Angeles, CA - USA.

Primer design

We used a nested PCR approach for measurement of RNA molecules in saliva. Outer primers (OF - outer forward; OR - outer reverse) were designed for cDNA synthesis and pre-amplification. Inner primers (IF - inner forward; IR - inner reverse) were designed for qPCR measurement of the cDNA targets. Primer pairs were designed using the NCBI/Primer-BLAST software (www.ncbi.nlm.nih.gov/tools/primer-blast). Three genes were used as saliva internal reference: GAPDH (OF: CATTGCCCTCAACGACCACTT; OR:

ATGTGGGCCATGAGGTCCA; IF: ACCACTTTGTCAAGCTCATTTCCT; IR: CACCCTGTTGCTGTAGCCAAAT), ACTB (OF: GATCATTGCTCCTCCTGAGC; OR: CGGACTCGTCATACTCCTGC; IF: CTCCTGAGCGCAAGTACTCC; IR: ATRACTCCTGCTTGCTGATCCAC) and RPS9 (OF: GCGCCATATCAGGGTCCGCA; OR: CCCCGTAGGGAGAGCGCAG; IF: GCCATATCAGGGTCCGCAAGCA; IR: TCTGGGAATCCAGGCGGACAA). Only samples that exhibited specific qPCR products for these 3 genes were used in the study. Ct values of all target genes were normalized according to RPS9 gene expression.

The target RNAs measured in saliva of the studied individuals were: IL-1 β (OF: GTACCTGTCCTGCGTGTTGAAAG; OR: TTCTATCTTGTTGAAGACAAATCGCTT; IF: TGTTGAAAGATGATAAGCCCACTCT; IR: CAAATCGCTTTTCCATCTTCTTCT), IL-8 (OF: TTTCTGATGGAAGAGAGCTCTGTCT; OR: ATCTTCACTGATTCTTGGATACCACA; IF: CCAAGGAAAAGTGGGTGCAG; IR: CTTGGATACCACAGAGAATGAATTTTT), SAT1 (OF: GCACTGGACTCCGGAAGGACAC; OR: ACTTGCCAATCCACGGGTCA; IF: GGACTCCGGAAGGACACAGCATTG; IR: GCCAATCCACGGGTCATAGGTA), OAZ1 (OF: TGCGAGCCGACCATGTC; OR: CCCCGGTCTCACAATCTCAA; IF: TCTTCATTTGCTTCCACAAGAACC; IR: TCAAAGCCCAAAAAGCTGAAG), DUSP1 (OF: CCTGTGGAGGACAACCACAAG; OR: GCCTGGCAGTGGACAAACA; IF: CAGACATCAGCTCCTGGTTCAA; IR:

CAAACACCCTTCCTCCAGCAT), S100P (OF: GCACGCAGACCCTGACCA;
 OR: CGTCCAGGTCCTTGAGCAATT; IF:
 GCTGATGGAGAAGGAGCTACCA; IR: TTGAGCAATTTATCCACGGCAT),
 H3F3A (OF: AGCGTCTGGTGCGAGAAATT; OR:
 GCACACAGGTTGGTGTCTTCAA; IF: CGCTTCCAGAGCGCAGCTAT; IR:
 TCTTCAAAAAGGCCAACCAGAT).

Direct saliva transcriptome analysis

A multiplex cDNA synthesis and pre-amplification approach was performed directly in 4µl of saliva for each sample. This technique was developed by our group and is called Direct Saliva Transcriptome Analysis (DSTA) (LEE et al. 2011). The reactions were performed in 10µl volume with a pool of outer primers at 50µM each and the SuperScript III Taq (Life Technologies, Carlsbad, CA) and 2X Reaction Mix. The thermocycler program used is described following: 2 minutes at 60°C, 30 minutes at 50°C, 2 minutes at 95°C, and 15 cycles of 15 seconds at 95°C, 30 seconds at 50°C, 10 seconds at 60°C and 10 seconds at 72°C and cooling at 4°C. The PCR products were treated with 4µl of ExoSAP-IT (Affymetrix, Santa Clara, CA) for 15 minutes at 37°C and then heated to 80°C for 15 minutes.

The target transcripts were quantified from 2µl of pre-amplified cDNA via singleplex qPCR using the Roche LightCycler 480 instrument (Roche, San Francisco, CA). The qPCR reactions were done in a 10µl volume containing 50µM of each inner primer pair and 2X SYBR Green qPCR Mix. The qPCR was carried out under the following conditions: 95°C for 5 minutes and 40

cycles of 10 seconds at 95°C, 10 seconds at 60°C, 10 seconds at 72°C. All of the qPCR reactions were performed in duplicate.

Salivary protein detection

Salivary IL-8 and IL-1 β proteins were measured using specific ELISA kits (Thermo Fisher) according to manufacturer's recommendations. For measurement, saliva samples were diluted in PBS according to recommendation of a previous study (ELASHOFF et al. 2012). For IL-8, saliva was diluted 1:8 and for IL-1 β it was diluted 1:3. For interleukin values correction, total salivary protein was measured using Bradford's method (Bio-Rad) using saliva dilutions of 1:3. All samples were assayed in duplicates using a microplate reader and the results were expressed in picograms per milliliter.

Statistical analysis

Missing values were replaced by the $\frac{1}{2}$ of minimum of the variable. Using the raw data, nonparametric ANOVA (Kruskal-Wallis Test) was performed to test the overall difference in biomarkers between three groups. If the overall difference is significant, Wilcoxon rank sum test were used to perform two specific comparisons: OSCC vs. controls, OSCC vs. PMOD. The mRNA markers were normalized to RPS9 gene expression. The protein markers were normalized to the total protein followed by standardization. ROC analysis was performed after running univariate Fractional Polynomial model (FP) (ROYSTON e ALTMAN 1994; SAUERBREI et al. 2006). The marker

with highest AUC value was used as the anchor marker. The relationships between the anchor marker and other markers were checked by Spearman correlation analysis and visualized by 2-D plots. Whether adding a second marker will increase the discriminatory power of the anchor marker was checked by 2 methods: ROC analysis after running the FP models and logistic regression models. The AUC values and AIC values were compared for different two marker-models. Additionally, multivariate analyses including all salivary markers and risk factors exposure were carried out after running FP and logistic regression models. Statistical analysis was carried out using R version 3.0.2 (R Foundation, Vienna, Austria).

Results

Clinical and demographical data from patients included in this study are described in the **Table 1**. No difference was observed among groups considering gender and age. Ethanol intake as well as betel nut chewing was higher in OSCC and PMOD patients than in controls. Tobacco use was higher in the PMOD group than among OSCC patients and controls.

The expression values for the analyzed salivary mRNAs and proteins are described in the **Table 2**. Significant differences were observed in the expression level of DUSP1 between OSCC patients and controls (p-value = 0.0123) as well as between OSCC patients and PMOD patients (p-value= 0.0422). The concentration of IL-8 protein (IL-8p) was significantly higher in OSCC patients when compared to controls (p-value < 0.0001) and PMOD patients (p-value < 0.0001) (**Figure 1a**). Similarly, salivary IL-1 β protein (IL-

1 β p) concentration was significantly higher in OSCC patients than in controls (p-value < 0.01) and PMOD patients (p-value 0.004).

Using univariate fractional polynomial models, the IL-8p gave the highest AUC (**AUC= 0.749**) in distinguishing OSCC patients from controls after running the model on each biomarker. The univariate model with IL-8p as predictor also had the lowest AIC (AIC=150.99). Similarly, the protein marker IL-8p gave a highest AUC (**AUC= 0.721**) and the lowest AIC (AIC=150.88) in distinguishing OSCC from PMOD group (**Table 3**) (**Figure 1b**).

Using the salivary expression of IL-8p as an anchor marker and applying the 2-marker fractional polynomial models, the combination of IL-8p and IL-1 β gave the highest AUC (**AUC= 0.817**) in distinguishing OSCC patients from controls. This two-marker model also had the lowest AIC (AIC=138.28) (**Figure 2a and 2b**). When the sensitivity of the test was fixed at 0.9 in distinguishing OSCC patients from controls this two-marker model gave the highest maximized specificity (MaxSpec = 0.56). In contrast, the combination of IL-8p and H3F3A gave a highest AUC (**AUC= 0.752**) and the lowest AIC (AIC=141.34) in distinguishing OSCC patients from PMOD patients (**Table 4**) (**Figure 2c and 2d**). When the sensitivity of the test was fixed at 0.9 in distinguishing OSCC patients from PMOD patients, this 2-makers model gave the highest maximized specificity (MaxSpec = 0.45).

Multivariate models analysis considering expression of salivary biomarkers and risk factor exposure were carried out using the fractional polynomial model and the logistic model (Table 5). Using the same marker combination, fractional polynomial model gave the highest AUC and the lowest AIC values

when compared to logistic model (Figure 3). Areca nut chewing, drinking and smoking habits associated to salivary expression of IL-1 β transcript was the best variable combination for distinction between PMOD patients and controls (AUC = 0.785). When sensitivity was fixed at 0.9 the MaxSpec obtained was 0.9. Areca nut chewing associated to salivary expression of DUSP1 transcript, IL-1 β p and IL-8p revealed the best combinatory effect for differentiation between OSCC patients and controls (AUC = 0.872). Fixing sensitivity at 90% MaxSpec obtained was 0.63. For separation between OSCC and PMOD patients, smoking associated to salivary H3F3A transcript and IL-8p expression showed up as the best combinatory makers (AUC=0.802). When sensitivity was set at 0.9 MaxSpec obtained was 0.53.

Discussion

South and Central Asia has one of the highest incidences and mortality rates of OSCC in the world. The consumption of areca nut (betel) associated or not with tobacco is the main risk factor for OSCC development among these people. Taiwan has one of the highest consumptions of areca nut, explaining why this country has one of the biggest prevalence of PMOD in the world (12.7%) and a high incidence rate of oral cancer (WARNAKULASURIYA 2009a; JOHNSON et al. 2011).

Areca nut usage is an important cause of PMOD, as well as tobacco. However, beyond the development of leukoplakia and erythroplakia, it can also induce a very distinct form of PMOD called oral submucous fibrosis (OSF) (LEE et al. 2003; JOHNSON et al. 2011). OSF is more prevalent among Southeast Asians, being rare in the western countries. This may

indicate that the pathways involved in the oral cancer development among Southeast Asians might diverge from occidental patients. However, even with these ethnic and behavior variations, the biomarkers developed in the western population and applied in this study showed a good performance for discrimination between oral cancer and controls in Taiwanese individuals, revealing high AUC values and high sensitivity. These biomarkers were challenged before in different populations from USA and Serbia, in which tobacco and alcohol consumption are the main etiological factor for OSCC (BRINKMANN et al. 2011; ELASHOFF et al. 2012). In these studies the performance of the biomarkers were similar from those observed in the present manuscript.

In this study, the proteomic markers had better performance than transcriptomic markers in distinguishing oral cancer cases from PMOD and controls when considered individually. The salivary IL-8 protein alone model performed best among the univariate models, always giving the highest AUC values than other individual marker. In other validation studies in western populations, IL-8 and IL-1 β proteins were also potential salivary biomarkers for oral cancer detection, showing sensitivity and specificity ranging from 70% to 80% and an AUC value around 0.7 (BRINKMANN et al. 2011; ELASHOFF et al. 2012). Although salivary IL-1 β protein was not considered a good marker for OSCC detection in our investigation, IL-8 showed AUC value (0.73) similar to other studies, confirming the reproducibility of this marker across different oral cancer populations.

However, as observed by others (LI et al. 2004; BRINKMANN et al. 2011; ELASHOFF et al. 2012), the combination of proteomic and transcriptomic

markers revealed the best discriminatory effect between oral cancer and non-cancer individuals. According to LI et al. (2004), salivary IL-8, IL-1 β , SAT and OAZ1 mRNA detection formed the best combinatory markers for OSCC detection. ELASHOFF et al. (2012) tested these biomarkers in 5 cohorts of patients and controls. They observed that for some of the cohorts the combination of biomarkers with the best performance for oral cancer diagnosis would change, giving AUC values varying from 0.75 to 0.86. However, the IL-8 and SAT mRNAs were present in all the 5 different combinations, suggesting that these biomarkers are the most consistent ones. For BRINKMANN et al. (2011) the best discriminatory markers for OSCC detection was the combination of IL-1 β protein and the SAT1 and DUSP1 mRNAs revealing an AUC value of 0.86. In our study, we obtained an AUC of 0.817 that is in the range of AUC values obtained in previous studies, but using just two markers, IL-8 protein combined to IL-1 β mRNA.

Beyond univariate and two-marker analysis we carried out a multivariate analysis including all possible combination of salivary analytes and also considering exposure to risk factors related to OSCC and PMOD development. Combining these parameters we were able to generate a set of predictors of great value for disease detection. The distinction between OSCC and controls as well as OSCC and PMOD was greatly dependent on risk factor exposure status. Using such information we observed a significant increase in test accuracy, achieving an AUC of 0.87 for differentiation between OSCC and controls and 0.80 for differentiation between OSCC and PMOD individuals.

One of the most important approaches for improve survival and decrease morbidity in oral cancer patients is the detection of early stage cancer. Detection of PMOD is of extreme importance, but its discrimination from early stage oral cancer can be challenging. Currently, the most efficient available method of diagnosis of these potentially malignant lesions and discrimination from oral cancer is biopsy followed by histopathological examination. Beyond the need of highly trained personal to do such exams, it is costly, time consuming and is associated with patient distress and risk since it is a surgical intervention. The development of a salivary biomarker with potential of discriminatory diagnosis between these two entities would be of great value. In this present study we showed, for the first time, that salivary biomarkers have discriminatory effect for discrimination between malignant and PMOD. High expression of IL-1 β transcript associated to consumption of betel, alcohol and tobacco generated the highest AUC value for differentiation between PMOD and controls. Importantly, for this combination of parameters we obtained specificity and sensibility of 90%, indicating a high accuracy test for detection of PMOD. This finding is of great importance for screening purposes in populations exposed to these risk factors, since the measurement of just one salivary biomarker would give a very accurate indication of PMOD diagnosis.

For some of our analysis, setting high sensitivity values (90%) lead to low maximum specificity. This occurred for differentiation between OSCC and controls and OSCC and PMOL. Although low specificity represents a limitation of the proposed biomarker combination, we believe that for a screening approach, high sensitivity is the most important parameter, since it

provides a low number of false negative cases and select candidate cases for complementary clinical evaluation.

Another potential limitation of our findings is the lack of periodontal evaluation of studied cases. Inflammatory diseases such as periodontitis are one of the most common pathologies in oral cavity, representing the most common inflammatory disease in humans (BATCHELOR 2014; CHAPPLE 2014). Considering that some of our candidate biomarkers are cytokines (IL-8 and IL-1 β), one may suggest that inflammatory diseases in oral cavity may represent a confusion factor in our analysis. Furthermore OSCC, PMOD and periodontitis share the same etiological factors, such as tobacco (DYE 2012; PETERSEN e OGAWA 2012). However, LISA CHENG et al. (2014) compared the salivary IL-8 protein expression between OSCC patients and patients with periodontitis. They observed that IL-8 salivary levels were significantly higher in OSCC when compared to chronic periodontitis patients ($p < 0.001$) and healthy controls ($p = 0.014$). Also, mean expression in chronic periodontitis patients was lower (0.58 ± 0.26 pg/mL) than in healthy controls (0.80 ± 0.41 pg/mL). This may indicate that inflammatory conditions may have little effect on our results. Moreover, inflammatory reaction elicited in periodontitis and in OSCC is of different nature, since microbes are the main players in the induction of the former (LISA CHENG et al. 2014).

An important aspect of this work is the use of direct saliva transcriptome analysis (DSTA) technique that allows the measurement of salivary transcripts directly from saliva with no need of RNA extraction prior analysis (LEE et al. 2011). This approach is of utmost importance for implementation of such biomarkers in a clinical practice. RNA extraction is involved with

higher costs, need of trained personal, and is time consuming. In this work we demonstrated that salivary RNA and protein could be measured directly from saliva, with no need of prior treatment. We believe that in the near future these biomarkers could be measured using portable technologies permitting its use in an ambulatory environment (27).

We concluded that proteomic and transcriptomic salivary biomarkers are of great value for oral cancer and PMOD detection in Taiwanese population. Salivary analytes and status of risk factors exposure related to oral carcinogenesis emerged as the best combination of variables for OSCC and PMOL detection. Also, for the first time, we demonstrated that the salivary analytes have discriminatory power for PMOD diagnosis, representing a potential tool for early detection of patients in risk of oral cancer development.

Acknowledgements

This study was supported by NIDCR/NIH research grants R01 DE17170 and NIH/UH2 TR000923 to D.T.W. Wong; MOST 102-2628-B-182A-012-MY3 from the Ministry of Science and Technology, Taiwan, to Kai-Ping Chang. Frederico Omar Gleber-Netto was supported by the grants 2011/13315-4 and 2013/09142-2 from São Paulo Research Foundation (FAPESP).

Tables

Table 1 - Clinical characteristics of subjects enrolled in the study								
		OSCC		PMOD		Controls		p
		Subjects	%	Subjects	%	Subjects	%	
Gender	Male	57	95	57	95	55	91.7	0.678
	Female	3	5	3	5	5	8.3	
Age	< 39	6	10	12	20	18	30	0.923
	40 - 49	21	35	15	25	11	18.3	
	50 - 59	23	38.3	21	35	14	23.3	
	60 - 69	7	11.7	9	15	10	16.7	
	> 70	3	5	3	5	7	11.7	
Ethanol consumption	Yes	41	68.33	38	63.3	25	41.7	0.007
	No	19	31.67	22	36.7	35	58.3	
Betel nut chewing	Yes	52	86.67	48	80	28	46.7	<
	No	8	13.33	12	20	32	53.3	
Tobacco consumption	Yes	50	83.33	58	96.7	49	81.7	0.02
	No	10	16.67	2	3.3	11	18.3	
Tumor size	T1	20	33.33	-	-	-	-	-
	T2	19	31.67	-	-	-	-	
	T3	5	8.33	-	-	-	-	
	T4a	14	23.33	-	-	-	-	
	T4b	2	3.33	-	-	-	-	
Lymph node status	N0	34	56.67	-	-	-	-	-
	N1	8	13.33	-	-	-	-	
	N2a	1	1.67	-	-	-	-	
	N2b	15	25	-	-	-	-	
	N2c	2	3.33	-	-	-	-	
Distant metastasis	M0	60	100	-	-	-	-	-
	I	18	30	-	-	-	-	
Clinical Stage	II	10	16.67	-	-	-	-	-
	III	6	10	-	-	-	-	
	IVa	24	40	-	-	-	-	
	IVb	2	3.33	-	-	-	-	

*ANOVA test;

Table 2 - Expression values for salivary mRNAs and proteins according to the studied groups

Variable	OSCC patients				Controls				PMOD patients				Kruskal- Wallis Test	Wilcoxon Two-Sample Test	
	Mean	SD	Min	Max	Mean	SD	Min	Max	Mean	SD	Min	Max	p-value	p-value OSCC vs. Control	p-value OSCC vs. PMOD
IL-1b	28.80	4.63	21.10	41.00	27.57	4.29	20.80	41.00	28.73	4.70	21.20	41.00	0.28		
IL-8	27.34	3.88	20.10	38.40	27.19	3.97	20.40	41.00	28.09	4.35	20.70	41.00	0.50		
SAT1	27.87	3.23	21.10	36.90	27.55	3.65	21.20	38.3	27.97	2.92	21.40	35.00	0.61		
OAZ1	23.77	3.71	15.90	32.30	24.15	3.92	17.80	33.30	24.11	3.94	17.30	33.10	0.96		
DUSP1	34.99	5.91	22.60	41.00	32.36	6.10	23.40	41.00	32.81	5.95	23.10	41.00	0.03	0.0123	0.0422
S100P	35.38	4.44	25.70	41.00	34.19	4.49	25.50	41.00	35.66	4.30	25.00	41.00	0.19		
H3F3A	20.60	4.21	12.20	33.30	21.00	4.76	13.80	33.00	20.60	3.85	13.00	32.90	0.99		
GAPDH	17.56	3.45	9.80	28.10	17.64	3.20	11.70	26.90	18.11	3.40	11.50	28.50	0.60		
ACTB	23.02	3.77	13.50	32.70	23.25	3.79	16.50	32.70	23.64	4.26	15.40	34.70	0.88		
RPS9	22.62	3.87	15.30	33.00	23.10	3.66	15.20	31.90	23.77	4.10	17.30	36.20	0.40		
IL-8_pg/ml	283.75	262.33	0.00	1048.72	127.79	110.84	0.00	654.89	140.35	155.13	0.00	1048.72	<.0001*	<.0001*	<.0001*
IL-1b_pg/ml	101.03	112.96	0.00	419.49	48.07	42.01	0.00	161.99	39.66	28.00	1.61	145.83	0.0061*	0.01*	0.004*
Total_protein_mg/ml	1.21	0.64	0.07	2.94	1.08	0.52	0.17	2.61	0.96	0.44	0.28	2.14	0.11		

Table 3 - ROC analysis using univariate fractional polynomial model of salivary transcriptomic and proteomic markers and risk factors exposure

Univariate model	PMOD vs. Controls		OSCC vs. controls		OSCC vs. PMOD	
	AUC	AIC	AUC	AIC	AUC	AIC
Variable						
IL-8	0.467	166.58	0.449	166.33	0.518	170.13
IL-1 β	0.542	170.11	0.721	151.49	0.569	168.63
OAZ1	0.508	170.32	0.519	170.33	0.576	168.07
SAT1	0.483	170.23	0.643	162	0.563	168.52
DUSP1	0.563	168.77	0.649	162.8	0.651	162.24
S100P	0.552	169.59	0.597	167.17	0.542	169.48
H3F3A	0.564	166.6	0.524	170.33	0.589	165.18
IL-1 β p	0.646	159.09	0.637	159.71	0.655	156.75
IL-8p	0.498	170.16	0.749	150.99	0.721	150.88
Total protein	0.568	168.72	0.551	166.06	0.607	164.39
Smoking	0.575	162.74	0.515	168.76	0.56	163.56
Drinking	0.608	164.66	0.639	159.5	0.531	168.45
Areca nut	0.667	155.6	0.707	144.3	0.541	167.48

Table 4 - ROC analysis of 2-marker fractional polynomial model using the salivary expression of IL-8 protein as an anchor marker

Two markers model	OSCC vs. Controls		OSCC vs. PMOD	
	AUC	AIC	AUC	AIC
IL-8p + IL-1 β	0.817	138.28	0.739	150.7
IL-8p + OAZ1	0.751	152.97	0.734	148.76
IL-8p + SAT1	0.794	145.06	0.738	150.98
IL-8p + DUSP1	0.754	148.62	0.726	153.75
IL-8p + S100P	0.767	150.54	0.723	152.16
IL-8p + H3F3A	0.75	152.86	0.752	141.34
IL-8p + IL-1 β p	0.74	152.67	0.685	155.25

Table 5 - Multivariate model ROC analyses of the optimal models using FP and logistic models

		PMOD vs. Controls	OSCC vs. Controls	OSCC vs. PMOD
Model		IL-1 β + ArecaNut + Drinking + Smoking	DUSP1 + IL-1 β p + IL-8p + ArecaNut	H3F3A + IL-8p + Smoking
Fractional polynomial model	AUC	0.785	0.872	0.802
	AIC	145.59	120.16	135.32
Logistic model	AUC	0.716	0.81	0.771
	AIC	156.32	137.28	146.07

Figures

Fig.1 - a) Salivary expression of IL-8 protein among the studied groups. IL-8 expression was significantly higher in cancer patients compared to controls and PMOD patients. Whiskers represent median. b) ROC curves for salivary IL-8 protein. The AUC value for cancer patients distinguishing from controls and PMOD patients was 0.749 and 0.721, respectively.

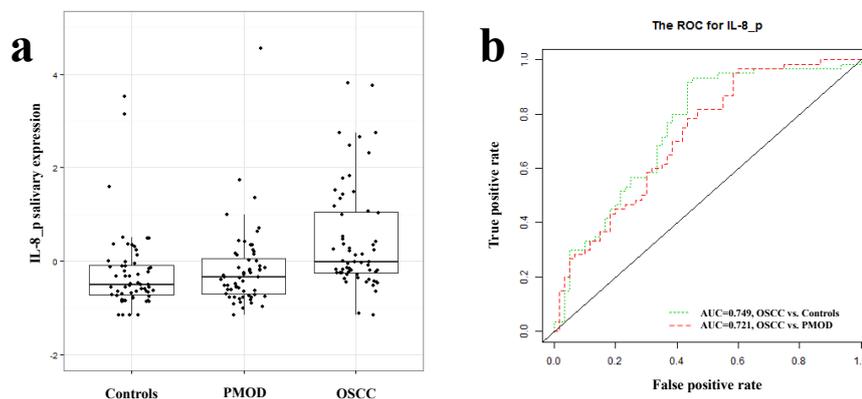


Fig.2 - a) ROC curves for combination of salivary IL-8 protein and IL-1 β mRNA. The AUC value for cancer patients distinguishing from controls and PMOD patients was 0.81 and 0.73, respectively. b) 2D scatterplot showing the correlation between the expression of salivary IL-8 protein and IL-1 β mRNA in controls and oral cancer patients. c) ROC curves for combination of salivary IL-8 protein and H3F3A mRNA. The AUC value for cancer patients distinguishing from controls and PMOD patients was 0.75 and 0.75, respectively. d) 2D scatterplot showing the correlation between the expression of salivary IL-8 protein and H3F3A mRNA in controls and oral cancer patients.

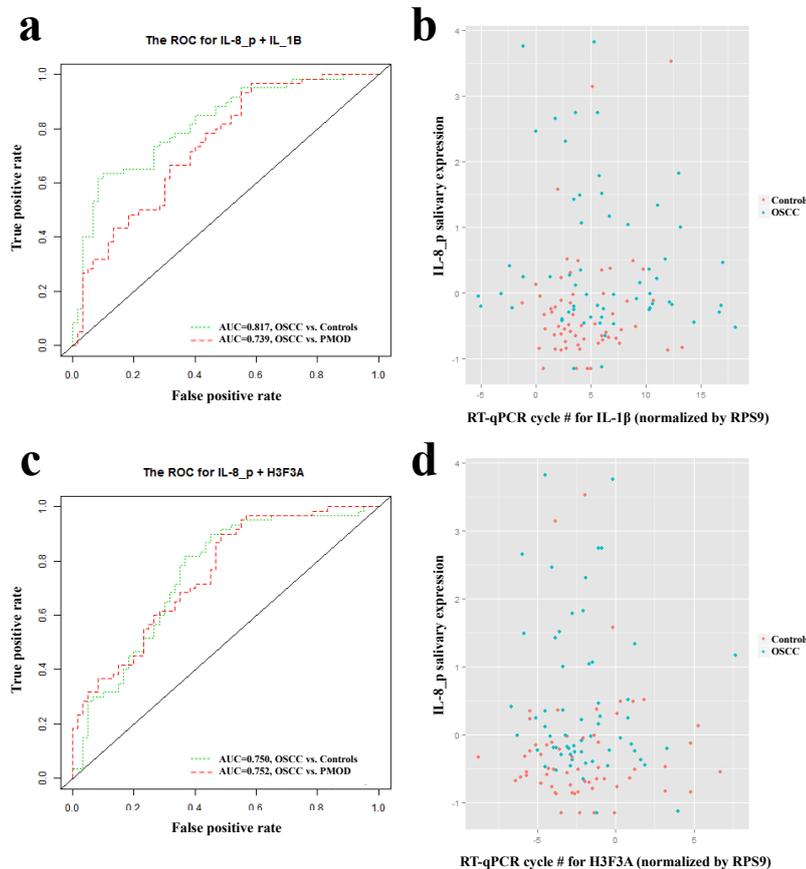
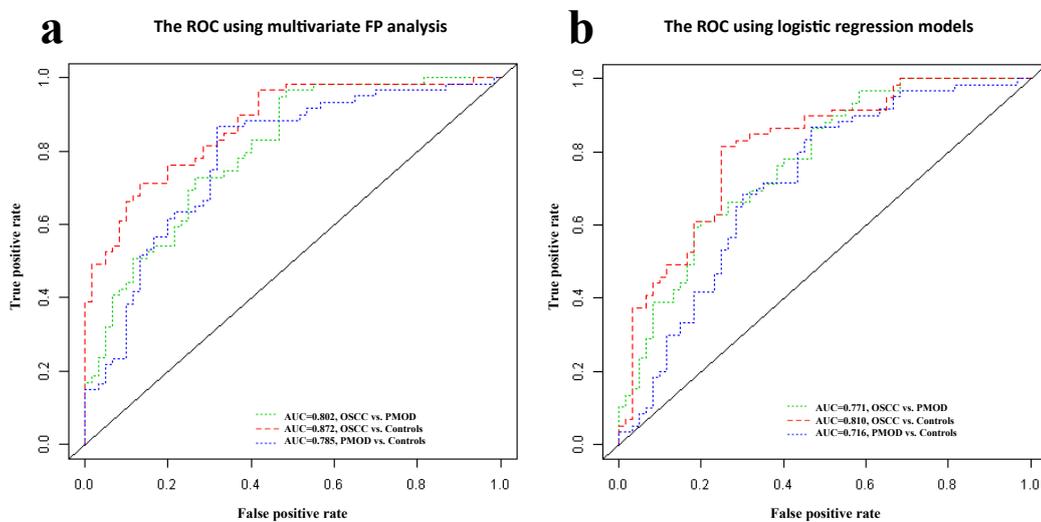


Fig.3 - a) ROC curves generated by multivariate fractional polynomial analysis using three different combinations of variables. Blue line: IL-1 β + Areca Nut + Drinking + Smoking for differentiation between PMOD and controls (AUC = 0.802). Red line: DUSP1 + IL-1 β p + IL-8p + Areca Nut for differentiation between OSCC and controls (AUC = 0.872). Green line: H3F3A + IL-8p + Smoking for differentiation between OSCC and PMOD (AUC = 0.785). b) ROC curves generated by logistic regression models using three different combinations of variables. Blue line: IL-1 β + Areca Nut + Drinking + Smoking for differentiation between PMOD and controls (AUC = 0.716). Red line: DUSP1 + IL-1 β p + IL-8p + Areca Nut for differentiation between OSCC and controls (AUC = 0.810). Green line: H3F3A + IL-8p + Smoking for differentiation between OSCC and PMOD (AUC = 0.771).



Manuscript references

1. Carvalho AL, Nishimoto IN, Califano J a, Kowalski LP. Trends in incidence and prognosis for head and neck cancer in the United States: a site-specific analysis of the SEER database. *Int J Cancer* 2005;114:806-16.
2. Van Harten MC, de Ridder M, Hamming-Vrieze O, Smeele LE, Balm AJM, van den Brekel MWM. The association of treatment delay and prognosis in head and neck squamous cell carcinoma (HNSCC) patients in a Dutch comprehensive cancer center. *Oral Oncol* 2014;50:282-90.
3. American Cancer Society. *Cancer Facts & Figures*. Cancer Facts Fig. Atlanta: American Cancer Society; 2012. page 68.
4. Califano J, van der Riet P, Westra W, Nawroz H, Clayman G, Piantadosi S, et al. Genetic progression model for head and neck cancer: implications for field cancerization. *Cancer Res* 1996;56:2488-92.
5. Leemans CR, Braakhuis BJM, Brakenhoff RH. The molecular biology of head and neck cancer. *Nat Rev Cancer* 2011;11:9-22.
6. Braakhuis BJM, Tabor MP, Kummer JA, Leemans CR, Brakenhoff RH. A genetic explanation of Slaughter's concept of field cancerization: evidence and clinical implications. *Cancer Res* 2003;63:1727-30.
7. Napier SS, Speight PM. Natural history of potentially malignant oral lesions and conditions: an overview of the literature. *J Oral Pathol Med* 2008;37:1-10.
8. Moyer VA. Screening for oral cancer: U.S. Preventive Services Task Force recommendation statement. *Ann Intern Med* 2014;160:55-60.
9. Neville BW, Day T a. Oral Cancer and Precancerous Lesions. *CA Cancer J Clin* 2002;52:195-215.
10. McGurk M, Chan C, Jones J, O'regan E, Sherriff M. Delay in diagnosis and its effect on outcome in head and neck cancer. *Br J Oral Maxillofac Surg.* 2005;43:281-4.

11. Park NJ, Zhou H, Elashoff D, Henson BS, Kastratovic D a, Abemayor E, et al. Salivary microRNA: discovery, characterization, and clinical utility for oral cancer detection. *Clin Cancer Res* 2009;15:5473-7.
12. Wei F, Lin C-C, Joon A, Feng Z, Troche G, Lira ME, et al. Non-Invasive Saliva-Based EGFR Gene Mutation Detection in Lung Cancer Patients. *Am J Respir Crit Care Med* 2014;1-52.
13. Arellano-Garcia ME, Hu S, Wang J, Henson B, Zhou H, Chia D, et al. Multiplexed immunobead-based assay for detection of oral cancer protein biomarkers in saliva. *Oral Dis* 2008;14:705-12.
14. Elashoff D, Zhou H, Reiss J, Wang J, Xiao H, Henson B, et al. Prevalidation of salivary biomarkers for oral cancer detection. *Cancer Epidemiol Biomarkers Prev* 2012;21:664-72.
15. Nagler RM. Saliva as a tool for oral cancer diagnosis and prognosis. *Oral Oncol* 2009;45:1006-10.
16. Li Y, John M, Zhou X, Kim Y. Salivary transcriptome diagnostics for oral cancer detection. *Clin Cancer Res* 2004;10:8442-50.
17. Bernardes VF, Gleber-Netto FO, Sousa SF, Silva TA, Abreu MHNG, Aguiar MCF. EGF in saliva and tumor samples of oral squamous cell carcinoma. *Appl Immunohistochem Mol Morphol* 2011;19:528-33.
18. Kim JK, Zhou H, Nabili V, Wang MB, Abemayor E, Wong DTW. Utility of multiple sampling in reducing variation of salivary interleukin-8 and interleukin-1 β mRNA levels in healthy adults. *Head Neck* 2013;35:968-73.
19. Brinkmann O, Kastratovic D a, Dimitrijevic M V, Konstantinovic VS, Jelovac DB, Antic J, et al. Oral squamous cell carcinoma detection by salivary biomarkers in a Serbian population. *Oral Oncol* 2011;47:51-5.
20. Cheng Y-SL, Rees T, Jordan L, Oxford L, O'Brien J, Chen H-S, et al. Salivary endothelin-1 potential for detecting oral cancer in patients with oral lichen planus or oral cancer in remission. *Oral Oncol* 2011;47:1122-6.

21. Hoffmann T, Sonkoly E. Aberrant cytokine expression in serum of patients with adenoid cystic carcinoma and squamous cell carcinoma of the head and neck. *Head Neck* 2007;29:472-8.
22. Korostoff A, Reder L, Masood R, Sinha UK. The role of salivary cytokine biomarkers in tongue cancer invasion and mortality. *Oral Oncol.* 2011;47:282-7.
23. Punyani SR, Sathawane RS. Salivary level of interleukin-8 in oral precancer and oral squamous cell carcinoma. *Clin Oral Investig* 2013;17:517-24.
24. Lee YH, Zhou H, Reiss JK, Yan X, Zhang L, Chia D, et al. Direct saliva transcriptome analysis. *Clin Chem.* 2011;57:1295-302.
25. Royston P, Altman DG. Regression Using Fractional Polynomials of Continuous Covariates: Parsimonious Parametric Modelling. *Appl Stat* 1994;43:429.
26. Sauerbrei W, Meier-Hirmer C, Benner A, Royston P. Multivariable regression model building by using fractional polynomials: Description of SAS, STATA and R programs. *Comput Stat Data Anal* 2006;50:3464-85.
27. Warnakulasuriya S. Global epidemiology of oral and oropharyngeal cancer. *Oral Oncol* 2009;45:309-16.
28. Johnson NW, Jayasekara P, Amarasinghe a a HK. Squamous cell carcinoma and precursor lesions of the oral cavity: epidemiology and aetiology. *Periodontol* 2000 2011;57:19-37.
29. Lee C-H, Ko Y-C, Huang H-L, Chao Y-Y, Tsai C-C, Shieh T-Y, et al. The precancer risk of betel quid chewing, tobacco use and alcohol consumption in oral leukoplakia and oral submucous fibrosis in southern Taiwan. *Br J Cancer.* 2003;88:366-72.
30. Chapple ILC. Time to take periodontitis seriously. *BMJ* 2014;348:g2645-g2645.
31. Batchelor P. Is periodontal disease a public health problem? *Br Dent J* 2014;217:405-9.

32. Dye B. Global periodontal disease epidemiology. *Periodontol 2000* 2012;58:10-25.
33. Petersen P, Ogawa H. The global burden of periodontal disease: towards integration with chronic disease prevention and control. *Periodontol 2000* 2012;60:15-39.
34. Lisa Cheng Y-S, Jordan L, Gorugantula LM, Schneiderman E, Chen H-S, Rees T. Salivary interleukin-6 and -8 in patients with oral cancer and patients with chronic oral inflammatory diseases. *J Periodontol* 2014;85:956-65.
35. Wei F, Yang J, Wong DTW. Detection of exosomal biomarker by electric field-induced release and measurement (EFIRM). *Biosens Bioelectron.* 2013;44:115-21.

8 REFERÊNCIAS BIBLIOGRÁFICAS

[ACS] American Cancer Society. **Cancer facts & figures**. Available form: <URL:<http://www.cancer.org/acs/groups/content/@editorial/documents/document/acspc-044552.pdf>> [2015 nov 12]

Alkureishi LW, Burak Z, Alvarez JA, et al. Joint practice guidelines for radionuclide lymphoscintigraphy for sentinel node localization in oral/oropharyngeal squamous cell carcinoma. **Ann Surg Oncol** 2009; 16:3190-210.

Anders S, Pyl PT, Huber W. HTSeq-a Python framework to work with high-throughput sequencing data. **Bioinformatics** 2014; 31:166-9.

Arellano-Garcia ME, Hu S, Wang J, et al. Multiplexed immunobead-based assay for detection of oral cancer protein biomarkers in saliva. **Oral Dis** 2008; 14:705-12.

Bagan J, Sarrion G, Jimenez Y. Oral cancer: clinical features. **Oral Oncol** 2010; 46:414-7.

Bagnardi V, Rota M, Botteri E, et al. Alcohol consumption and site-specific cancer risk: a comprehensive dose-response meta-analysis. **Br J Cancer** 2015; 112:580-93.

Batchelor P. Is periodontal disease a public health problem? **Br Dent J** 2014; 217:405-9.

Belcher R, Hayes K, Fedewa S, Chen AY. Current treatment of head and neck squamous cell cancer. **J Surg Oncol** 2014; 110:551-74.

Bernardes VF, Gleber-Netto FO, Sousa SF, Silva TA, Abreu MHNG, Aguiar MCF. EGF in saliva and tumor samples of oral squamous cell carcinoma. **Appl Immunohistochem Mol Morphol** 2011; 19:528-33.

Boccardo E. HPV-mediated genome instability: at the roots of cervical carcinogenesis. **Cytogenet Genome Res** 2010; 128:57-65.

Braakhuis BJM, Brakenhoff RH, Leemans CR. Second field tumors: a new opportunity for cancer prevention? **Oncologist** 2005a; 10:493-500.

Braakhuis BJM, Leemans CR, Brakenhoff RH. Expanding fields of genetically altered cells in head and neck squamous carcinogenesis. **Semin Cancer Biol** 2005b; 15:113-20.

Braakhuis BJM, Tabor MP, Kummer JA, Leemans CR, Brakenhoff RH. A genetic explanation of Slaughter's concept of field cancerization: evidence and clinical implications. **Cancer Res** 2003; 63:1727-30.

Brackenbury WJ. Voltage-gated sodium channels and metastatic disease. **Channels (Austin)** 2012; 6:352-61.

Bradley PJ, Ferlito A, Silver CE, et al. Neck treatment and shoulder morbidity: still a challenge. **Head Neck** 2011; 33:1060-7.

Brinkmann O, Kastratovic D A, Dimitrijevic M V, et al. Oral squamous cell carcinoma detection by salivary biomarkers in a Serbian population. **Oral Oncol** 2011; 47:51-5.

Califano J, Van der Riet P, Westra W, et al. Genetic progression model for head and neck cancer: implications for field cancerization. **Cancer Res** 1996; 56:2488-92.

Carvalho AL, Nishimoto IN, Califano J A, Kowalski LP. Trends in incidence and prognosis for head and neck cancer in the United States: a site-specific analysis of the SEER database. **Int J Cancer** 2005; 114:806-16.

Carvalho MC DA CG DE, Silva DCG DA. Sequenciamento de DNA de nova geração e suas aplicações na genômica de plantas. **Ciência Rural** 2010; 40:735-744.

Chapple ILC. Time to take periodontitis seriously. **BMJ** 2014; 348:g2645-g2645.

Cheng Y-SL, Rees T, Jordan L, et al. Salivary endothelin-1 potential for detecting oral cancer in patients with oral lichen planus or oral cancer in remission. **Oral Oncol** 2011; 47:1122-6.

Chiang W-C, Wong Y-K, Lin S-C, Chang K-W, Liu C-J. Increase of MMP-13 expression in multi-stage oral carcinogenesis and epigallocatechin-3-gallate suppress MMP-13 expression. **Oral Dis** 2006; 12:27-33.

Chung CH, Parker JS, Karaca G, et al. Molecular classification of head and neck squamous cell carcinomas using patterns of gene expression. **Cancer Cell** 2004; 5:489-500.

D'Cruz AK, Vaish R, Kapre N, et al. Elective versus Therapeutic Neck Dissection in Node-Negative Oral Cancer. **N Engl J Med** 2015; 1-9.

Da Silva SD, Marchi FA, Xu B, et al. Predominant Rab-GTPase amplicons contributing to oral squamous cell carcinoma progression to metastasis. **Oncotarget** 2015; 12:

De Bree R, Takes RP, Castelijns JA, et al. Advances in diagnostic modalities to detect occult lymph node metastases in head and neck squamous cell carcinoma. **Head Neck** 2014; 1-11.

De Camargo Cancela M, Voti L, Guerra-Yi M, Chapuis F, Mazuir M, Curado MP. Oral cavity cancer in developed and in developing countries: population-based incidence. **Head Neck** 2010; 32:357-67.

Del Mónaco SM, Marino GI, Assef Y A, Damiano AE, Kotsias B A. Cell migration in BeWo cells and the role of epithelial sodium channels. **J Membr Biol** 2009; 232:1-13.

Dionne KR, Warnakulasuriya S, Zain RB, Cheong SC. Potentially malignant disorders of the oral cavity: current practice and future directions in the clinic and laboratory. **Int J Cancer** 2015; 136:503-15.

Döbrossy L. Epidemiology of head and neck cancer: magnitude of the problem. **Cancer Metastasis Rev** 2005; 24:9-17.

DuPont JJ, Hill MA, Bender SB, Jaisser F, Jaffe IZ. Aldosterone and vascular mineralocorticoid receptors: regulators of ion channels beyond the kidney. **Hypertension** 2014; 63:632-7.

Dye B. Global periodontal disease epidemiology. **Periodontol 2000** 2012; 58:10-25.

Elashoff D, Zhou H, Reiss J, et al. Prevalidation of salivary biomarkers for oral cancer detection. **Cancer Epidemiol Biomarkers Prev** 2012; 21:664-72.

Erler JT, Bennewith KL, Nicolau M, et al. Lysyl oxidase is essential for hypoxia-induced metastasis. **Nature** 2006; 440:1222-6.

Fiske JL, Fomin VP, Brown ML, Duncan RL, Sikes R A. Voltage-sensitive ion channels and cancer. **Cancer Metastasis Rev** 2006; 25:493-500.

Franco EL, Kowalski LP, Oliveira B V, et al. Risk factors for oral cancer in Brazil: a case-control study. **Int J Cancer** 1989; 43:992-1000.

Freier K, Joos S, Flechtenmacher C, et al. Tissue microarray analysis reveals site-specific prevalence of oncogene amplifications in head and neck squamous cell carcinoma. **Cancer Res** 2003; 63:1179-82.

Fuller PJ. Novel interactions of the mineralocorticoid receptor. **Mol Cell Endocrinol** 2015; 408:33-7.

Gilkes DM, Semenza GL, Wirtz D. Hypoxia and the extracellular matrix: drivers of tumour metastasis. **Nat Rev Cancer** 2014; 14:430-9.

Hanahan D, Weinberg R A. Hallmarks of cancer: The next generation. **Cell** 2011; 144:646-674.

Hensen EF, De Herdt MJ, Goeman JJ, et al. Gene-expression of metastasized versus non-metastasized primary head and neck squamous cell carcinomas: a pathway-based analysis. **BMC Cancer** 2008; 8:168.

Hoffmann TK, Sonkoly E, Homey B, et al. Aberrant cytokine expression in serum of patients with adenoid cystic carcinoma and squamous cell carcinoma of the head and neck. **Head Neck** 2007; 29:472-8.

Homann N, Tillonen J, Rintamäki H, Salaspuro M, Lindqvist C, Meurman JH. Poor dental status increases acetaldehyde production from ethanol in saliva: A possible link to increased oral cancer risk among heavy drinkers. **Oral Oncol** 2001; 37:153-8.

Huang L, Zheng M, Zhou Q-M, et al. Identification of a gene-expression signature for predicting lymph node metastasis in patients with early stage cervical carcinoma. **Cancer** 2011; 117:3363-73.

Huang Q, Yu GP, McCormick S A, et al. Genetic differences detected by comparative genomic hybridization in head and neck squamous cell carcinomas from different tumor sites: construction of oncogenetic trees for tumor progression. **Genes Chromosomes Cancer** 2002; 34:224-33.

Hübbers CU, Akgül B. HPV and cancer of the oral cavity. **Virulence** 2015; 6:244-8.

Irié T, Aida T, Tachikawa T. Gene expression profiling of oral squamous cell carcinoma using laser microdissection and cDNA microarray. **Med Electron Microsc** 2004; 37:89-96.

Johnson NW, Jayasekara P, Amarasinghe A A HK. Squamous cell carcinoma and precursor lesions of the oral cavity: epidemiology and aetiology. **Periodontol 2000** 2011; 57:19-37.

Kademani D. Oral cancer. **Mayo Clin Proc** 2007; 82:878-87.

Kang C-J, Chen Y-J, Liao C-T, et al. Transcriptome profiling and network pathway analysis of genes associated with invasive phenotype in oral cancer. **Cancer Lett** 2009; 284:131-40.

Karim-Kos HE, De Vries E, Soerjomataram I, Lemmens V, Siesling S, Coebergh JWW. Recent trends of cancer in Europe: a combined approach of incidence, survival and mortality for 17 cancer sites since the 1990s. **Eur J Cancer** 2008; 44:1345-89.

Kashiwazaki H, Hassan NMM, Hamada J-I, et al. Gene expression profile changes correlated with lymph node metastasis in oral squamous cell carcinoma. **Odontology** 2008; 96:38-43.

Kato Y, Uzawa K, Saito K, et al. Gene expression pattern in oral cancer cervical lymph node metastasis. **Oncol Rep** 2006; 16:1009-14.

Kim JK, Zhou H, Nabili V, Wang MB, Abemayor E, Wong DTW. Utility of multiple sampling in reducing variation of salivary interleukin-8 and interleukin-1 β mRNA levels in healthy adults. **Head Neck** 2013; 35:968-73.

Koo K, Barrowman R, McCullough M, Iseli T, Wiesenfeld D. Non-smoking non-drinking elderly females: a clinically distinct subgroup of oral squamous cell carcinoma patients. **Int J Oral Maxillofac Surg** 2013; 42:929-33.

Korostoff A, Reder L, Masood R, Sinha UK. The role of salivary cytokine biomarkers in tongue cancer invasion and mortality. **Oral Oncol** 2011; 47:282-287.

Lallemant B, Evrard A, Chambon G, et al. Gene expression profiling in head and neck squamous cell carcinoma: Clinical perspectives. **Head Neck** 2010; 32:1712-9.

Lambert R, Sauvaget C, De Camargo Cancela M, Sankaranarayanan R. Epidemiology of cancer from the oral cavity and oropharynx. **Eur J Gastroenterol Hepatol** 2011; 23:633-41.

Lawrence MS, Sougnez C, Lichtenstein L, et al. Comprehensive genomic characterization of head and neck squamous cell carcinomas. **Nature** 2015; 517:576-582.

Lee C-H, Ko Y-C, Huang H-L, et al. The precancer risk of betel quid chewing, tobacco use and alcohol consumption in oral leukoplakia and oral submucous fibrosis in southern Taiwan. **Br J Cancer** 2003; 88:366-372.

Lee YH, Zhou H, Reiss JK, et al. Direct saliva transcriptome analysis. **Clin Chem** 2011; 57:1295-1302.

Leemans CR, Braakhuis BJM, Brakenhoff RH. The molecular biology of head and neck cancer. **Nat Rev Cancer** 2011; 11:9-22.

Leemans CR, Tiwari R, Nauta JJP, Van der Waal I, Snow GB. Recurrence at the Primary Site in Head and Neck Cancer and the Significance of Neck Lymph Node Metastases as a Prognostic Factor. **Cancer** 1994; 73:187-190.

Leivonen S-K, Ala-Aho R, Koli K, Grénman R, Peltonen J, Kähäri V-M. Activation of Smad signaling enhances collagenase-3 (MMP-13) expression and invasion of head and neck squamous carcinoma cells. **Oncogene** 2006; 25:2588-600.

Leusink FKJ, Van Es RJJ, De Bree R, et al. Novel diagnostic modalities for assessment of the clinically node-negative neck in oral squamous-cell carcinoma. **Lancet Oncol** 2012; 13:e554-61.

Li Y, John M, Zhou X, Kim Y. Salivary transcriptome diagnostics for oral cancer detection. **Clin Cancer Res** 2004; 10:8442-8450.

Li Y, Zhang J, Hong S. ANO1 as a marker of oral squamous cell carcinoma and silencing ANO1 suppresses migration of human SCC-25 cells. **Med Oral Patol Oral Cir Bucal** 2014; 19:e313-9.

Lingen MW, Xiao W, Schmitt A, et al. Low etiologic fraction for high-risk human papillomavirus in oral cavity squamous cell carcinomas. **Oral Oncol** 2013; 49:1-8.

Lisa Cheng Y-S, Jordan L, Gorugantula LM, Schneiderman E, Chen H-S, Rees T. Salivary interleukin-6 and -8 in patients with oral cancer and patients with chronic oral inflammatory diseases. **J Periodontol** 2014; 85:956-65.

Love MI, Huber W, Anders S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. **Genome Biol** 2014; 15:550.

Martinez EZ, Louzada-Neto F, Pereira BB. Analysis of diagnostic tests using ROC curves. **Cad Saúde Coletiva** 2003; 11:7-31.

Matsuo K, Kawano K. Immunohistochemical distribution and morphometric analysis of aquaporin-3 in oral squamous cell carcinoma. **Int J Oral Maxillofac Surg** 2014; 43:13-21.

McGurk M, Chan C, Jones J, O'regan E, Sherriff M. Delay in diagnosis and its effect on outcome in head and neck cancer. **Br J Oral Maxillofac Surg** 2005; 43:281-4.

Méndez E, Fan W, Choi P, et al. Tumor-specific genetic expression profile of metastatic oral squamous cell carcinoma. **Head Neck** 2007; 29:803-14.

Ministério da Saúde. Instituto Nacional de Câncer José Alencar Gomes da Silva. **Estimativa 2014: incidência de câncer no Brasil**. Rio de Janeiro: INCA; 2014.

Montero PH, Yu C, Palmer FL, et al. Nomograms for preoperative prediction of prognosis in patients with oral cavity squamous cell carcinoma. **Cancer** 2014; 120:214-21.

Moyer VA. Screening for oral cancer: U.S. Preventive Services Task Force recommendation statement. **Ann Intern Med** 2014; 160:55-60.

Nagata M, Fujita H, Ida H, et al. Identification of potential biomarkers of lymph node metastasis in oral squamous cell carcinoma by cDNA microarray analysis. **Int J Cancer** 2003; 106:683-9.

Nagler RM. Saliva as a tool for oral cancer diagnosis and prognosis. **Oral Oncol** 2009; 45:1006-10.

Napier SS, Speight PM. Natural history of potentially malignant oral lesions and conditions: an overview of the literature. **J Oral Pathol Med** 2008; 37:1-10.

Neville BW, Damm DD, Allen CM, Bouquot JE. **Oral and maxillofacial pathology**. 3rd ed. Philadelphia, W. B. Saunders, 2009. Epithelial pathology; p.410-23.

Neville BW, Day T A. Oral Cancer and Precancerous Lesions. **CA Cancer J Clin** 2002; 52:195-215.

Nguyen ST, Hasegawa S, Tsuda H, et al. Identification of a predictive gene expression signature of cervical lymph node metastasis in oral squamous cell carcinoma. **Cancer Sci** 2007; 98:740-6.

O'Donnell RK, Kupferman M, Wei SJ, et al. Gene expression signature predicts lymphatic metastasis in squamous cell carcinoma of the oral cavity. **Oncogene** 2005; 24:1244-51.

Oliveira LR, Ribeiro-Silva A, Zucoloto S. Perfil da incidência e da sobrevida de pacientes com carcinoma epidermóide oral em uma população brasileira. **J Bras Patol Med Lab** 2006; 42:385-392.

Ow TJ, Myers JN. Current management of advanced resectable oral cavity squamous cell carcinoma. **Clin Exp Otorhinolaryngol** 2011; 4:1-10.

Palazzo AF, Lee ES. Non-coding RNA: what is functional and what is junk? **Front Genet** 2015; 6:1-11.

Park NJ, Zhou H, Elashoff D, et al. Salivary microRNA: discovery, characterization, and clinical utility for oral cancer detection. **Clin Cancer Res** 2009; 15:5473-7.

Patel SG, Shah JP. TNM staging of cancers of the head and neck: striving for uniformity among diversity. **CA Cancer J Clin** 2005; 55:242-58.

Pateromichelakis S, Farahani M, Phillips E, Partridge M. Molecular analysis of paired tumours: time to start treating the field. **Oral Oncol** 2005; 41:916-26.

Pedersen NJ, Jensen DH, Hedbäck N, et al. Staging of early lymph node metastases with the sentinel lymph node technique and predictive factors in T1/T2 oral cavity cancer: A retrospective single-center study. **Head Neck** 2015; 1-8.

Petersen P, Ogawa H. The global burden of periodontal disease: towards integration with chronic disease prevention and control. **Periodontol** 2000 2012; 60:15-39.

Pimenta Amaral TM, Da Silva Freire AR, Carvalho AL, Pinto CAL, Kowalski LP. Predictive factors of occult metastasis and prognosis of clinical stages I and II squamous cell carcinoma of the tongue and floor of the mouth. **Oral Oncol** 2004; 40:780-6.

Pires FR, Ramos AB, Oliveira JBC DE, Tavares AS, Luz PSR DA, Santos TCRB DOS. Oral squamous cell carcinoma: clinicopathological features from 346 cases from a single oral pathology service during an 8-year period. **J Appl Oral Sci** 2013; 21:460-7.

Po Wing Yuen A, Lam KY, Lam LK, et al. Prognostic factors of clinically stage I and II oral tongue carcinoma-A comparative study of stage, thickness, shape, growth pattern, invasive front malignancy grading, Martinez-Gimeno score, and pathologic features. **Head Neck** 2002; 24:513-20.

Prevarskaya N, Skryma R, Shuba Y. Calcium in tumour metastasis: new roles for known actors. **Nat Rev Cancer** 2011; 11:609-18.

Punyani SR, Sathawane RS. Salivary level of interleukin-8 in oral precancer and oral squamous cell carcinoma. **Clin Oral Investig** 2013; 17:517-24.

Radoï L, Luce D. A review of risk factors for oral cavity cancer: the importance of a standardized case definition. **Community Dent Oral Epidemiol** 2013; 41:97-109, e78-91.

Rampias T, Sasaki C, Weinberger P, Psyrris A. E6 and e7 gene silencing and transformed phenotype of human papillomavirus 16-positive oropharyngeal cancer cells. **J Natl Cancer Inst** 2009; 101:412-23.

Reis EM, Ojopi EPB, Alberto FL, et al. Large-scale transcriptome analyses reveal new genetic marker candidates of head, neck, and thyroid cancer. **Cancer Res** 2005; 65:1693-9.

Reuschenbach M, Kansy K, Garbe K, et al. Lack of evidence of human papillomavirus-induced squamous cell carcinomas of the oral cavity in southern Germany. **Oral Oncol** 2013; 49:937-942.

Rikardsen OG, Bjerkli I, Uhlin-Hansen L, Hadler-Olsen E, Steigen SE. Clinicopathological characteristics of oral squamous cell carcinoma in Northern Norway: a retrospective study. **BMC Oral Health** 2014; 14:103.

Roepman P, Kemmeren P, Wessels LF A, Slootweg PJ, Holstege FCP. Multiple robust signatures for detecting lymph node metastasis in head and neck cancer. **Cancer Res** 2006; 66:2361-6.

Roepman P, Wessels LF A, Kettelarij N, et al. An expression profile for diagnosis of lymph node metastases from primary head and neck squamous cell carcinomas. **Nat Genet** 2005; 37:182-186.

Royston P, Altman DG. Regression using fractional polynomials of continuous covariates: parsimonious parametric modelling. **Appl Stat** 1994; 43:429.

Saba NF, Goodman M, Ward K, et al. Gender and ethnic disparities in incidence and survival of squamous cell carcinoma of the oral tongue, base of tongue, and tonsils: a surveillance, epidemiology and end results program-based analysis. **Oncology** 2011; 81:12-20.

Sano D, Myers JN. Metastasis of squamous cell carcinoma of the oral tongue. **Cancer Metastasis Rev** 2007; 26:645-62.

Sauerbrei W, Meier-Hirmer C, Benner A, Royston P. Multivariable regression model building by using fractional polynomials: Description of SAS, STATA and R programs. **Comput Stat Data Anal** 2006; 50:3464-3485.

Schegg B, Hülsmeyer AJ, Rutschmann C, Maag C, Hennet T. Core glycosylation of collagen is initiated by two beta(1-O)galactosyltransferases. **Mol Cell Biol** 2009; 29:943-52.

Schmalbach CE, Chepeha DB, Giordano TJ, et al. Molecular profiling and the identification of genes associated with metastatic oral cavity/pharynx squamous cell carcinoma. **Arch Otolaryngol Head Neck Surg** 2004; 130:295-302.

Schwab A, Fabian A, Hanley PJ, Stock C. Role of ion channels and transporters in cell migration. **Physiol Rev** 2012; 92:1865-913.

Shah JP, Gil Z. Current concepts in management of oral cancer--surgery. **Oral Oncol** 2009; 45:394-401.

Shrime MG, Bachar G, Lea J, et al. Nodal ratio as an independent predictor of survival in squamous cell carcinoma of the oral cavity. **Head Neck** 2009; 31:1482-8.

SLAUGHTER DP, SOUTHWICK HW, SMEJKAL W. Field cancerization in oral stratified squamous epithelium; clinical implications of multicentric origin. **Cancer** 1953; 6:963-8.

Sparano A, Weinstein G, Chalian A, Yodul M, Weber R. Multivariate predictors of occult neck metastasis in early oral tongue cancer. **Otolaryngol Head Neck Surg** 2004; 131:472-6.

Spiro RH, Huvos AG, Wong GY, Spiro JD, Gnecco CA, Strong EW. Predictive value of tumor thickness in squamous carcinoma confined to the tongue and floor of the mouth. **Am J Surg** 1986; 152:345-50.

Steeg PS. Tumor metastasis: mechanistic insights and clinical challenges. **Nat Med** 2006; 12:895-904.

Steiglitz BM, Keene DR, Greenspan DS. PCOLCE2 encodes a functional procollagen C-proteinase enhancer (PCPE2) that is a collagen-binding protein differing in distribution of expression and post-translational modification from the previously described PCPE1. **J Biol Chem** 2002; 277:49820-30.

Stockand JD. New ideas about aldosterone signaling in epithelia. **Am J Physiol Renal Physiol** 2002; 282:F559-76.

Sugahara K, Michikawa Y, Ishikawa K, et al. Combination effects of distinct cores in 11q13 amplification region on cervical lymph node metastasis of oral squamous cell carcinoma. **Int J Oncol** 2011; 39:761-9.

Tarazona S, García-Alcalde F, Dopazo J, Ferrer A, Conesa A. Differential expression in RNA-seq: a matter of depth. **Genome Res** 2011; 21:2213-23.

Thomas AM, Gleber-Netto FO, Fernandes GR, et al. Alcohol and tobacco consumption affects bacterial richness in oral cavity mucosa biofilms. **BMC Microbiol** 2014; 14:250.

Tommasino M. The human papillomavirus family and its role in carcinogenesis. **Semin Cancer Biol** 2014; 26:13-21.

Tucker T, Marra M, Friedman JM. Massively parallel sequencing: the next big thing in genetic medicine. **Am J Hum Genet** 2009; 85:142-54.

Uchida K, Oga A, Nakao M, et al. Loss of 3p26.3 is an independent prognostic factor in patients with oral squamous cell carcinoma. **Oncol Rep** 2011; 26:463-9.

Van Harten MC, De Ridder M, Hamming-Vrieze O, Smeele LE, Balm AJM, Van den Brekel MWM. The association of treatment delay and prognosis in head and neck squamous cell carcinoma (HNSCC) patients in a Dutch comprehensive cancer center. **Oral Oncol** 2014; 50:282-90.

Van Hooff SR, Leusink FKJ, Roepman P, et al. Validation of a gene expression signature for assessment of lymph node metastasis in oral squamous cell carcinoma. **J Clin Oncol** 2012; 30:4104-10.

Vincent-Chong VK, Salahshourifar I, Karen-Ng LP, et al. Overexpression of MMP13 is associated with clinical outcomes and poor prognosis in oral squamous cell carcinoma. **ScientificWorldJournal** 2014; 2014:897523.

Wang Z, Gerstein M, Snyder M. RNA-Seq: a revolutionary tool for transcriptomics. **Nat Rev Genet** 2009; 10:57-63.

Warnakulasuriya S. Global epidemiology of oral and oropharyngeal cancer. **Oral Oncol** 2009a; 45:309-16.

Warnakulasuriya S. Causes of oral cancer--an appraisal of controversies. **Br Dent J** 2009b; 207:471-475.

Warnakulasuriya S, Reibel J, Bouquot J, Dabelsteen E. Oral epithelial dysplasia classification systems: predictive value, utility, weaknesses and scope for improvement. **J Oral Pathol Med** 2008; 37:127-33.

Warner GC, Reis PP, Jurisica I, et al. Molecular classification of oral cancer by cDNA microarrays identifies overexpressed genes correlated with nodal metastasis. **Int J Cancer** 2004; 110:857-68.

Wei F, Lin C-C, Joon A, et al. Non-Invasive Saliva-Based EGFR Gene Mutation Detection in Lung Cancer Patients. **Am J Respir Crit Care Med** 2014; 1-52.

Wei F, Yang J, Wong DTW. Detection of exosomal biomarker by electric field-induced release and measurement (EFIRM). **Biosens Bioelectron** 2013; 44:115-121.

Weijers M, Leemans C-R, Aartman I-H, Karagozoglu K-H, Van-der-Waal I. Oral cancer trends in a single head-and-neck cancer center in the Netherlands; decline in T-stage at the time of admission. **Med Oral Patol Oral y Cir Bucal** 2011; 16:914-918.

Weise JB, Rudolph P, Heiser A, et al. LOXL4 is a selectively expressed candidate diagnostic antigen in head and neck cancer. **Eur J Cancer** 2008; 44:1323-31.

Woolgar JA. Pathology of the NO neck. **Br J Oral Maxillofac Surg** 1999; 37:205-9.

Wünsch-Filho V. The epidemiology of oral and pharynx cancer in Brazil. **Oral Oncol** 2002; 38:737-46.

Xia J, Broadhurst DI, Wilson M, Wishart DS. Translational biomarker discovery in clinical metabolomics: an introductory tutorial. **Metabolomics** 2013; 9:280-299.

Xie C, Mao X, Huang J, et al. KOBAS 2.0: a web server for annotation and identification of enriched pathways and diseases. **Nucleic Acids Res** 2011; 39:W316-22.

Yu T, Wu Y, Helman JI, Wen Y, Wang C, Li L. CXCR4 promotes oral squamous cell carcinoma migration and invasion through inducing expression of MMP-9 and MMP-13 via the ERK signaling pathway. **Mol Cancer Res** 2011; 9:161-72.

Zennaro MC, Souque A, Viengchareun S, Poisson E, Lombès M. A new human MR splice variant is a ligand-independent transactivator modulating corticosteroid action. **Mol Endocrinol** 2001; 15:1586-98.

Zhou X, Temam S, Oh M, et al. Global expression-based classification of lymph node metastasis and extracapsular spread of oral tongue squamous cell carcinoma. **Neoplasia** 2006; 8:925-32.

Anexo 1 - Carta de aprovação do Comitê de Ética em Pesquisa do A.C. Camargo Cancer Center

**Hospital
A.C. Camargo**
Centro de Tratamento, Ensino e Pesquisa em Câncer

**Comitê de Ética em
Pesquisa - CEP**

São Paulo, 12 de Dezembro de 2011.

Ao
Dr. Emmanuel Dias-Neto

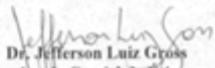
Ref.: Projeto de Pesquisa n°. 1621/11
"Estudo em larga-escala de alterações transcricionais relacionadas à ocorrência de metástases no câncer oral".

Os membros do Comitê de Ética em Pesquisa em Seres Humanos da Fundação Antonio Prudente – Hospital do Câncer - A.C. Camargo/SP, em sua última reunião de 29/11/2011, **aprovaram** a realização do estudo em referência e tomaram conhecimento dos seguintes documentos:

- Folha de Rosto para Pesquisa Envolvendo Seres Humanos;
- Termo de Compromisso do Pesquisador com as Resoluções do Conselho Nacional de Saúde;
- Termo de Dispensa do Consentimento Livre e Esclarecido;
- Declaração sobre os Dados Coletados, Publicação dos Dados e Propriedade das Informações Geradas;
- Declaração sobre o Uso e Destino do Material Biológico, Publicação dos Dados e Propriedade das Informações Geradas;
- Orçamento Financeiro Detalhado;
- Declaração de Infraestrutura e instalações do Laboratório de Genômica Médica - CIPE;
- Declaração de Ciência e Comprometimento do Biobanco do Hospital A.C. Camargo;
- Declaração sobre a Publicação dos Dados de Pesquisa Utilizando Amostras Fornecidas e Processadas pelo Biobanco do Hospital A.C. Camargo;
- Declaração de Ciência e Comprometimento da Faculdade de Medicina de São José do Rio Preto;
- Declaração de Ciência e Comprometimento do Departamento de Cirurgia de Cabeça e Pescoço e Otorrinolaringologia.

Informações a respeito do andamento do referido projeto deverão ser encaminhadas à assistente do CEP dentro de 12 meses.

Atenciosamente,


Dr. Jefferson Luiz Gross
1º Vice-Coordenador do Comitê de Ética em Pesquisa

1/1

Fundação Antonio Prudente - CNPJ/MF N. 60.961.968/0001-06
Rua Prof. Antônio Prudente, 211 - Liberdade - São Paulo, SP - 01509-900
Telefone: (11) 2189-5000
www.accamargo.org.br

Anexo 2 - Trabalhos Publicados Durante o Doutorado

Gleber-Netto FO, Braakhuis BJ, Triantafyllou A, Takes RP, Kelner N, Rodrigo JP, Strojan P, Vander Poorten V, Rapidis AD, Rinaldo A, Brakenhoff RH, Ferlito A, Kowalski LP. Molecular events in relapsed oral squamous cell carcinoma: Recurrence vs secondary primary tumor. *Oral Oncol.* 2015 Aug; 51(8):738-44. doi: 10.1016/j.oraloncology.2015.04.016. PMID: 25987307.

Thomas AM, **Gleber-Netto FO**, Fernandes GR, Amorim M, Barbosa LF, Francisco AL, de Andrade AG, Setubal JC, Kowalski LP, Nunes DN, Dias-Neto E. Alcohol and tobacco consumption affects bacterial richness in oral cavity mucosa biofilms. *BMC Microbiol.* 2014 Oct 3; 14:250. doi: 10.1186/s12866-014-0250-2. PMID: 25278091.

Bernardes VF, **Gleber-Netto FO**, Sousa SF, Rocha RM, Aguiar MC. EGFR status in oral squamous cell carcinoma: comparing immunohistochemistry, FISH and CISH detection in a case series study. *BMJ Open.* 2013 Jan 28; 3(1). pii: e002077. doi: 10.1136/bmjopen-2012-002077. PMID: 23358562

do Carmo MA, **Gleber-Netto FO**, Romano ML, Caldeira PC, de Aguiar MC. Clinical and demographic overlaps among immunologically mediated oral diseases: a challenge for clinicians. *Gen Dent.* 2014 Jan-Feb; 62(1):67-72. PMID: 24401354.