



UNIVERSIDADE FEDERAL DA BAHIA
FACULDADE DE ODONTOLOGIA
DOUTORADO EM ODONTOLOGIA E SAÚDE



DANIELA ADORNO FARIAS

**ANÁLISE DE ALTERAÇÕES GENÔMICAS POR MEIO DO
SEQUENCIAMENTO DO EXOMA COMPLETO EM UMA
SERIE DE CASOS DE DISPLASIAS EPITELIAIS ORAIS**

Salvador, BA – Brasil

2019

UNIVERSIDADE FEDERAL DA BAHIA
FACULDADE DE ODONTOLOGIA
DOUTORADO EM ODONTOLOGIA E SAÚDE

DANIELA ADORNO FARIAS

**ANÁLISE DE ALTERAÇÕES GENÔMICAS POR MEIO DO
SEQUENCIAMENTO DO EXOMA COMPLETO EM UMA
SERIE DE CASOS DE DISPLASIAS EPITELIAIS ORAIS**

Tese de Doutorado apresentada ao Programa de Pós-Graduação em Odontologia e Saúde, da Faculdade de Odontologia da Universidade Federal da Bahia, como parte dos requisitos para obtenção do título de Doutora em Odontologia.

Orientador: Prof. Dr. Jean Nunes Dos Santos

Coorientador: Prof. Dr. Ricardo Fernândes Ramires

Salvador, BA - Brasil
2019

Ficha catalográfica elaborada pelo Sistema Universitário de Bibliotecas (SIBI/UFBA), com os dados fornecidos pelo(a) autor(a).

Farias, Daniela Adorno

ANÁLISE DE ALTERAÇÕES GENÔMICAS POR MEIO DO SEQUENCIAMENTO DO EXOMA COMPLETO EM UMA SÉRIE DE CASOS DE DISPLASIAS EPITELIAIS ORAIS/ Daniela Adorno Farias. -- Salvador, 2019.
92 f.

Orientador: Prof. Dr. Jean Nunes dos Santos.
Coorientador: Prof. Dr. Ricardo Fernández Ramires.

Tese (Doutorado - Programa de Pós-Graduação em Odontologia e Saúde) -- Universidade Federal da Bahia, Faculdade de Odontologia, 2019.

1. Displasia epitelial oral. 2. Alterações genômicas. 3. Sequenciamento do exoma completo. 4. Sequenciamento de nova geração. I. dos Santos, Jean Nunes. II. Ramires, Ricardo Fernández.
III. Título.




UNIVERSIDADE FEDERAL DA BAHIA
FACULDADE DE ODONTOLOGIA
PROGRAMA DE PÓS-GRADUAÇÃO EM
ODONTOLOGIA E SAÚDE

TERMO DE APROVAÇÃO

C.D. DANIELA ADORNO FARIAS

**“ANÁLISE DE ALTERAÇÕES GENÔMICAS POR MEIO DO
SEQUENCIAMENTO DO EXOMA COMPLETO EM UMA SÉRIE DE
CASOS DE DISPLASIAS EPITELIAIS ORAIS”**


BANCA EXAMINADORA:



Prof. Dr. Jean Nunes dos Santos (Orientador)
Professor da Universidade Federal da Bahia – Faculdade de Odontologia



Prof. Dra. Flávia Caló de Aquino Xavier (Examinador Interno)
Professora da Universidade Federal da Bahia – Faculdade de Odontologia



Prof. Dra. Iguaracyra Barreto de Oliveira Araújo (Examinador Externo)
Professora da Universidade Federal da Bahia – Faculdade de Medicina



Prof. Dr. Atson Carlos de Souza Fernandes (Examinador Externo)
Professor da Escola Bahiana de Medicina e Saúde Pública



Prof. Dr. José Cláudio Casali da Rocha (Examinador Externo)
Hospital Erasto Gaetner, Curitiba - PR

DEDICATÓRIA

Dedico este trabalho aquelas pessoas que de forma espontânea e incondicional me entregaram sempre apoio e amor: minha mãe e meu pai (*em memória*).

AGRADECIMENTOS

Este trabalho de pesquisa é resultado de uma trajetória que comecei em 2015.2 e durante estes 4 anos nunca estive sozinha. Meu sincero agradecimento aos professores, orientadores e amigos: **Jean Nunes dos Santos e Ricardo Fernández Ramires.**

Prof. Jean, muito obrigada pela confiança, paciência e carinho que depositou em mim desde o começo. Concluo este trabalho com a esperança de levar um pouco da sua objetividade, competência e praticidade para tudo.

Prof. Ricardo, obrigada por fazer este caminho da “genética” mais agradável. Não poderia avançar sem poder contar com a sua genialidade e alegria em cada dia de trabalho. Acredito hoje ser um pouquinho “patogeneticista” graças a você.

Mãe, esta minha conquista é todinha sua! Obrigada pelo exemplo de ser uma pessoa íntegra, de fazer o bem aos outros e de nunca se cansar de buscar o crescimento pessoal e profissional.

Obrigada a meu **irmão Marcelo**, por estar preocupado e zelando para que eu esteja sempre bem. Estes pequenos cuidados são fundamentais em todo caminho que sigo. Finalmente, a meus sobrinhos, cunhada, tias e a toda minha família que deposita em mim tanto amor e confiança em tudo que faço. Sem vocês a minha vida seria muito mais complicada.

Aos meus amigos e colegas da pós-graduação: muito obrigada por todos os momentos de convivência, aprendizado e risadas. Obrigada **Ernesto, Ana Carol, Joaquim, Nêssa, Livia, Maíra, Getúlio e Rosângela!**

Tayla, obrigada pela amizade, confiança e ensinamentos constantes. Agradeço a **Sueli** pelas risadas matinais e todo apoio que sempre me deu. **Edilson e Mirian**, meu muito obrigada por todo apoio entregue.

Agradeço aos **Professores da Pós-graduação** que contribuíram para o meu crescimento e formação.

Agradecimento especial a **CAPES** pelo apoio financeiro durante o Doutorado e pela Bolsa Sanduiche que permitiu poder avançar no projeto de pesquisa.

Ao **FONDECYT** pelo financiamento para a realização deste projeto de pesquisa.

Aos **pacientes** que depositam sempre confiança e prestam todo apoio quando solicitamos que participem como voluntários em trabalhos de pesquisa.

Aos meus colegas, dirigentes e amigos da *Universidad de Chile*. Obrigada por todo apoio e confiança durante toda esta trajetória. Sem vocês seria impossível realizar este curso.

A **María José** por todo apoio profissional no laboratório com as amostras e pela amizade.
A **Aby** por ter ajudado com todas as fichas dos pacientes, com um sorriso sempre aberto.
O trabalho fica mais fácil quando realizamos com amigos!

E finalmente, obrigada aos **meus amigos**! Como poderia avançar nos meus projetos pessoais e profissionais se não tivesse vocês? Sou uma felizarda por estar morando fora do Brasil e poder encontrar amigos que são também “família”. De voltar a Salvador e encontrar as minhas amigas de sempre da Faculdade e do colégio. De poder viajar para outros países e me encontrar com amigos que moram longe, mas que sempre estão presentes. Tenho a certeza que o fechamento deste ciclo na minha vida também é momento de felicidade para vocês.

Deixo aqui registrado o meu Muito Obrigada a todos!

Adorno-Farias, Daniela. Análise de alterações genômicas por meio do sequenciamento do exoma completo em uma serie de casos de displasias epiteliais orais. [Tese de Doutorado]. Salvador: Faculdade de Odontologia da Universidade Federal da Bahia; 2019.

RESUMO

As lesões clínicas associadas com o desenvolvimento do carcinoma oral de células escamosas, neoplasia maligna oral de maior importância, são conhecidas como desordens orais potencialmente malignas e, em muitos casos, correspondem histopatologicamente a displasias epiteliais orais. Atualmente é considerado a divisão destas em baixo e alto grau para melhorar a concordância entre os patologistas. As bases genômicas das displasias epiteliais orais são desconhecidas e não existe um método confiável de avaliação de risco de transformação maligna. Entendendo que mutações somáticas são responsáveis pela transformação da mucosa displásica a câncer, e que estas variações genômicas poderiam representar marcadores objetivos para determinar potencial de transformação maligna, realizou-se sequenciamento do exoma completo em dez amostras de displasia epitelial oral de pacientes brasileiros e chilenos. Foram identificadas 42 variantes capazes de produzir alterações de alto impacto nas estruturas aminoacídicas de 39 genes. Deste total de variantes, seis são novas e foram identificadas nos genes: *FAM198B*, *CBWD5*, *PKD1L3*, *LRRC37A2*, *GAREM1* e *GIPC1*. Observou-se mais de uma variante de efeito deletéreo nos genes *TP53I11* e *CELAI*. As amostras de displasias epiteliais de alto grau mostraram uma tendência a acumular maior número de variantes de sequência de efeitos deletéreos que aquelas de baixo grau. Foram identificadas variantes exclusivas em amostras de displasias de baixo e alto graus. A variante SNV de efeito deletério localizada no gene *SIX1*, considerado um oncogene, foi observada em todas as amostras. Variantes de sequência nos genes *OR6C1*, *GABRG3*, *KRT24*, *MRPL27* e *LAMA5* foram identificados em todas as amostras de displasias de baixo grau, e nos genes: *ACTN2*, *TP53I11*, *GLYCTK* e *CELAI* foram observadas em todas as amostras de displasias de alto grau. Sendo assim, estes genes são sugeridos como possíveis marcadores de displasias epiteliais orais. É importante realizar novos estudos com um maior número de amostras que possam confirmar ou refutar as informações obtidas no presente estudo.

Palavras-chave: displasia epitelial oral; alterações genômicas; sequenciamento do exoma completo; sequenciamento de nova geração.

Adorno-Farias, Daniela. Genomic changes analysis in a series of cases of oral epithelial dysplasias by whole exome sequencing. [Tese de Doutorado]. Salvador: Faculdade de Odontologia da Universidade Federal da Bahia; 2019

ABSTRACT

Clinical conditions associated with the development of oral squamous cell carcinoma, the most important oral cancer, are known as oral potentially malignant disorders and, in many cases, correspond to oral epithelial dysplasias. Nowadays, the division of these in low and high grades is considered to improve agreement among pathologists. The genetic basis of oral epithelial dysplasias is unknown and there is no reliable method of evaluating malignant transformation risk. We understand that somatic mutations are responsible for the transformation of the dysplastic mucosa to cancer. In addition, these genomic variations could represent objective markers to determine potential for malign transformation. In the present study, We performed whole exome sequencing in ten samples of oral epithelial dysplasia of Brazilian and Chilean patients. We identified 42 variants able to produce high-impact alterations in the amino acid structures of 39 genes. Of these total variants, six are new and have been identified in the genes: *FAM198B*, *CBWD5*, *PKD1L3*, *LRRC37A2*, *GAREM1* and *GIPC1*. More than one deleterious effect variant was observed in the *TP53III1* and *CELA1* genes. High-grade epithelial dysplasia samples showed a tendency to accumulate more deleterious effect sequence variants than low-grade ones. Exclusive variants have been identified in low and high grade dysplasia samples. The deleterious effect variant SNV located on the *SIX1* gene, considered an oncogene, was observed in all samples. Sequence variants in the *OR6C1*, *GABRG3*, *KRT24*, *MRPL27* and *LAMA5* genes were identified in all low-grade dysplasia samples, and in the genes: *ACTN2*, *TP53III1*, *GLYCTK* and *CELA1* were observed in all high-grade dysplasia samples. Thus, these genes are suggested as possible markers of oral epithelial dysplasias. Further studies with more samples should be performed with to confirm or refute the information obtained in the present study.

Keywords: oral epithelial dysplasia; genomic alterations; whole exome sequencing; next-generation sequencing.

SUMÁRIO

| | | |
|----------|---|-----------|
| 1 | INTRODUÇÃO | 13 |
| 2 | REVISÃO DA LITERATURA | 16 |
| | 2.1 CÂNCER ORAL..... | 16 |
| | 2.2 DESORDENS ORAIS POTENCIALMENTE MALIGNAS..... | 17 |
| | 2.3 LEUCOPLASIA E ERITROPLASIA: ASPECTOS CLÍNICOS..... | 18 |
| | 2.4 DISPLASIA EPITELIAL ORAL..... | 20 |
| | 2.5 DOPM: MARCADORES MOLECULARES..... | 22 |
| | 2.6 SEQUENCIAMENTO DE NOVA GERAÇÃO EM DOPM..... | 26 |
| | 2.7 TRATAMENTO COCE E DOPM..... | 28 |
| 3 | OBJETIVOS | 29 |
| | 3.1 OBJETIVO GERAL..... | 29 |
| | 3.2 OBJETIVOS ESPECÍFICOS..... | 29 |
| 4 | METODOLOGIA | 30 |
| | 4.1 CONSIDERAÇÕES ÉTICAS..... | 30 |
| | 4.2 CARACTERIZAÇÃO DO ESTUDO E DAS AMOSTRAS..... | 30 |
| | 4.3 EXTRAÇÃO DE DNA GENÔMICO..... | 31 |
| | 4.4 SEQUENCIAMENTO DO EXOMA COMPLETO..... | 31 |
| | 4.4.1 PREPARAÇÃO DAS AMOSTRAS E CONTRUÇÃO DAS BIBLIOTECAS..... | 31 |
| | 4.4.2 SEQUENCIAMENTO E OBTENÇÃO DE DADOS..... | 32 |
| | 4.4.3 ANÁLISE BIOINFORMÁTICA DOS DADOS..... | 32 |
| | 4.5 ANÁLISE ESTATÍSTICA..... | 34 |
| 5 | RESULTADOS | 35 |
| 6 | DISCUSSÃO | 47 |
| 7 | CONCLUSÃO | 55 |
| | REFERÊNCIAS | 57 |
| | ANEXO A: Parecer do comitê de ética de pesquisa..... | 64 |
| | ANEXO B: Tabelas e Figuras suplementares..... | 66 |

LISTA DE ABREVIATURAS E SIGLAS

- AT: Conteúdo de adenina-timina.
- COCE: Carcinoma oral de células escamosas.
- DAG: displasia epitelial de alto grado.
- DBG: Displasia epitelial de bajo grado.
- DEO: Displasia epitelial oral.
- DNA: ácido desoxirribonucleico.
- dNTPs: Desoxirribonucleotídeos Fosfatados.
- DOPM: Desordens orais potencialmente malignas.
- GC: Conteúdo de guanina-citosina.
- HGS: Sequenciamento em larga escala (*High throughput sequencing*).
- HNC: Câncer de cabeça e pescoço (*Head and neck cancer*).
- HNSCC: Carcinoma de células escamosas de cabeça e pescoço (*Head and neck squamous cell carcinoma*).
- IARC: *International Agency for Research on Cancer*.
- INDEL: Variantes de sequência de inserção ou deleção.
- LOH: Perda de heterozigidade.
- LVP: Leucoplasia verrucosa proliferativa.
- miRNA: Também conhecido como microRNA. Pequenas moléculas de ácido ribonucleico.
- NGS: Sequenciamento de nova geração (*Next Generation sequencing*).
- OMS: Organização Mundial de Saúde.
- PCGA: Atlas do genoma do pré-câncer (*Pre-Cancer Genome Atlas*).
- PCR: Reação em cadeia da polimerase.
- Q20: Proporção de leituras que possuem pontuação de qualidade de *phred* acima de 20.
- Q30: Proporção de leituras que possuem pontuação de qualidade de *phred* acima de 30.
- RNA: Ácido ribonucleico.
- RTA: Análise em tempo real (*Real Time analysis*).
- SNP: Polimorfismo de nucleotídeo simples (*Single nucleotide polymorphism*).
- SNV: Variante de nucleotídeo simples (*Single nucleotide variant*).
- TCGA: Atlas do genoma do Câncer (*Cancer Genome Atlas*).
- VNC: variações do número de cópia.

WES: sequenciamento do exoma completo (*Whole Exome Sequence*).

1 INTRODUÇÃO

O Câncer oral é uma patologia que apresenta expressiva e crescente relevância internacional e distintos aspectos têm colaborado para tal condição: a) aumento da incidência; b) pelo encaminhamento de pacientes a tratamento em etapas tardias; c) pela morbidade significativa e d) pela taxa de sobrevivência inferior a 50% aos cinco anos. Particularmente, é dada maior atenção ao carcinoma oral de células escamosas (COCE), que representa aproximadamente 90% dos casos de tumores da cavidade bucal (DIONNE et al., 2015; EL-NAGGAR et al., 2017), e é um dos cânceres mais comuns no mundo entre homens, somando mais de 300.000 novos casos anualmente (WARNAKULASURIYA, 2009; FERLAY et al., 2015). Essa tendência negativa desta patologia é atribuída a vários fatores, incluindo o estágio tardio do diagnóstico, a cancerização em campo e a agressividade biológica inerente (propensão para o crescimento invasivo e disseminação metastática dos gânglios linfáticos) (NIKITAKIS et al., 2018a).

Sendo um câncer de fácil acesso clínico ao diagnóstico, é importante desenvolver estratégias para o diagnóstico precoce e preventivo, melhorando assim, o panorama atual. As lesões que precedem o COCE, são classificadas como desordens orais potencialmente malignas (DOPM) (EL-NAGGAR et al., 2017) que representam alterações morfológicas específicas, nas quais algumas podem ter um maior potencial de sofrer transformação maligna e também podem indicar um maior risco de desenvolvimento de câncer em qualquer lugar da mucosa oral aparentemente normal (WARNAKULASURIYA; JOHNSON; VAN DER WAAL, 2007; DIONNE et al., 2015; EL-NAGGAR et al., 2017). As DOPM que mais frequentemente estão associadas a transformação maligna na mucosa oral são: leucoplasia e eritroplasia, que em muitos casos correspondem histologicamente a displasia epitelial oral (DEO), nos seus diferentes graus (DIONNE et al., 2015; MELLO et al., 2018). O diagnóstico das displasias epiteliais é subjetivo, o que dificulta a avaliação destas em relação ao risco de transformação maligna (KUJAN et al., 2007; WARNAKULASURIYA et al., 2008; FLESKENS; SLOOTWEG, 2009 ; DIONNE et al., 2015). Para minimizar esta subjetividade é sugerido a utilização do sistema binário que divide estas em baixo e alto grau (KUJAN et al., 2006; EL-NAGGAR et al., 2017).

Não existem parâmetros clínicos objetivos que podem ser associados para

solucionar esta limitação da avaliação histológica na determinação do risco de malignidade. Certas características das lesões como: tamanho maior a 200mm³, textura não homogênea, coloração vermelha ou mista (branco-vermelha), localização em língua ou assoalho bucal, indicam associação com maior risco de transformação maligna. No entanto, a biopsia segue sendo a indicação padrão ouro para avaliar a presença de displasia epitelial, que quando de alto grau também indica alto risco de transformação para COCE. (SPEIGHT; KHURRAM; KUJAN, 2018)

A eliminação cirúrgica e outros tratamentos não invasivos em lesões que apresentam alterações displásicas não eliminam o risco de transformação maligna, e todos os estudos que visam detecção de marcadores moleculares, para ajudar a melhorar os problemas expostos realizados até o momento, não são aplicáveis na clínica (DIONNE et al., 2015; NIKITAKIS et al., 2018b). Vários desafios existem, incluindo dificuldades contínuas no diagnóstico, prognóstico e tratamento dos DOPMs (NIKITAKIS et al., 2018a).

Evidências na literatura indicam que a transição do epitélio normal para DEO e a COCE é o resultado de um acúmulo de alterações genéticas e epigenéticas (LINGEN et al., 2011; GRAVELAND et al., 2013; DIONNE et al., 2015). Esta complexa relação do ponto de vista molecular ainda não foi esclarecida na literatura atual, o que possivelmente explica muitos dos fracassos expostos relacionados com estas patologias.

Com muitas perguntas ainda em aberto em relação a estas temáticas, entendemos que buscar novas estratégias metodológicas atuais, como é o caso do sequenciamento de nova geração, que possibilita um sequenciamento do DNA em larga escala (GOODWIN; MCPHERSON; MCCOMBIE, 2016) e que já vem sendo usada em vários tipos de cânceres, incluindo o COCE (STRANSKY et al., 2012; ER et al., 2015; LAWRENCE et al., 2015; AL-HEBSHI et al., 2016), podem ser úteis para as DEOs. Esperamos encontrar que nas DEOs existam mutações somáticas que caracterizam esta etapa de possível transformação maligna.

Sabendo da escassez de publicações com uso do sequenciamento em larga escala em DOPM e ausência do uso desta tecnologia na população latinoamericana, realizamos o presente estudo que teve como objetivo identificar alterações genômicas, a partir da realização do sequenciamento do exoma completo, em dez amostras de DEO de baixo e alto graus, de pacientes com diagnóstico clínico de leucoplasia e/ou eritroplasia. O conhecimento de variações no DNA destas amostras poderia indicar novos genes

associados a transformação maligna e novos alvos terapêuticos para lesões com DEO. Sendo uma metodologia cara e de complexa interpretação, a realização de um trabalho colaborativo viabiliza integrar tecnologia, diferentes conhecimentos e experiências para dar respostas a pacientes chilenos e brasileiros.

2 REVISÃO DE LITERATURA

2.1 CÂNCER ORAL

De acordo as estimativas da *International Agency for Research on Cancer (IARC)*, em 2018 existirão 18.1 milhões de novos casos de câncer no mundo e um total de 9.6 milhões de mortes por esta causa (BRAY et al., 2018; FERLAY et al., 2018a). Deste total de novos casos, o câncer oral e de lábio representa aproximadamente 2% dos casos de morte estimados para 2018 (BRAY et al., 2018).

O carcinoma de células escamosas é a neoplasia maligna mais frequente da cavidade oral e representa um dos tumores malignos mais comuns entre homens no mundo (EL-NAGGAR et al., 2017). Esta patologia tem sido reconhecida como uma grande ameaça para a saúde pública, com diagnósticos tardios, tratamentos invasivos e alta taxa de mortalidade antes dos cinco anos (SCULLY; BAGAN, 2009; WARNAKULASURIYA, 2009). Apresenta uma distribuição geográfica variável nas diferentes regiões do mundo, tendo uma alta concentração de casos no sul da Ásia (BRAY et al., 2018). Em países de alto risco, como Sri Lanka, Índia, Paquistão e Bangladesh, o câncer bucal é o câncer mais comum em homens, representando até 25% dos novos casos (WARNAKULASURIYA, 2009).

Na América latina, o Brasil tem a taxa mais alta de incidência de câncer oral (WARNAKULASURIYA, 2009; FERLAY et al., 2018b) e a sua população masculina apresenta o terceiro mais alto risco no mundo para apresentar câncer de boca (WARNAKULASURIYA, 2009). No Brasil, o câncer de boca representa o quinto e 12º posições em frequência entre todos as neoplasias malignas, em homens e mulheres, respectivamente (INCA, 2017). O Chile, não classificado como país de alto risco para câncer oral (WARNAKULASURIYA, 2009), apresenta uma taxa de morbidade de 1,6% entre todos os tipos de câncer (RIERA; MARTÍNEZ, 2005). Este tipo de câncer é duas a quatro vezes mais frequentes em homens que mulheres na maioria dos grupos étnicos (SCULLY; BAGAN, 2009; WARNAKULASURIYA, 2009). Muitos estudos relatam que o maior número de casos desta patologia tem origem depois dos 45 ou 50 anos de idade e identificam vários fatores de associação com o aparecimento do carcinoma oral de células escamosas (COCE), no entanto, hábito de fumar cigarro e

consumo excessivo de álcool são os principais fatores descritos (SCULLY; BAGAN, 2009; WARNAKULASURIYA, 2009; EL-NAGGAR et al., 2017).

Entre as localizações na cavidade bucal, as regiões de língua e assoalho bucal são consideradas as mais frequentes de todos os casos de COCE (SCULLY, 2014; EL-NAGGAR et al., 2017). Pela sua localização é considerado um câncer de fácil acesso para inspeção e diagnóstico, mas geralmente os pacientes são encaminhados para avaliação e tratamento em estágios avançados (WARNAKULASURIYA, 2009; EL-NAGGAR et al., 2017). Apesar das inúmeras pesquisas que abordam a patogênese e o tratamento dos COCEs, o índice de sobrevivência aos cinco anos não tem aumentado muito nas últimas décadas, permanecendo próxima aos 50% (WARNAKULASURIYA; JOHNSON; VAN DER WAAL, 2007; NIKITAKIS et al., 2018a).

Ainda que o COCE possa se originar *de novo*, aceita-se que as DOPM representam um aumento do risco para o desenvolvimento desta neoplasia (WARNAKULASURIYA, 2009, 2018; MELLO et al., 2018). Sendo assim, estas serão comentadas a seguir.

2.2 DESORDENS ORAIS POTENCIALMENTE MALIGNAS

As DOPM são alterações clínicas que acarretam risco para o desenvolvimento de câncer na cavidade oral, seja em lesões precursoras identificadas clinicamente ou na mucosa oral clinicamente normal (EL-NAGGAR et al., 2017). Esta terminologia, introduzida na literatura a partir de uma reunião coordenada pela OMS em 2005 (WARNAKULASURIYA; JOHNSON; VAN DER WAAL, 2007) e publicada oficialmente por esta organização em 2017, substituiu as terminologias condições pré-cancerosas e lesões pré-cancerosas (KRAMER et al., 1978; WARNAKULASURIYA; JOHNSON; VAN DER WAAL, 2007), porque não existe a certeza que todas as lesões pré-cancerosas eventualmente se transformem em COCE (WARNAKULASURIYA, 2018). De fato, até o momento o motivo pelo qual estas lesões podem acarretar um risco para a transformação do câncer não é conhecido, porém entender este processo é muito importante para o conhecimento da prevenção do câncer. Recentemente foi proposto a terminologia “lesões epiteliais orais potencialmente pré-malignas”, que inclui o conceito daquelas lesões que tem um potencial para se tornar malignas, mas que mesmo antes de uma transformação maligna, permanecem sendo pré-malignas

(WARNAKULASURIYA, 2018). No entanto, esta última terminologia ainda não foi incorporada pela OMS (EL-NAGGAR et al., 2017).

As DOPM incluem: eritroplasia, leucoplasia, fibrose oral submucosa, disqueratose congênita, queratose em usuários de tabaco não fumado, lesões palatinas associadas ao hábito de fumar invertido, candidíase crônica, líquen plano, lúpus eritematoso discoide, glossite sifilítica e queilite actínica (EL-NAGGAR et al., 2017). Deste grupo, as lesões frequentemente associadas ao desenvolvimento do câncer oral são leucoplasia e eritroplasia em países ocidentais, sendo que a primeira é a mais encontrada na prática clínica (DIONNE et al., 2015; MELLO et al., 2018; WARNAKULASURIYA, 2018). A incidência e a prevalência das DOPM não foram descritas amplamente, pois diferenças regionais de estilo de vida e condições naturais podem influenciar a prevalência destas lesões (DIONNE et al., 2015; MELLO et al., 2018). Um recente estudo de revisão sistemática e meta-análise reporta uma prevalência global para DOPM de 4.47%, identificada mais comumente em homens (60%), sendo as populações da Ásia, América do sul e Caribe as mais afetadas (MELLO et al., 2018).

Muitos agentes etiológicos tem sido relacionados ao desenvolvimento das DOPM, sendo o consumo de tabaco fumado e álcool os principais fatores associados (WARNAKULASURIYA, 2009). O uso de tabaco mastigável ou tabaco em combinações com outras substâncias como *paan* ou *betel quid* (folha de *betel*, noz de areca, limão e tabaco) representam um risco maior para o desenvolvimento do câncer oral (STEWART; WILD, 2014). A infecção por vírus, principalmente o papiloma vírus, também é reportada e a exposição solar é um fator importante para o desenvolvimento do câncer de lábio (WARNAKULASURIYA, 2009).

2.3 LEUCOPLASIA E ERITROPLASIA: ASPECTOS CLÍNICOS

A leucoplasia é uma placa de coloração predominantemente branca com risco questionável, que exclui outras doenças ou desordens conhecidas que não acarretam risco para o câncer (WARNAKULASURIYA; JOHNSON; VAN DER WAAL, 2007; WARNAKULASURIYA, 2018). O seu diagnóstico muitas vezes é difícil e é recomendado inicialmente excluir lesões como queratose friccional, queratose em rebordo alveolar, leucoedema e nevo branco esponjoso. Algumas lesões induzidas pelo

tabaco como: estomatite nicotínica ou lesões associadas a usuários de cachimbo e fumantes invertidos são tradicionalmente separadas de leucoplasia, embora sejam lesões de coloração branca e também associadas ao uso do tabaco. (AXÉLL et al., 1996; CALIFANO et al., 1996; VAN DER WAAL, 2015; WARNAKULASURIYA, 2018)

A leucoplasia não está associada a nenhum agente físico ou químico, com exceção do tabaco, álcool ou *betel quid* (AXÉLL et al., 1996; REIBEL, 2003; WARNAKULASURIYA, 2018). Normalmente esta lesão é diagnosticada depois da quarta década de vida, é mais comum em homens e é até seis vezes mais frequente em fumantes do que não fumantes (WARNAKULASURIYA, 2018).

Foi estimado para as leucoplasias uma prevalência global de 2% (PETTI, 2003; DIONNE et al., 2015) a 4% (MELLO et al., 2018) e uma taxa de transformação maligna que varia de 1 a 18% (REIBEL, 2003; DIONNE et al., 2015). Geralmente é uma lesão assintomática, pode comprometer qualquer setor da mucosa oral e varia em tamanho. Está dividida em leucoplasia homogênea e não homogênea, sendo que esta última variedade envolve três tipos clínicos: eritroleucoplasia, nodular e verrucosa ou exofítica. (WARNAKULASURIYA, 2018)

A eritroleucoplasia está caracterizada por apresentar uma área de coloração mista, branca e eritematosa. A leucoplasia nodular está constituída por várias pápulas eritematosas ou brancas. Já a do tipo verrucosa ou exofítica apresenta uma superfície de aspecto corrugada ou verrucosa. Quando é observada uma placa extensa ou múltiplas placas com uma superfície verrucosa e exofítica é utilizado o termo leucoplasia verrucosa proliferativa (LVP) (PINDBORG et al., 1997; VAN DER WAAL, 2009; WARNAKULASURIYA, 2018). A LVP é uma forma distinta, rara e agressiva de DOPM (EL-NAGGAR et al., 2017), o seu diagnóstico somente deve ser feito em uma análise retrospectiva, pois caracteriza-se por uma lesão irreversível, persistente, de curso progressivo, resistente a maioria das terapias e com uma alta tendência a recorrência (VAN DER WAAL, 2015; EL-NAGGAR et al., 2017).

A eritroplasia, apesar de menos frequente que a leucoplasia, está relacionada a um maior risco de desenvolvimento maligno. Está definida como uma placa homogênea eritematosa, excluindo outras lesões conhecidas clínica ou patologicamente. (REICHART; PHILIPSEN, 2005; HOLMSTRUP, 2018; WARNAKULASURIYA, 2018). O diagnóstico diferencial deve ser realizado principalmente com a candidíase eritematosa, gengivite descamativa, lúpus discóide, líquen plano erosivo, penfigoide e

outras condições inflamatórias. É predominantemente vista na meia idade e idosos, sem predileção por sexo e a taxa de prevalência varia entre 0,02% e 0,83%. (REICHART; PHILIPSEN, 2005)

A eritroplasia normalmente é observada com maior frequência no palato mole, o assoalho da boca e a mucosa jugal (REICHART; PHILIPSEN, 2005). Esta lesão normalmente é menor a 1,5 cm de diâmetro, apresenta bordos irregulares e a superfície é brilhante, tendo algumas vezes um aspecto granular (WARNAKULASURIYA, 2018).

A etiologia e o comportamento biológico da leucoplasia e eritroplasia são desconhecidos, entretanto existe uma forte associação com o consumo de tabaco e o uso de álcool (REIBEL, 2003; WARNAKULASURIYA; JOHNSON; VAN DER WAAL, 2007; DIONNE et al., 2015; WARNAKULASURIYA, 2018) . Em certas regiões, principalmente em relação a leucoplasia, também está relatado associação o consumo de *betel quid* (AXÉLL et al., 1996; WARNAKULASURIYA, 2018). Outros fatores, como infecção por vírus papiloma humano e infecção por *Candida* foram relacionados, no entanto não foi identificada uma associação causal (SPEIGHT; KHURRAM; KUJAN, 2018; WARNAKULASURIYA, 2018). Algumas leucoplasias não estão associadas a um fator de risco conhecido e são consideradas idiopáticas (WARNAKULASURIYA, 2018).

Existem vários fatores clínicos que devem ser considerados para determinar um maior risco de transformação maligna destas lesões. As características clínicas que mostraram ter forte associação com um maior risco de transformação maligna são: lesões com superfície não homogênea, tamanho maior a 200 mm², coloração eritematosa ou branco-eritematosa e localizadas em bordos laterais de língua ou assoalho bucal. (SPEIGHT; KHURRAM; KUJAN, 2018). Na ausência de hábitos, a presença da leucoplasia tem uma indicação imediata de biópsia, para excluir a presença de DEO (VAN DER WAAL, 2015).

2.4 DISPLASIA EPITELIAL ORAL

A biópsia nas DOPM continua sendo o procedimento padrão para avaliar risco de transformação maligna (DIONNE et al., 2015; SPEIGHT; KHURRAM; KUJAN, 2018). O tecido extirpado das leucoplasias e ou eritroplasia podem corresponder a uma hiperqueratose, atrofia, displasia epitelial nos diferentes graus e COCE (DIONNE et al.,

2015; WARNAKULASURIYA et al., 2008). A DEO é um espectro de alterações epiteliais arquiteturais e citológicas causadas pelo acúmulo de alterações genéticas, associado a um risco aumentado de progressão para o carcinoma oral de células escamosas. O sistema de classificação mais utilizado e que é confirmado pela OMS, distingue entre DEO leve, moderada e severa (EL-NAGGAR et al., 2017). O carcinoma *in situ* (WARNAKULASURIYA et al., 2008), antes também incluído nesta classificação, foi sugerido a ser considerado como sinônimo de DEO severa (EL-NAGGAR et al., 2017).

Para melhorar a concordância entre os patologistas, em relação ao grau de displasia existe uma tendência em utilizar o sistema binário que divide as lesões em alto e baixo grau (KUJAN et al., 2006; EL-NAGGAR et al., 2017). Achados displásicos do epitélio escamoso estratificado são caracterizados por proliferação, maturação e diferenciação anormais das células epiteliais (EL-NAGGAR et al., 2017). Os critérios utilizados no diagnóstico de DEO (PINDBORG et al., 1997; REIBEL, 2003; WARNAKULASURIYA et al., 2008) são descritos na **Tabela 1**. Lesões de alto grau mostram ao menos quatro características arquitetônicas e cinco citológicas, enquanto que as lesões de baixo grau mostram apenas pequenas mudanças (KUJAN et al., 2006).

Tabela 1 Critérios usados no diagnóstico de DEO*.

| Arquitetura | Citologia |
|--|---|
| Estratificação epitelial irregular | Variação anormal no tamanho nuclear (anisonucleose) |
| Perda de polaridade das células basais | Variação anormal na forma nuclear (pleomorfismo nuclear) |
| Hiperplasia das células basais | Variação anormal no tamanho das células (anisocitose) |
| Papilas em forma de gota | Variação anormal da forma celular (pleomorfismo celular) |
| Aumento do número de figuras mitóticas | Relação nuclear-citoplasmática aumentada |
| Mitoses superficiais anormais | Tamanho nuclear aumentado |
| Queratinização pré-matura em células isoladas (disqueratose) | Figuras mitóticas atípicas |
| Pérolas de queratina dentro de papilas | Aumento do número e tamanho dos nucléolos Hiperchromasia |

*PINDBORG et al., 1997; REIBEL, 2003; WARNAKULASURIYA et al., 2008.

As evidências demonstram que quanto mais severa a displasia, maior a possibilidade de transformação maligna, o que torna importante ter uma adequada diferenciação das características histopatológicas de cada lesão (KUJAN et al., 2006; WARNAKULASURIYA et al., 2008, 2011; SCULLY, 2014; DIONNE et al., 2015; SPEIGHT; KHURRAM; KUJAN, 2018). A subjetividade no diagnóstico dos diferentes graus de DEO é um fator que dificulta o reconhecimento do potencial de transformação maligna. Muitos estudos demonstram grandes variações individuais e entre diferentes patologistas para diagnosticar quanto à presença ou ausência, e aos diferentes graus de displasia. Quando comparado o grau de malignização do epitélio oral através dos valores *kappa*, em geral, foi observada uma pobre ou moderada concordância entre os examinadores. (KUJAN et al., 2006; WARNAKULASURIYA et al., 2008; DOST et al., 2014)

Importante ressaltar que o conhecimento dos aspectos clínicos destas lesões, não aumentaram as taxas de concordância diagnóstica. (KUJAN et al., 2006; WOOLGAR, 2006; WARNAKULASURIYA et al., 2008). Esta variabilidade é particularmente problemática, quando diagnósticos incorretos podem levar a prognósticos diferentes para a mesma lesão, recebendo um sobre ou subtratamento. Por este e outros motivos, muitos pesquisadores tem tratado definir marcadores celulares e moleculares ou alterações no genoma para prognosticar o risco de transformação maligna (DIONNE et al., 2015; SPEIGHT; KHURRAM; KUJAN, 2018). Mas, nenhum estudo conseguiu melhorar até o momento o panorama atual que possa ser usado na rotina diagnóstica (SPEIGHT; KHURRAM; KUJAN, 2018).

A detecção de DEO continua sendo considerada a característica mais importante para a caracterização de alterações mucosas e para predição de transformação maligna (WARNAKULASURIYA, 2014; NIKITAKIS et al., 2018a). No entanto, esta característica não reflete necessariamente uma continuidade para a malignidade (NIKITAKIS et al., 2018a) e, além disso, a ausência desta não impossibilita a natureza de risco de transformação maligna das DOPM (LINGEN et al., 2011).

2.5 DOPM: MARCADORES CELULARES

O aspecto clínico e as alterações microscópicas que indicam a presença de DOPM

são provocadas por alterações moleculares específicas, genéticas e epigenéticas, que se acumulam ao longo do tempo, relacionadas com a transformação maligna (NIKITAKIS et al., 2018a). Muitas das moléculas investigadas estão associadas a vários processos oncogênicos críticos, incluindo as chamadas *hallmarks* do câncer, como proliferação celular, evasão da supressão do crescimento, resistência à apoptose, imortalidade replicativa, angiogênese, invasão e metástase. Nesta busca de biomarcadores, outros processos avaliados são: reprogramação do metabolismo energético de evasão da vigilância imunológica, instabilidade genômica e inflamação promotora de tumor (HANAHAN; WEINBERG, 2011). Um estudo de revisão sistemática identificou 113 biomarcadores em DOPM, dos quais os mais comuns foram p53, Ki-67, PCNA, perda de heterozigotidade (LOH) em 3p, 8q, 9p e 11p, além de uma variedade de ciclinas e metaloproteínas da matriz extracelular (MMP). Em relação a associação do risco de progressão maligna, foram sugeridos LOH em 3p e 9p, survivina, MMP-9 e conteúdo anormal de DNA. (SMITH et al., 2009)

Embora a ploidia do DNA não tenha associação a uma alteração genética específica, ela mede as alterações genéticas globais decorrentes da instabilidade genômica que promove uma diversidade de alterações moleculares associadas ao desenvolvimento do câncer, como deleções e ampliações de DNA. (DIONNE et al., 2015; NIKITAKIS et al., 2018a)

Até 80% dos casos de COCE apresentam conteúdo anormal de DNA e fortes evidências sugerem que a detecção de alterações semelhantes em leucoplasia e eritroplasia é positivamente correlacionada ao risco de transformação maligna, independentemente do grau de displasia (DIONNE et al., 2015). Bradley et al. (BRADLEY et al., 2010) encontrou variações de ploidia em 45% casos de displasia oral que evoluíram para carcinoma e apenas 12% dos casos que não progrediram. Foi determinado que as lesões que apresentaram aneuploidia tem 3.2 vezes mais possibilidades de transformação maligna e estão associadas a uma displasia com maior grau de severidade (ALAIZARI et al., 2018). Um outro estudo demonstrou que a aneuploidia e grau de DEO são características preditivas independentes para determinar risco de transformação maligna. No entanto, estas quando combinadas, tendem a aumentar o valor preditivo positivo (SPERANDIO et al., 2013).

A LOH ocorre como resultado da perda de material genômico, afetando uma única cópia remanescente do alelo funcional. Padrões de LOH foram encontrados em amostras

de COCE em vários locais em todos os cromossomos, exceto X e Y. Em relação a transformação maligna em DOPM, foram encontradas associações na presença de LOH em 3p e 9p (ROSIN et al., 2000; DIONNE et al., 2015; EL-NAGGAR et al., 2017; NIKITAKIS et al., 2018a), além de 4q e 17p (EL-NAGGAR et al., 2017). Dados de uma coorte prospectiva sugerem que DOPM com LOH em 3p e/ou 9p aumentam até 22.6 vezes o risco de transformação maligna (ZHANG et al., 2012), evidenciando-se acúmulo de dano genético nestas lesões (LINGEN et al., 2011). Foram feitas outras associações de genes relacionados com o câncer nas localizações associadas a LOH em DOPM, entre estas: 3p14 (gene *FHIT*), 9p21 (gene *p16^{INK4a}/CDKN2A*) e 17p (gene *TP53*). (NIKITAKIS et al., 2018a)

As variações do número de cópias (CNV) são duplicações ou deleções que existem em grandes partes do genoma (DIONNE et al., 2015). Em relação a progressão de DOPM a COCE foram observados ganhos principalmente em 3q, 5p, 7p, 8q, 11q e 20q, já as perdas foram frequentemente descritas em 3p, 8p, 9p e 18q. Embora nas VNC, as perdas sejam menos frequentes que os ganhos, estas podem ser eventos primários na causa do câncer (SALAHSHOURIFAR et al., 2014). O fenótipo -8p/+3q/+8q/+20 previamente relatado em COCE (BHATTACHARYA et al., 2011) foi identificado em lesões displásicas, sugerindo que esse padrão de instabilidade genômica pode ocorrer nas fases prévias ao desenvolvimento do COCE e pode ser um marcador útil de progressão tumoral (DIONNE et al., 2015).

Vários genes/proteínas emergiram como potenciais marcadores de DEO e de transformação maligna das DOPM. Os exemplos mais extensivamente estudados incluem p53, Ciclina D1 e podoplanina (DIONNE et al., 2015). Um estudo recente mostrou relação entre uma superexpressão de p53, Ki-67, p16, β -cetenina, c-jun, c-met, IMP3, COX-2, podoplanina e CA9, com o aumento da possibilidade de transformação maligna para leucoplasias. Este estudo também reportou que a combinação de p53 e CA9 com fatores clínicopatológicos, incluindo idade e grau de displasia, fornece melhor acurácia na predição de malignidade (ZHANG et al., 2017). A MMP-9 apesar de ser um marcador útil de prognóstico em COCE foi pouco estudada em DOPM.

É importante ressaltar que um processo acumulativo de mutações em DOPM, ainda desconhecidas, podem evoluir para um fenótipo invasivo maligno progressivo. Tais alterações genéticas resultam na inativação de múltiplos genes supressores de tumores e na ativação de proto-oncogenes, incluindo *CDKN2A*, *TP53*, *CCND1*, *FHIT*, *RASSF1A*,

receptor do fator de crescimento epidérmico (*EGFR*) e *RBI*. Outros genes que também contribuem para o processo de carcinogênese e metástase são os que regulam os fatores de crescimento endotelial vascular (VEGF e VEGF-C), as citocinas, como interleucina-6 (IL-6) e IL-8, e seus receptores, bem como os genes anti-apoptóticos da família do Bcl-2 e survivina (PEREZ-ORDONEZ; BEAUCHEMIN; JORDAN, 2006; DIONNE et al., 2015). Entretanto, nenhum dos marcadores reportados é específico, e também se encontram alterados na maioria dos tumores e processos histológicos prévios a malignização, eliminando a utilização dos mesmos na prática clínica. (DIONNE et al., 2015; NIKITAKIS et al., 2018a)

As alterações epigenéticas envolvem modificações de DNA e histonas que não são codificadas na seqüência do DNA, embora essas mudanças não sejam hereditárias (EGGER et al., 2004). Três sistemas estão envolvidos: hipermetilação do DNA, silenciamento pós-transcricional associado ao RNA e modificação de histonas (LINGEN et al., 2011). Destes, a metilação do DNA parece ser a mais importante, sendo a hipermetilação observada em COCE e DOPM, porém sem associação com o grau de DEO e com predição de transformação maligna (LINGEN et al., 2011; NIKITAKIS et al., 2018a). Recentes estudos revisaram moléculas envolvidas no mecanismo de resposta a danos do DNA, já que uma disfunção deste mecanismo pode estar associada a transformação maligna, observando principalmente uma associação do aumento da expressão de γ H2 AX, forma fosforilada de uma das mais comuns histonas (H2AX), com a possibilidade de transformação maligna. (NIKITAKIS et al., 2018b; ZHU et al., 2018)

Os Micro-RNAs (miRNAs) são pequenas moléculas de RNA, não codificantes, que atuam como reguladores pós-transcricional da expressão gênica. Estas moléculas têm sido envolvidas na patogênese de uma variedade de cânceres, direcionados a vários oncogenes ou genes supressores de tumor, estando sobre-expressados em 70% dos cânceres orais. No entanto, existem poucos estudos que avaliam a associação destes em DOPM. Até o momento, o aumento de miR-31 foi frequentemente observado em leucoplasias com alto risco de transformação maligna. (HUNG et al., 2016; NIKITAKIS et al., 2018a).

2.6 SEQUENCIAMENTO DE NOVA GERAÇÃO EM DOPM

O incompleto entendimento da história natural das DOPM bem como sua composição genética torna difícil a classificação antecipada de casos positivos para transformação maligna. Um objetivo fundamental da investigação sobre a carcinogênese, com implicações para a prevenção do câncer, é determinar a ordem das mutações que ocorrem em uma neoplasia à medida que progride a partir do tecido normal ao tumor. Esta análise tem sido amplamente influente e usada como modelo em vários tipos de câncer, incluindo câncer de mama, o câncer de pulmão, e melanoma. Um modelo utilizado para entender a malignização é a sequência linear de mutações que devem ocorrer em ordem, começando nas células normais. Por tratar-se de um modelo evolutivo, cada mutação tem ao menos uma que a sucede e, sua presença, aumenta as chances de encontrar a próxima, sucessivamente (CALIFANO et al., 1996; BRADLEY et al., 2010). Portanto, o acúmulo de alterações genéticas conduz a uma transformação maligna (CALIFANO et al., 1996). Pesquisadores relataram que a maioria das alterações genéticas ocorre antes da carcinogênese e que os perfis de expressão gênica diferem entre mucosa normal, DOPM e COCE (HA et al., 2003).

O desenvolvimento de DOPM e COCE são processos de múltiplas etapas, envolvendo alterações genéticas devido a fatores exógenos ou do próprio indivíduo. Uma ferramenta de excelência utilizada no diagnóstico de outras lesões benignas e tumores malignos é a estratificação molecular. Esta caracterização utiliza tecnologia de última geração baseada no sequenciamento e estabelecimento de mutações que sejam típicas, ou seja, repetidamente encontradas nas mesmas lesões. (BRADLEY et al., 2010; GRAVELAND et al., 2013)

A partir de 2005, surgiram os primeiros equipamentos de sequenciamento de DNA conhecidos como sequenciamento de nova geração (NGS, *Next-Generation Sequencing*), permitindo uma nova abordagem de sequenciamento em larga escala (HGS – *High throughput sequencing*) (GOODWIN; MCPHERSON; MCCOMBIE, 2016). O Sequenciamento do Exoma completo (WES, *Whole Exome Sequencing*) e do genoma completo mostraram-se úteis para reconhecer mutações *driver* no câncer de cabeça e pescoço (BHATIA; BURTNESS, 2017). O recente estudo do Atlas do Genoma do Câncer (TCGA) mostrou que o carcinoma de células escamosas de cabeça e pescoço tem um conjunto de anormalidades genômicas, com poucos genes *driver* comuns

(STRANSKY et al., 2012; LAWRENCE et al., 2015; WOOD et al., 2017). No COCE foram identificadas frequentes mutações em *TP53*, *PIK3CA*, *APC*, *MET*, *CDH1*, *FBXW7* (ER et al., 2015). Em pacientes com COCE, usuários de rapé Árabe (*Arabian snuff*), foram encontradas novas mutações em *NOTCH3*, *CSMD3*, *CRB1*, *CLTCL1*, *OSMR* e *TRPM2*, amplificação dos proto-oncogenes *FOSL1*, *RELA*, *TRAF6*, *MDM2*, *FRS2* e *BAG1*, além de deleção do supressor tumoral *SMARCC1* (AL-HEBSHI et al., 2016). Por outro lado, foi reportado que tumores da cavidade oral com resultados clínicos favoráveis apresentaram infrequentes alterações no número de cópias, em associação a mutações ativadoras de *HRAS* ou *PIK3CA* e com mutações inativadoras de *CASP8*, *NOTCH1* e *TP53* (LAWRENCE et al., 2015).

Ao contrário do extenso trabalho que tem sido feito em varias neoplasias em etapas avançadas, poucos estudos descrevem de forma abrangente as alterações genômicas encontradas em tecidos em etapas prévias ao câncer (CAMPBELL et al., 2016; SPIRA et al., 2017). No entanto, a correta caracterização das alterações moleculares em DOPM e as correspondentes alterações no microambiente associadas à progressão ajudariam o desenvolvimento de biomarcadores para detecção precoce e estratificação de risco, além de sugerir intervenções preventivas para reverter ou retardar o desenvolvimento do câncer. Pela complexidade e diversidade das alterações a serem determinadas, são necessárias metodologias mais abrangentes, como NGS (CAMPBELL et al., 2016; SHRIDHAR et al., 2016; BHATIA; BURTNES, 2017).

Em 2009 foi realizado pela primeira vez a análise do genoma completo de lesões com DEO. Este estudo demonstrou que existem desequilíbrios genômicos em lesões com maior risco de transformação maligna. Além disso, comprovou também, de forma pioneira, que lesões com DEO de baixo grau que evoluíram a COCE apresentam um perfil genômico que se assemelha mais as DEO de alto grau, do que aquelas lesões com o mesmo diagnóstico histopatológico, mas que não evoluíram a COCE. (GARNIS et al., 2009)

Utilizando o NGS para avaliar genes candidatos das vias apoptóticas intrínseca e extrínseca, foram determinados polimorfismos de nucleotídeo único (SNP, *Single nucleotide polymorphism*) nos genes *BCL2*, *CASP1* e *CASP10* que são capazes de modular risco de leucoplasia e COCE (DATTA et al., 2015). Wood et al. (WOOD et al., 2017) sugeriram que a maioria das alterações genômicas que conduzem ao câncer bucal ocorrem em etapas prévias a este, por meio do acúmulo aleatório gradual, em vez de um único

evento. Este estudo, através da técnica WES, avaliou 16 pacientes diagnosticados com COCE, utilizando diferentes setores da mesma amostra, com diagnósticos de displasias epiteliais de baixo e alto grau e COCE (WOOD et al., 2017). Em estudo similar anterior, o mesmo grupo de pesquisadores examinaram as relações clonais entre DEO e carcinoma em cinco pacientes, e sugerem que as relações entre estes são às vezes complexas e nem sempre são uma progressão simples (WOOD et al., 2015).

2.7 TRATAMENTO COCE E DOPM

Entender a diversidade genética é importante no sucesso ou fracasso de terapias para o câncer e DOPM. Até o momento, em etapas precoces ou tardias, o COCE é frequentemente tratado cirurgicamente, com radioterapia e com ou sem quimioterapia, no cenário adjuvante pós-operatório (SCULLY; BAGAN, 2009; EL-NAGGAR et al., 2017). Mesmo quando a melhor combinação das abordagens cirúrgicas e não cirúrgicas são usadas, mais de 50% dos pacientes experimentarão o fracasso. Em relação às DOPM o panorama atual não é diferente, não existe um consenso na literatura quanto à conduta clínica em relação a patologias como leucoplasia e eritroplasia. São descritos cirurgias com margem livre de lesão (ZHANG et al., 2016), uso do laser de alta potência para lesões extensas (THOMSON et al., 2017), controle de fatores de risco (ZHANG et al., 2016; AWADALLAH et al., 2018), terapia fotodinâmica, utilização de ácido retinoico, licopeno, consumo de chá verde, suplemento com vitaminas A, C, D, E e K, dieta equilibrada, controles clínicos seriados, entre outros (ZHANG et al., 2016; CHAU et al., 2017; AWADALLAH et al., 2018). No entanto, ainda não foram encontrados alvos terapêuticos para um tratamento que diminua o risco de transformação maligna de DEO para COCE (WARNAKULASURIYA et al., 2011; DIONNE et al., 2015; NIKITAKIS et al., 2018b).

Muitos desafios persistem, incluindo dificuldades na prevenção da progressão maligna e tratamento das DOPM. Para uma melhoria destas dificuldades existentes é necessário um melhor entendimento dos mecanismos moleculares da carcinogênese associados, que permitirá o desenvolvimento nas áreas de diagnóstico, prognóstico, marcadores preditivos e novas terapias alvos específicas. (LINGEN et al., 2011; NIKITAKIS et al., 2018a)

3 OBJETIVOS

3.1 OBJETIVO GERAL

Identificar alterações genômicas presentes em displasias epiteliais orais, de baixo e alto graus, de pacientes brasileiros e chilenos.

3.2 OBJETIVOS ESPECÍFICOS

- Identificar e caracterizar as variantes de sequência de DNA encontradas nas amostras do estudo, relacionando-as com o diagnóstico histopatológico.
- Analisar e caracterizar as variantes que produzem alto impacto biológico das amostras do estudo, relacionando-as aos dados clínicos e histopatológicos.
- Identificar os genes associados as displasias epiteliais orais de baixo e alto grau.

4 METODOLOGIA

4.1 CONSIDERAÇÕES ÉTICAS

Este estudo faz parte de um projeto de pesquisa financiado pelo *Fondo Nacional de Desarrollo Científico y Tecnológico (FONDECYT)* do Chile, identificado pelo número 11140281. Foi avaliado e aprovado pelos Comitês de Ética e Institucional de Biossegurança da Faculdade de Odontologia da Universidade do Chile e Comitê de Ética da Faculdade de Odontologia da Universidade Federal de Pelotas (**Anexo A**). A pesquisa aqui apresentada foi realizada em total conformidade com os princípios éticos locais e a Declaração de Helsinque.

4.2 CARACTERIZAÇÃO DO ESTUDO E DAS AMOSTRAS

Esta pesquisa trata-se de um estudo transversal e descritivo. A partir das bases de dados do Serviço de Anatomia Patológica da Faculdade de Odontologia da Universidade do Chile, do Laboratório de Patologia Oral da Faculdade de Odontologia da Universidade Federal de Pelotas e do Laboratório de Patologia Cirúrgica da Faculdade de Odontologia da Universidade Federal da Bahia, foram analisados blocos de amostras preservadas em parafina com diagnóstico histopatológico de displasia epitelial oral nos diferentes graus. As lâminas histopatológicas foram coradas com hematoxilina-eosina e reavaliadas por um patologista bucal através de microscopia de luz. As amostras foram reclassificadas em displasia epitelial de baixo (DBG) ou alto (DAG) graus, de acordo ao sistema binário descrito por Kujan et al (KUJAN et al., 2006) e pela OMS (EL-NAGGAR et al., 2017).

De um universo com 114 amostras de displasias epiteliais da mucosa oral, foram selecionadas 45 amostras que apresentavam qualidade, do ponto de vista anátomo patológico e quantidade de tecido suficientes para realizar o estudo. Foi considerado a preservação de amostra residual em cada um dos blocos utilizado, que ficou disponível em banco de dados de origem. As amostras selecionadas estavam associadas ao diagnóstico clínico de leucoplasia (homogênea e não homogênea) ou eritroplasia. A avaliação clínica das amostras selecionadas foi realizada por um patologista bucal ou

estomatologista.

Foram considerados as seguintes características clínicas das amostras avaliadas: sexo, idade, ingestão de álcool, consumo de tabaco, região de estabelecimento (Brasil ou Chile), localização anatômica da biopsia realizada e diagnóstico clínico. Para a realização do diagnóstico clínico, as leucoplasias foram classificadas em homogêneas ou não homogêneas, sendo estas últimas subclassificadas em eritroleucoplasia, nodular, verrucosa ou verrucosa proliferativa. O diagnóstico clínico baseou-se nos critérios de Van der Waal, 2009 (VAN DER WAAL, 2009) e Warnakulasuriya, 2018 (WARNAKULASURIYA, 2018).

4.3 EXTRAÇÃO DE DNA GENÔMICO

Todas as 45 amostras de tecido incluído em parafinas foram submetidas ao isolamento de DNA, de acordo as instruções do fabricante (Puregene® DNA Purification Tissue Kit – Gentra Systems, Inc., Minneapolis, Minnesota). O rendimento de DNA foi de 0.2 a 2.0µg. Depois de processar cada amostra, foram obtidos 20µl de solução e dessa quantidade utilizados 1,0µl, os quais foram adicionados a 99µl de água *Milli-Q®*, para a leitura do DNA no espectrofotômetro, sendo checada a quantidade e a pureza de cada amostra de DNA (DU-640, Beckman, Palo Alto, CA, EUA).

4.4 SEQUENCIAMENTO DO EXOMA COMPLETO

4.4.1 PREPARAÇÃO DAS AMOSTRAS E CONSTRUÇÃO DAS BIBLIOTECAS

As amostras de DNA foram enviadas para realização do sequenciamento do exoma completo na *Humanizing Genomics Macrogen*, Macrogen Korea, Seoul, Rep. Korea. Inicialmente, as 45 amostras passaram por um controle de qualidade interno para checar integridade e concentração por *PicoGreen* das moléculas de DNA. Destas amostras, foram selecionadas 10 para o sequenciamento, pois tinham qualidade superior que as outras amostras. Do total de amostras incluídas no estudo, 6 tinham diagnóstico de DBG

e 4 de DAG. O Kit SureSelectXT Library Prep Kit (Agilent) foi utilizado para preparar a biblioteca e captura da sequência do exoma das amostras de DNA. A biblioteca de sequenciamento foi preparada por fragmentação aleatória em cada uma das amostras de DNA, seguido pela ligação no adaptador 5' e 3'. Alternativamente, *tagmentation* combina as reações de fragmentação e ligação em um único passo, aumentando significativamente a eficiência do processo de preparação da biblioteca. Os fragmentos ligados ao adaptador foram então amplificados por PCR e purificados em gel. O protocolo de classificação pós captura, SureSelectXT Target Enrichment System for Illumina Version B.2, April 2015, foi utilizado para assegurar a alta eficiência e cobertura. O sequenciamento das bibliotecas do exoma foi realizado usando 101bp *paired-end reads* em um sequenciador HiSeq-2500 (Illumina®) com planejamento de profundidade de sequenciamento de destino de pelo menos 100x, o que é considerado ser suficiente para detectar a maioria das variantes.

4.4.2 SEQUENCIAMENTO E OBTENÇÃO DE DADOS

Para geração de *cluster*, a biblioteca é carregada em uma célula de fluxo, onde os fragmentos são capturados em uma placa com oligos ligados à superfície, complementares aos adaptadores da biblioteca. Cada fragmento é então amplificado em *clusters* de clones através da amplificação em ponte (*bridge amplification*). Quando a geração dos *clusters* é concluída, os moldes estão prontos para o sequenciamento.

A tecnologia *Illumina*® *SBS* usa um método baseado em terminador reversível que detecta bases simples à medida que elas são incorporadas em fitas de modelo de DNA. Como todos os 4 dNTPs reversíveis, ligados ao terminador, são persistentes durante cada ciclo de sequenciamento, a competição natural minimiza o viés de incorporação e reduz consideravelmente as taxas de erros brutos em comparação com outras tecnologias. O resultado é um sequenciamento base-por-base altamente preciso que praticamente elimina erros específicos do contexto de sequência, mesmo em regiões de sequência repetitiva e homopolímeros. Estes dados obtidos do sequenciamento são convertidos em dados brutos (*raw data*) para a análise. O sequenciador *Illumina*® gera imagens brutas utilizando software de controle de sequenciamento para domínio do sistema e chamadas básicas através de um software de análise primária integrado chamado RTA (*Real Time*

Analysis). O binário BCL (*base calls*) é convertido em FASTQ utilizando o pacote illumina bcl2fastq. Os adaptadores não são removidos dos *reads*.

O número total de bases, leituras, GC (%), Q20 (%) e Q30 (%) foi calculado para as 10 amostras (**Anexo B: Tabela S1, Figura S1, Figura S2, Figura S3**). Cada par de sequências ficaram disponíveis em arquivos tipo *fastq.gz*.

4.4.3 ANÁLISE BIOINFORMÁTICA DOS DADOS

A análise dos dados contou com a colaboração do BioinfoGP (Spanish National Biotechnology Centre, CNB-CSIC. Espanha). O fluxo de trabalho de análise baseou-se nos protocolos descritos GATK (versão 4.0.2.1) (MCKENNA et al., 2010; VAN DER AUWERA et al., 2013), seguido pela quantificação das variantes detectadas, testes estatísticos comparativos por grupo com base na presença ou ausência de cada variante, anotação funcional, filtragem de variantes e formato dos dados finais.

A qualidade das sequências brutas foi analisada usando o software *FastQC* (versão 0.11.7) (ANDREWS, 2010) e não foi detectada nenhum problema de qualidade. Logo, as sequências brutas foram alinhadas pelo *BWA-MEM* (v0.7.17) (LI, 2013) com o genoma de referência humano (*Ensembl release GRCh38.91* (ZERBINO et al., 2018)), utilizando parâmetros por defeito. As rotinas *MarkDuplicates*, *BaseRecalibrator* e *ApplyBQSR* do GATK foram aplicadas para detectar duplicações de leitura e recalibrar as qualidades de alinhamento, respectivamente. A recalibração foi baseada no padrão ouro de 1000 genomas fornecido pelo GATK.

As funções GATK *HaplotypeCaller* e *GenotypeGVCF* foram usadas para SNP/indel e genotipagem. Anotações de recalibração e filtragem de variantes foram adicionadas. A recalibração com as rotinas GATK *VariantRecalibrator* e *ApplyVQSR* foi baseada no *HapMap* (build 3.3) (CONSORTIUM., 2003), 1000 genomas (fase 1) de alta confiança, *omni* (v2.5) SNPs (1000 Genomes Project Consortium, 2012), e *dbSNP* (build 146) (SHERRY et al., 2001). As variantes de cada amostra foram então combinadas em um único arquivo *gvcf* para análise comparativa. A rotina *CaseControl* incluída no *SnpSift* (CINGOLANI et al., 2012) foi usada para detectar variantes com ocorrência diferencial entre amostras de alto e baixo risco. Os valores de *p-value* foram obtidos para diferentes modelos genéticos: dominante, recessivo, alélico, genotípico/codominante e tendência

Cochran-Armitage.

Cada variante foi anotada usando *Variant Effect Predictor* (MCLAREN et al., 2016) da versão 91 do *Ensembl*, com a opção *-everything* e o *Condel algorithm plugin*. As anotações foram obtidas de: *Ensembl* - incluindo variantes e anotações funcionais - (91), 1000 genomas (fase 3) (SUDMANT et al., 2015), *Cosmic* (FORBES et al., 2017), *ClinVar* (LANDRUM et al., 2016), ESP (NHLBI EXOME SEQUENCING PROJECT (ESP), 2017), HGMDPUBLI (STENSON et al., 2003), dbSNP (SHERRY et al., 2001), *Gencode* (HARROW et al., 2012), *Genebuild* (AKEN et al., 2016), *gnomAD* (LEK et al., 2016), *Polyphen* (ADZHUBEI et al., 2010), *regbuild*, *SIFT* (NG; HENIKOFF, 2001) e *Condel* (GONZÁLEZ-PÉREZ; LÓPEZ-BIGAS, 2011).

Os efeitos das variantes de sequência também foram avaliadas usando os programas computacionais: Anotação de Proteína Através da Relação Evolutiva (PANTHER) (MI et al., 2019), STITCH (SZKLARCZYK ET AL., 2016) e Mutações Patológicas (PMut) (LÓPEZ-FERRANDO et al., 2017). As informações de sequência para as variantes encontradas serão submetidas ao banco de dados SNP (dbSNP) (<https://www.ncbi.nlm.nih.gov/snp>).

4.5 ANÁLISE ESTATÍSTICA

A relação entre a quantidade de variantes por amostra e por grupo foi analisada pelos testes do qui-quadrado (X^2) e de Fisher. A correlação entre total de variantes e grau de displasia será determinada pelo teste de *Sperman*. Os cálculos estatísticos foram realizados utilizando o GraphPad Prism 6.03 (San Diego, CA, EUA). Significância estatística foi estabelecida em $p < 0,05$.

5 RESULTADOS

Os dados clínicos e diagnósticos histopatológicos das amostras estão sistematizados na **Tabela 2** e **Figura 1**. A média de idade dos 10 pacientes foi de 54 anos. Através do sequenciamento de cada uma das 10 amostras do presente estudo, um total de 3.055.651 variantes foram identificadas, sendo 2.099.808 (68,7%) variantes existentes e 955.843 (31,3%) novas. As classes destas variantes estão descritas na **Tabela 3**, sendo que 90,4% destas foram variantes de nucleotídeo único (SNV, *Single Nucleotide Variant*).

Para cada variante determinada, ao mapeá-las com genomas de referência, foram determinadas suas consequências, sendo as mais frequentes localizadas em regiões intrônicas (52 %) e intergênicas (29%), descritas na **Figura 2** e **Tabela S2 (Anexo B)**. Os cromossomos 1, 2 e 3 concentraram a maior quantidade de variantes. A quantidade de variantes em relação a sua localização por cromossomo pode ser observada nas **Figura 3** e **Tabela S3 (Anexo B)**. A distribuição das variantes em cada cromossomo está detalhada na **Figura S4 (Anexo B)**. Em relação a posição das variantes nas proteínas afetadas, pode-se observar que o maior número de variantes afetou, principalmente, o último decil e um menor número de variantes se localizaram nos decis centrais (**Figura 4**).

Tabela 2. Características clínicas e diagnóstico histopatológico das amostras.

| Am | Idade | Sexo | Tabaco | Álcool | Local. | Dx. Clínico | Dx. Hist. | Origem | Tipo biópsia |
|----|-------|------|--------|--------|------------------|-----------------------|-----------|--------|--------------|
| 1 | 56 | F | Sim | Sim | Língua | Leucoplasia homogênea | DBG | Chile | I |
| 2 | 47 | F | Não | Sim | Língua | Leucoplasia verrucosa | DBG | Chile | E |
| 3 | 38 | F | Sim | Sim | Mucosa Jugal | Leucoplasia homogênea | DBG | Brasil | I |
| 4 | 51 | F | Não* | Não | Palato | Leucoplasia verrucosa | DBG | Chile | E |
| 5 | 49 | F | Sim | Sim | Língua | Leucoplasia verrucosa | DBG | Chile | E |
| 6 | 52 | M | Sim | Não | Rebordo alveolar | Leucoplasia verrucosa | DBG | Chile | E |
| 7 | 54 | F | Sim | Sim | Rebordo alveolar | Leucoeritro. | DAG | Chile | I |
| 8 | 69 | F | Sim | Sim | Assoalho de boca | Leucoplasia homogênea | DAG | Chile | E |
| 9 | 82 | M | Não* | Não* | Língua | Leucoeritro. | DAG | Brasil | I |
| 10 | 38 | M | Sim | Sim | Mucosa Jugal | Leucoplasia homogênea | DAG | Brasil | I |

Am., Amostra; Local., Localização; Dx., Diagnóstico; Hist., Histológico; Leucoeritro., Leucoeritroplasia; DAG, Displasia epitelial de alto grau; DBG, Displasia epitelial de baixo grau; * Deixou o hábito faz mínimo 5 anos; I, Incisional; E, Excisional.

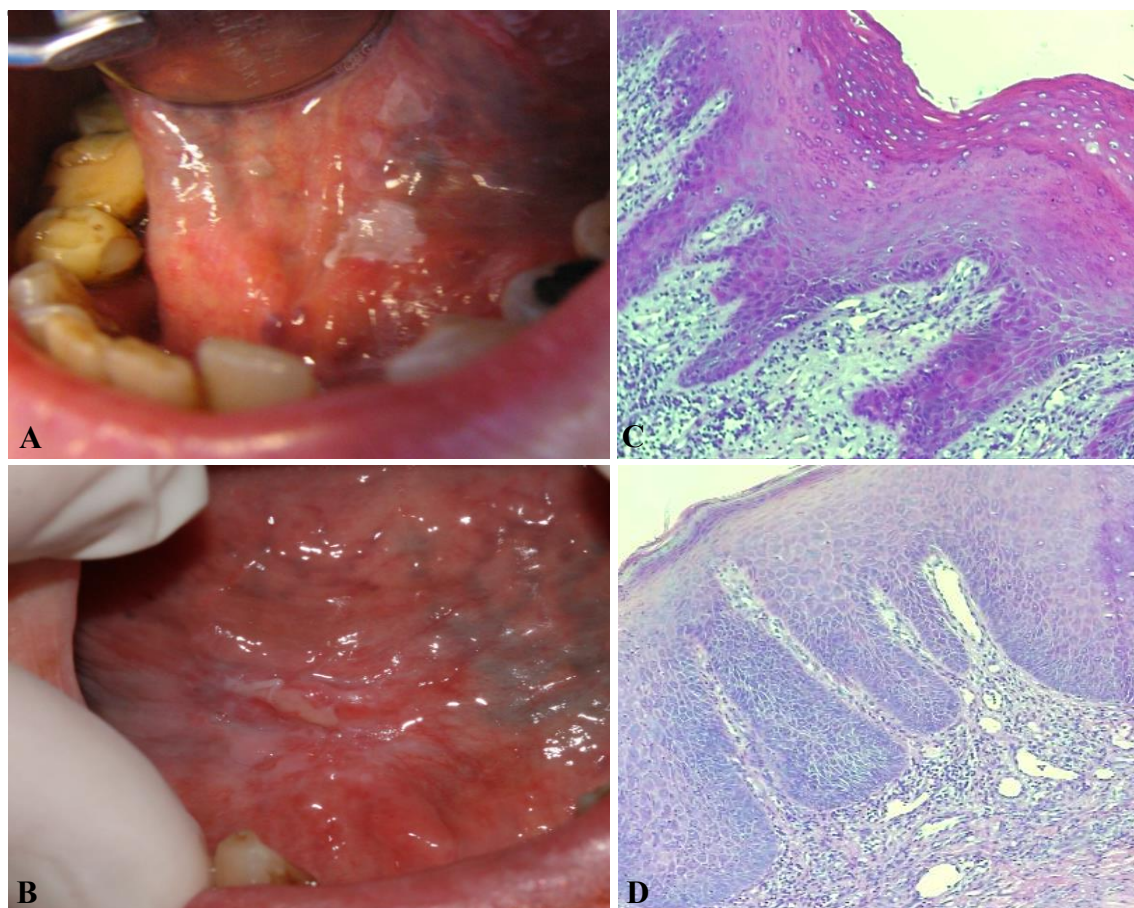


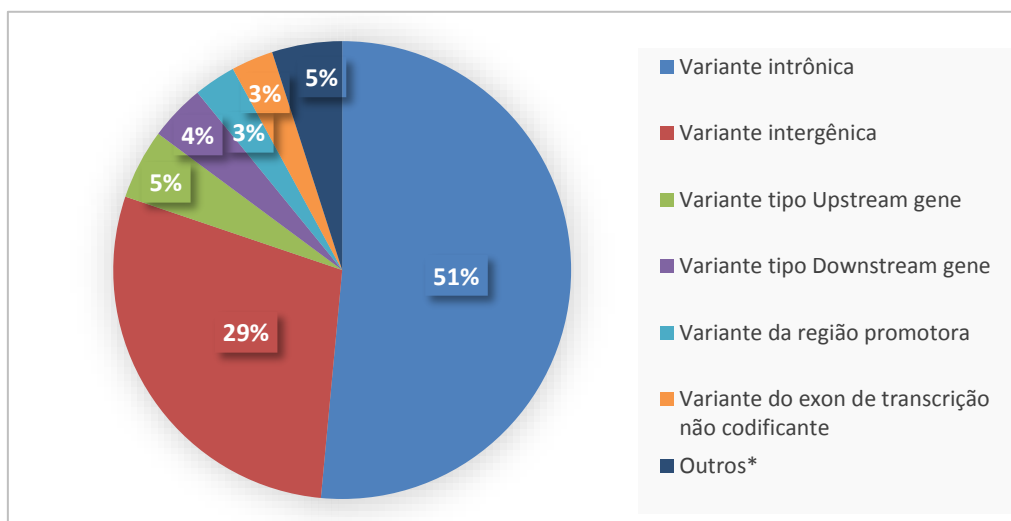
Figura 1 Imagens clínicas e histopatológicas representativas. **A** Imagem clínica do paciente n° 1 com diagnóstico de Leucoplasia homogênea; **B** Imagem clínica do paciente n°8 com diagnóstico de leucoeritroplasia; **C** Imagem histopatológica da amostra n° 3 com diagnóstico de DBG; **D** Imagem histopatológica da amostra n° 8 com diagnóstico de DAG.

Tabela 3 Classes das variantes pontuais identificadas

| Classe da Variante* | Nº de variantes (%) |
|------------------------|---------------------|
| SNV | 2.761.210 (90,4%) |
| Deleção | 166.612 (5,45%) |
| Inserção | 117.375 (3,84%) |
| Alteração de sequência | 9.124 (0,3%) |
| Indel | 1.330 (0,04%) |

* Estabelecida em *Variant Effect Predictor (VEP)* (MCLAREN et al., 2016);

Indel, Inserção e deleção; SNV., variante de nucleotídeo único



* Descritos em tabela complementar (Tabela S2, Anexo 3)

Figura 2 Localizações genômicas das consequências calculadas para as variantes identificadas.

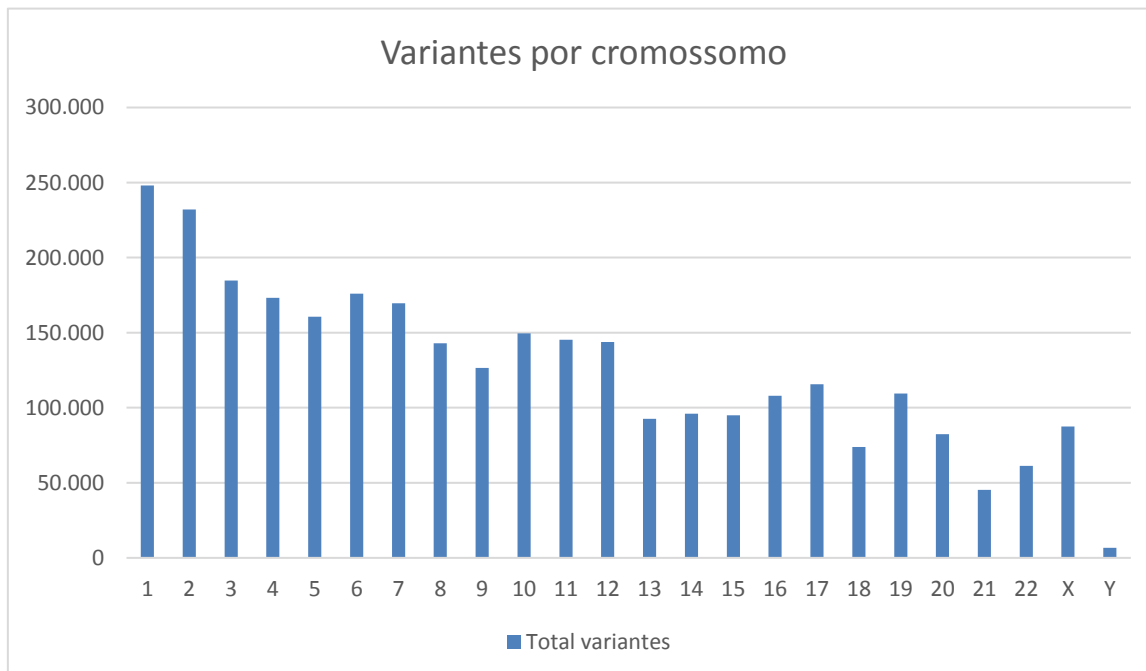


Figura 3 Total de variantes encontradas em cada um dos cromossomos autossômicos e sexuais.

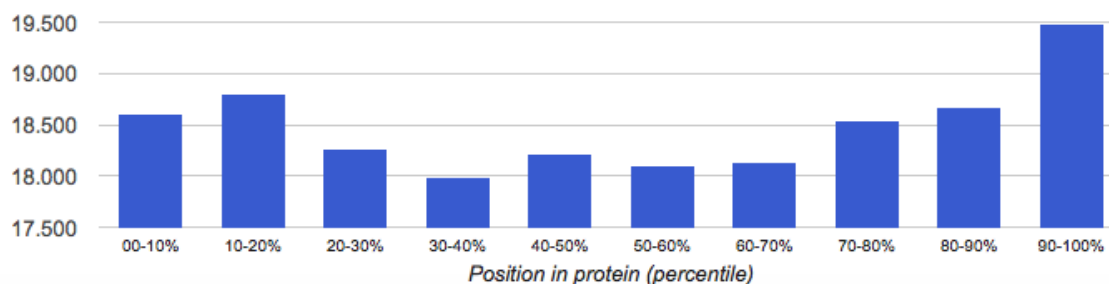


Figura 4 Distribuição dos decis da proteína em relação a presença de variantes.

Analisando as variantes encontradas por amostra, foi observado que o paciente 10 apresentou o maior número de variantes SNV ($p < 0,0029^*$, HET; $p = 0,136$ HOM) e INDEL (variantes de inserção e deleção; $p = 0,77$, HET; $p = 0,81$, HOM), em heterozigose e homozigose. Por outro lado, a amostra 2, de DBG, teve o menor número de SNV (heterozigose e homozigose; $p > 0,05$) e INDEL em heterozigose ($p > 0,05$). A menor quantidade de variantes INDEL em homozigose foi observada na amostra 5 ($p > 0,05$). A média total das variantes SNV e INDEL foi mais alta no grupo de DAG que no grupo de DBG ($p = 0,0003^*$ e $p = 0,40$, respectivamente). Separadamente, a média total das variantes SNV e INDEL, heterozigotos e homozigotos, foram maiores no grupo DAG ($p = 0,475$, $p = 0,4975$, $p = 0,803$ e $p = 0,91$ para SNV e INDEL, heterozigotos e homozigotos, respectivamente). Estes resultados podem ser observados na **Tabela 4**.

De acordo as predições dos programas *SIFT* (NG; HENIKOFF, 2001) e *PolyPhen* (ADZHUBEI et al., 2010), 11 a 21% das variantes são provavelmente deletérias e 60 a 75% são benignas (**Figura 5** e **Tabela S4- Anexo B**). Do total de variantes observadas e analisadas ($n = 3.055.651$), 1.069 foram previstas com maiores possibilidades de causar alterações de baixo, moderado e alto impacto nas estruturas aminoacídicas. Estas variantes produzem alterações em um total de 774 genes, sendo que em 416 genes a probabilidade do efeito da modificação é de baixo risco, em 319 é de risco moderado e em 39 é de alto risco (**Tabela S5, Anexo B**). Do total de genes alterados das amostras de DEO, as funções moleculares que agrupam uma maior quantidade de genes são as de união e atividades catalíticas, com um total de 196 (25%) e 171 (22%) genes, respectivamente (**Figura 6**). Foram observadas participações deste total de genes em mais de 60 vias de sinalizações, no entanto as que agrupam um maior número de genes são: a via de sinalização *Wnt* e da integrina, relacionadas a 15 e 14 genes, respectivamente (**Tabela 5**).

Tabela 4 Total de variantes por amostra e por grau de DEO.

| Amostra | Tipo DE | SNV | SNV | SNV | INDEL | INDEL | INDEL |
|-----------|------------|----------------|----------------|----------------|---------------|---------------|---------------|
| | | Het. | Hom. | Total | Het. | Hom | Total |
| 1 | DBG | 118.683 | 266.874 | 385.557 | 30.844 | 15.780 | 46.624 |
| 2 | DBG | 100.488 | 223.701 | 324.189 | 26.095 | 13.472 | 39.567 |
| 3 | DBG | 126.435 | 281.636 | 408.071 | 31.654 | 15.169 | 46.823 |
| 4 | DBG | 127.014 | 280.353 | 407.367 | 31.271 | 14.747 | 46.018 |
| 5 | DBG | 102.957 | 261.329 | 364.286 | 26.975 | 12.225 | 39.200 |
| 6 | DBG | 104.476 | 241.417 | 345.893 | 26.933 | 13.330 | 40.263 |
| \bar{X} | DBG | 113.342 | 229.218 | 372.560 | 28.962 | 14.121 | 43.083 |
| 7 | DAG | 110.266 | 256.388 | 366.654 | 27.944 | 13.942 | 41.886 |
| 8 | DAG | 139.676 | 324.735 | 464.411 | 33.500 | 16.555 | 50.055 |
| 9 | DAG | 128.667 | 297.701 | 426.368 | 31.723 | 14.981 | 46.704 |
| 10 | DAG | 294.046 | 455.377 | 749.423 | 45.268 | 27.030 | 72.298 |
| \bar{X} | DAG | 168.164 | 333.550 | 501.714 | 34.609 | 18.127 | 52.736 |

DE, Displasia Epitelial Oral; DBG, Displasia de Baixo Grau; DAG, Displasia de Alto Grau; SNV, Variante de um único nucleotídeo; Het., Heterozigoto; Hom., Homozigoto; INDEL, inserções e deleções

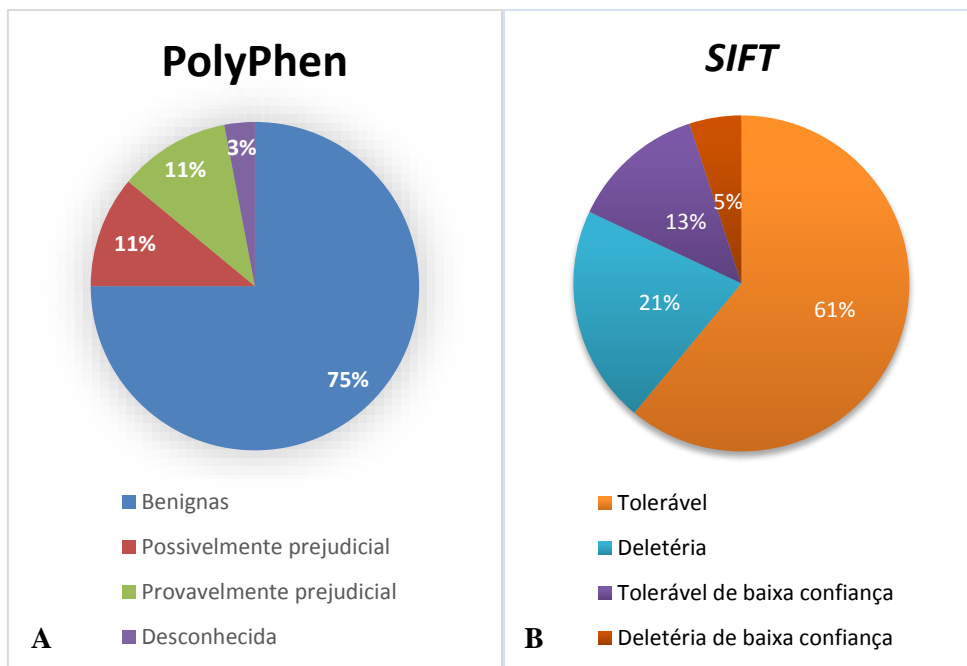


Figura 5 **A** Resultados das análises das variantes identificadas, de acordo aos programas *PolyPhen* (ADZHUBEI et al., 2010); **B** *SIFT* (NG; HENIKOFF, 2001).

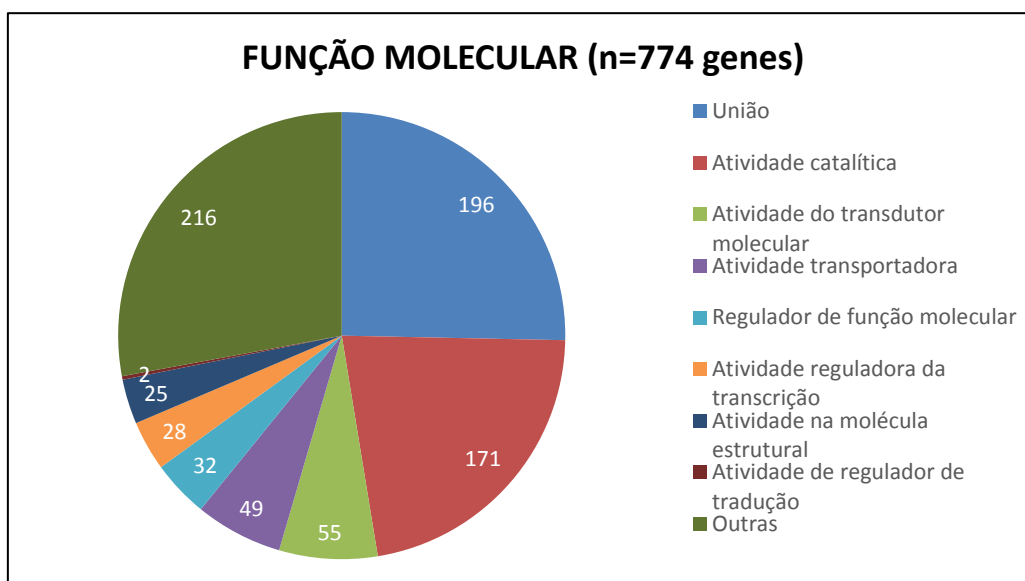


Figura 6 Análise da função molecular dos genes obtidos no estudo, através da ferramenta *PANTHER Classification System* (MI et al., 2019).

Tabela 5 Análise das vias de sinalizações maiormente representadas a partir dos 774 genes associados a variantes de baixo, moderado e alto impacto na estrutura aminoacídica. Utilizou-se a ferramenta *PANTHER Classification System* (MI et al., 2019).

| Via de sinalização* | Total de Genes |
|--|-----------------------|
| Via de sinalização <i>Wnt</i> | 15 |
| Via de sinalização Integrinas | 14 |
| Via de sinalização da proteína G heterotrimérica | 13 |
| Via do receptor do hormônio liberador de gonadotrofina | 10 |
| Via da presenilina da doença de Alzheimer | 9 |
| Via de sinalização da caderina | 8 |
| Angiogênese | 7 |
| Doença de <i>Huntington</i> | 7 |
| Mapa de sinalização CCKR | 6 |
| Via de sinalização FGF | 6 |
| Via de sinalização do receptor 1 e 3 da acetilcolina muscarínica | 6 |
| Via de sinalização do receptor de acetilcolina nicotínico | 6 |
| Via de sinalização do receptor <i>Toll</i> | 6 |
| Outras vias** | < 5 |

* Um total de 77 vias foram relacionadas aos 774 genes identificados no estudo; **Outras vias de sinalizações relacionadas (Total = 63) agrupam cinco ou menos genes.

Foi estimado que um total de 42 variantes encontradas produzem um efeito de alto impacto nas estruturas aminoacídicas de 39 genes. A lista destas variantes com informações detalhadas pode ser observada na **Tabela S6 (Anexo B)**, em que se destaca 6 variantes não descritas. Cada variante afeta um gene, mas o *CELA1* e *TP53I11* estão afetados por 3 e 2 variantes, respectivamente. Do total das variantes que causam alto impacto, 23 (59%) foram exônicas, 21 (50%) foram SNV e 23 (55%) incluem *frameshift_variant* (nomenclatura atribuída por *Esembl* (MCLAREN et al., 2016)) como consequência funcional calculada. Em relação a localização, os cromossomos 11, 12 e 19 concentraram um maior número de variantes de alto impacto (n= 6, 6 e 5 variantes, nesta ordem).

A **Tabela S7 (Anexo B)** caracteriza cada uma das variantes que produzem efeito de alto impacto, em relação aos dados clínicos e diagnóstico histopatológico das amostras

nas quais estas foram identificadas. A maioria das variantes (n=34, 81%) está presente no mínimo em 2 amostras e um total de 14 variantes (33%) foram observadas em pelo menos metade das amostras, estas impactam os genes: *ACTN2*, *GLYCTK*, *TP53III1*, *WNK1*, *CELA1*, *OR6C1*, *SIX1*, *GABRG3*, *KRT24*, *MRPL27* e *LAMA5*. A variante nº 26, que impacta o gene *SIX1*, está presente em todas as amostras.

A **Tabela 6** resume o total de variantes de alto impacto por cada amostra, onde é possível observar que as amostras 9 e 10, com diagnóstico de DAG, acumulam um maior número de variantes. A média de variantes observadas nas amostras de DAG foi mais alta que a do grupo de DBG ($p > 0,05$).

Tabela 6 Total de variantes classificadas, de acordo ao VEP (MCLAREN et al., 2016), como de alto impacto (n= 42) por amostra.

| Amostra | Tipo DE | Var. Het. | Var. Hom. | Total Var. |
|-----------|------------|-----------|-----------|------------|
| 1 | DBG | 6 | 9 | 15 |
| 2 | DBG | 7 | 10 | 17 |
| 3 | DBG | 6 | 15 | 21 |
| 4 | DBG | 5 | 9 | 14 |
| 5 | DBG | 8 | 8 | 16 |
| 6 | DBG | 5 | 12 | 17 |
| \bar{x} | DBG | 6 | 11 | 17 |
| 7 | DAG | 7 | 11 | 18 |
| 8 | DAG | 15 | 6 | 21 |
| 9 | DAG | 11 | 17 | 28 |
| 10 | DAG | 13 | 13 | 26 |
| \bar{x} | DAG | 11 | 12 | 23 |

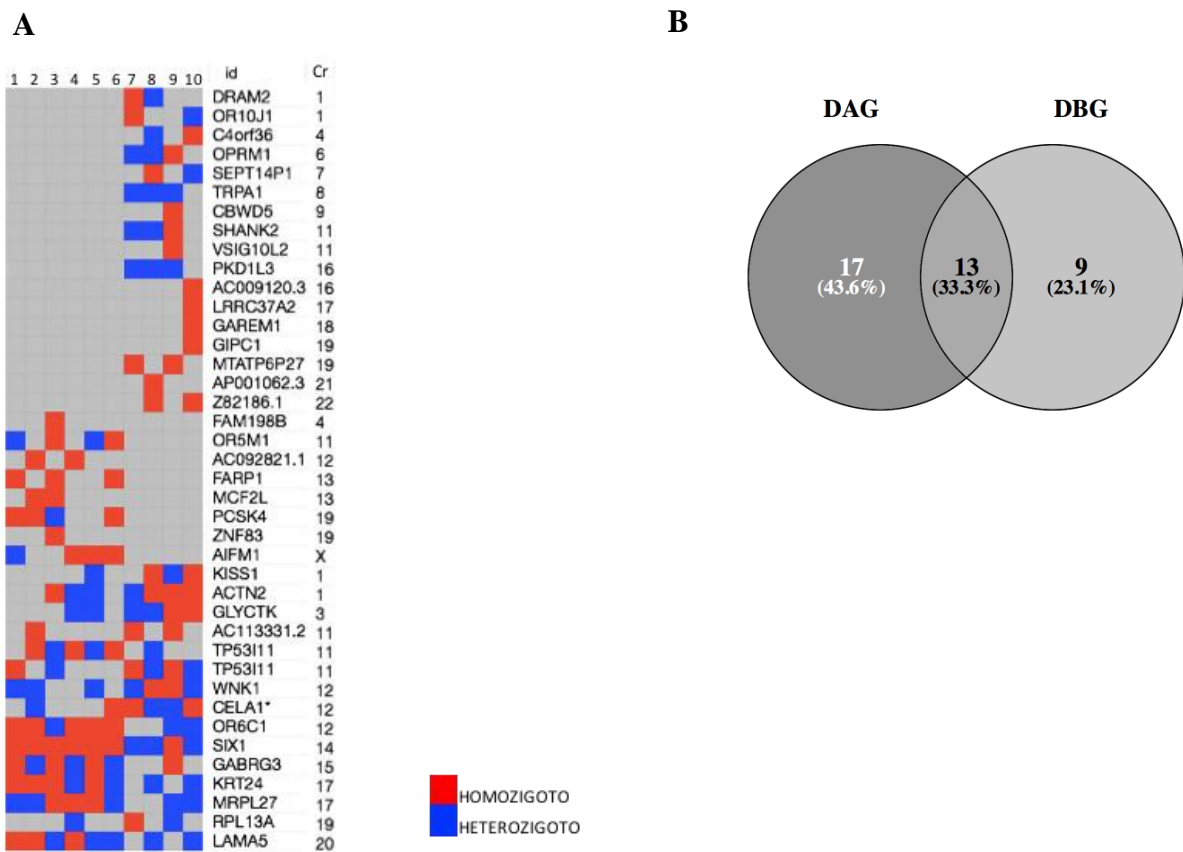
DE, Displasia Epitelial Oral; VAR., Variante; DBG, Displasia de Baixo Grau; DAG, Displasia de Alto Grau; Het., Heterozigoto; Hom., Homozigoto.

Na **Figura 7A** é possível observar os genes acometidos pelas variantes que produzem efeito de alto impacto aminoacídico, localização cromossômica, amostras afetadas e o genótipo. Do total de genes identificados no estudo, impactados por variantes

observadas nas amostras, 23,1% destes foram observados exclusivamente nas amostras de DBG e 43,6% somente nas amostras de DAG (**Figura 7B**). Os genes com variantes identificados nas amostras DAG foram: *DRAM2*, *OR10J1*, *C4orf36*, *OPRM1*, *SEPT14P1*, *TRPA1*, *CBWD5*, *SHANK2*, *VSIG10L2*, *PKD1L3*, *AC009120.3*, *LRRC37A2*, *GAREM1*, *GIPC1*, *MTATP6P27*, *AP001062.3*, *Z82186.1*. Já aqueles observados em amostras de DBG foram: *FAM198B*, *OR5M1*, *AC092821.1*, *FARP1*, *MCF2L*, *PCSK4*, *ZNF83*, *AIFM1*. Um total de 13 genes (33,3%) alterados foram identificados em amostras de DAG e DBG (**Figura 7B**), são eles: *KISS1*, *ACTN2*, *GLYCTK*, *AC113331.2*, *TP53I11*, *WNK1*, *CELA1*, *OR6C1*, *SIX1*, *GABRG3*, *KRT24*, *MRPL27*, *RPL13A*, *LAMA5*.

A partir das informações genômicas previamente reportadas na literatura (AGRAWAL et al., 2011; STRANSKY et al., 2011; PICKERING et al., 2013; CANCER GENOME ATLAS NETWORK, 2015) e registradas na plataforma *cBioPortal* (MEMORIAL SLOAN KETTERING CANCER CENTER, 2019), foi realizado uma comparação dos 39 genes identificados no presente estudo (**Tabela S6, Anexo B**) com genes mutados em todos tipos de canceres de cabeça e pescoço (n=16807 genes), em amostras de carcinoma de células escamosas da região de cabeça e pescoço (HNSSC, n=16099 genes) e genes mutados reportados em COCE (n= 2656 genes). Os genes coincidentes em cada uma das comparações podem ser observados na **Tabela 8**.

Avaliando cada um dos genes em que foram identificados variantes que produzem efeito de alto impacto do presente estudo, com bases de dados de sequência e anotação de proteínas, além de outras referências da literatura, foi encontrado que o *MCF2L* (UNIPROT CONSORTIUM, 2019) e *SIX1* (COLETTA et al., 2008; XIE et al., 2018) representam oncogenes e o *KISS1* é descrito como um gene supressor de metástase (MITCHELL et al., 2007; TIAN et al., 2018).



* Três variantes estão associadas ao gene *CELA1* (Tabela S6, Anexo 3). As amostras (2, 6, 7, 8, 9 e 10) apresentam estas três variantes, com o mesmo genótipo esquematizado na figura.

Figura 7 **A** Distribuição dos genes mais severamente alterados em relação ao grupo, as amostras e genótipo (homo ou heterozigoto); **B** Porcentagem de genes específicos e compartilhados nos grupos DAG e DBG.

Tabela 8 Resultado das comparações de genes com variantes que produzem efeito de alto impacto, identificados em amostras de DEO do presente estudo (n=39 genes), com os grupos de genes mutados em amostras de HNC (n= 16807 genes), HNSCC (n=16099 genes) e de COCE (n=2656 genes).

| Genes com variantes em DEO | Genes mutados em HNC* | Genes mutados em HNSCC* | Genes mutados em COCE* |
|----------------------------|-----------------------|-------------------------|------------------------|
| <i>AC009120.3</i> | | | |
| <i>AC092821.1</i> | | | |
| <i>AC113331.2</i> | | | |
| <i>ACTN2</i> | X | X | |
| <i>AIFM1</i> | X | X | X |
| <i>AP001062.3</i> | | | |
| <i>C4orf36</i> | | | |
| <i>CBWD5</i> | X | X | |
| <i>CELA1</i> | | | |
| <i>DRAM2</i> | X | X | X |
| <i>FAM198B</i> | | | |
| <i>FARP1</i> | X | X | X |
| <i>GABRG3</i> | X | X | X |
| <i>GAREM1</i> | X | X | X |
| <i>GIPC1</i> | X | X | X |
| <i>GLYCTK</i> | X | X | |
| <i>KISS1</i> | | | |
| <i>KRT24</i> | X | X | X |
| <i>LAMA5</i> | X | X | X |
| <i>LRRC37A2</i> | X | X | X |
| <i>MCF2L</i> | X | X | X |
| <i>MRPL27</i> | X | X | X |
| <i>MTATP6P27</i> | | | |
| <i>OPRM1</i> | X | X | X |
| <i>OR10J1</i> | X | X | |
| <i>OR5M1</i> | X | X | |
| <i>OR6C1</i> | | | |
| <i>PCSK4</i> | X | X | |
| <i>PKD1L3</i> | | | |
| <i>RPL13A</i> | X | X | |
| <i>SEPT14P1</i> | | | |
| <i>SHANK2</i> | X | X | X |
| <i>SIX1</i> | X | X | |
| <i>TP53I11</i> | X | X | X |
| <i>TRPA1</i> | X | X | X |
| <i>VSIG10L2</i> | | | |
| <i>WNK1</i> | X | X | X |
| <i>Z82186.1</i> | | | |
| <i>ZNF83</i> | X | X | X |

DEO, Displasia epitelial oral; HNC, Câncer de cabeça e pescoço; HNSCC, Carcinoma de células escamosas da região de cabeça e pescoço; COCE, Carcinoma oral de células escamosas. Em laranja destaca-se genes mutados reportados em HNSCC e em amarelo genes mutados em carcinomas de células escamosas da cavidade oral e de outras regiões de cabeça e pescoço. *Informações de dados genômicos obtidos através da plataforma *cBioPortal* (MEMORIAL SLOAN KETTERING CANCER CENTER, 2019)

6 DISCUSSÃO

Com a escassez de publicações na literatura que utilize a técnica de NGS em amostras de DEO, o presente estudo vem a contribuir com a descrição de alterações genômicas no exoma de 10 amostras de pacientes chilenos e brasileiros, com lesões diagnosticadas como DEO de baixo e alto graus. Pela dificuldade de obtenção de tecidos viáveis e pelo tamanho pequeno das amostras dos DOPM para análise do genoma ou exoma completo (GARNIS et al., 2009; WOOD et al., 2015, 2017), estudos prévios que utilizaram esta metodologia usaram amostras de COCE e selecionaram áreas com DEO adjacentes ao tumor (WOOD et al., 2015, 2017). Como desconhecemos o perfil genômico dos DOPM, existe a possibilidade que tecidos com áreas de DEO, extraídos de amostras de COCE, possam apresentar alterações genéticas não próprias das lesões com DEO. Sendo assim, apesar das dificuldades encontradas na obtenção de amostras com qualidade e quantidade suficientes para a realização do NSG, o presente estudo incluiu amostras de pacientes diagnosticados clinicamente com diferentes tipos de leucoplasias que foram compatíveis a DEO de baixo e alto graus. Com amostras de DOPM, o presente estudo apresenta a vantagem de ter boa representatividade na correlação das informações genômicas encontradas em DEO com os aspectos clínicos.

Mesmo com um número reduzido de amostras, o perfil clínico dos pacientes foi representativo ao reportado na literatura (SPEIGHT; KHURRAM; KUJAN, 2018; WARNAKULASURIYA, 2018), com uma média de idade de 54 anos e uma maior quantidade de indivíduos fumantes e que consumem álcool. Além destas características, as localizações e diagnósticos clínicos das lesões foram diversas, sendo o maior número destas localizadas na língua e diagnosticadas como leucoplasia homogênea. Apesar de estar descrito uma maior frequência de leucoplasias em homens (WARNAKULASURIYA, 2018), no presente estudo predominaram amostras do sexo feminino.

Como esperado, um número alto de variantes já descritas e novas foram identificadas neste estudo, o que pode ser explicado pela variabilidade do genoma humano (STRANGER et al., 2007). No entanto, somente uma porcentagem inferior a 1% destas tem consequências localizadas em regiões de *splicing* (*splice_donor_variant* ou *splice_acceptor_variant*), que podem resultar em um códon de parada prematuro (*stop_lost*), um códon de início alterado (*start_lost*) ou conseguem alterar o marco de

leitura (*frameshift_variant*). Mudanças nestas regiões produzem um alto impacto na proteína, trazendo consequências como a perda da sua função ou alterações na estrutura, que associam-se a processos patológicos (MCLAREN et al., 2016).

Os cromossomos 1, 2 e 3 foram os que acumularam a maior quantidade de variantes, isso provavelmente ocorre porque estes cromossomos tem um maior tamanho, possuem uma maior quantidade de genes, aumentando a possibilidade de acumular alterações (ALBERTS et al., 2017). Prevendo o impacto das variantes nas proteínas, as porções centrais, concentraram uma menor quantidade de variantes e por outro lado, a porção final teve um maior número de variantes. Como a grande maioria das variantes identificadas no estudo tem um efeito benigno nas proteínas, é esperado encontrar alterações nas porções terminais destas, pois modificações nestas regiões normalmente não produzem alterações funcionais. As porções centrais das proteínas correspondem a regiões pericentroméricas, acumulando um menor número de genes exposto a modificações. (ALBERTS et al., 2017).

A maiorias das variantes identificadas no presente estudo foram tipo SNV, resultado que está de acordo com a literatura que descreve esta alteração como a mais comumente encontrada na avaliação de genoma ou exoma completo (SHIGEMIZU et al., 2013). Apesar das variantes INDEL tenham sido menos frequentes que as SNV, estas normalmente são extremamente importantes no NGS, pois estão implicadas em muitas doenças constitucionais e oncológicas (SEHN, 2015). A maior média significativa de variantes SNV foi observada no grupo das DAG, sendo a amostra nº 10 (DAG) associada ao maior número de variantes deste tipo. Este resultado vem a ser sustentado por trabalhos que mostram as lesões que antecedem o câncer como sendo caracterizadas por mudanças progressivas na sequência de DNA, expressão gênica, estruturas das proteínas e rearranjo microscópico (CAMPBELL et al., 2016; SPIRA et al., 2017; HU et al., 2018; SRIVASTAVA et al., 2018). Além disso, concorda com o estudo prévio que observou um menor número de mutações nas amostras de DBG quando comparado as DAG e COCE (WOOD et al., 2017).

A partir da análise de ferramentas que preveem possíveis impactos de uma substituição de aminoácidos na estrutura e função de uma proteína humana, usando considerações físicas e comparativas diretas (NG; HENIKOFF, 2001; ADZHUBEI et al., 2010; SUDMANT et al., 2015; MCLAREN et al., 2016), foram identificadas 1069 variantes que causam efeito de alto, moderado e baixo impacto nas estruturas

aminoacídicas de 774 genes das amostras de DEO do presente estudo. Uma maior quantidade destes genes estão relacionados a funções metabólicas de união e atividades catalíticas, e participam das vias de sinalização *Wnt* e integrina. Já foi descrito que a desregulação funcional da via de sinalização *Wnt* pode promover o desenvolvimento e progressão do câncer bucal, portanto esta é uma via de interesse para criar estratégias de tratamento para este câncer (GONZALEZ-MOLES et al., 2014; SHIAH; SHIEH; CHANG, 2016). Em relação as integrinas, principais componentes da adesão celular, tem sido reportado a implicação destas em quase todas as etapas da progressão do câncer desde o desenvolvimento do tumor primário até a metástase (HAMIDI; IVASKA, 2018).

No total de amostras de DEO foram identificadas 42 variantes que devem produzir um alto impacto nas estruturas aminoacídicas de 39 genes. A maioria das variantes são exômicas, de deleção e com consequências funcionais diversas, sendo a mais frequente uma mudança no marco de leitura (*frameshift_variant*). As variantes intrônicas, que representam um alto número também, são quase todas do tipo SNV. Apesar destas últimas estarem localizadas em regiões não codificantes, as consequências funcionais provocadas por estas variantes se localizam nas 2 bases das extremidades 3' ou 5' do intrón (*splice_acceptor_variant* e *splice_donor_variant*, respectivamente). A região extremidade 3' corresponde a secção do RNA mensageiro (RNAm) que segue imediatamente o códon de terminação da tradução e 5' é a região de um RNAm que está acima do códon de iniciação (ALBERTS et al., 2017). Quaisquer erro nestas regiões pode levar à remoção incorreta de intróns, causando alterações no quadro de leitura aberto, o que leva à diminuição do nível do RNAm específico e, portanto, à privação do nível de proteína, podendo resultar em metabolismo e/ou função celular aberrante (CHABOT; SHKRETA, 2016; ABRAMOWICZ; GOS, 2018).

O presente estudo identificou seis novas variantes, sendo três SNV que levam a consequências funcionais nas regiões de *splice_acceptor* e *splice_donor*, e a outra metade é de deleção, que levam a mudanças do marco de leitura. Os genes *GAREM1*, *GIPC1*, *LRR37A2* relacionados a estas variantes, estão associados a mutações descritas em carcinoma de células escamosas de cabeça e pescoço e COCE (AGRAWAL et al., 2011; STRANSKY et al., 2011; PICKERING et al., 2013; CANCER GENOME ATLAS NETWORK, 2015). O gene *CBWS*, também associado a uma variante nova, foi previamente identificado mutado em carcinomas de células escamosas de cabeça e pescoço (AGRAWAL et al., 2011; STRANSKY et al., 2011; PICKERING et al., 2013;

CANCER GENOME ATLAS NETWORK, 2015).

No presente estudo, assim como foi observado que as amostras de DAG acumularam uma maior quantidade de variantes no total (baixo, moderado e alto impactos), uma mesma tendência foi observada ao considerar somente as variantes de alto impacto, no entanto esta associação não foi significativa. Concordando assim com o estudo prévio que observou um menor número de mutações nas amostras de DBG quando comparado as DAG (WOOD et al., 2017). Em relação as características clínicas dos pacientes, não foi possível estabelecer uma relação com cada uma das variantes encontradas, pois o tamanho da amostra foi reduzida para este tipo de correlação. Como até o momento não existem estudos similares realizados com amostras clínicas representativas de DEO, a correlação das características clínicas e variações genômicas ainda não foi estabelecida na literatura.

No presente estudo foram identificadas variantes em genes exclusivos para os grupos DBG e DAG. O grupo de DAG acumula um maior número de variantes de alto impacto associadas, sendo algumas destas também identificadas em DBG. No grupo DAG, observa-se que as variantes que afetam os genes *OPRM1*, *TRPA1*, *SHANK2* e *PKDIL3* estão presentes nas mesmas amostras (7,8 e 9). Já no grupo DBG as variantes dos genes *OR5M1*, *FARP1*, *PCSK4*, *AIFM1* foram identificadas em mais da metade das amostras. Considerando variantes compartilhadas nos dois grupos, os genes afetados na maioria das amostras foram: *ACTN2*, *GLYCTK*, *TP53I11*, *WNK1*, *CELA1*, *OR6C1*, *SIX1*, *GABRG3*, *KRT24*, *MRPL27* e *LAMA5*. Destes, destaca-se que as alterações dos genes *OR6C1*, *GABRG3*, *KRT24*, *MRPL27* e *LAMA5* (Grupo de Genes 1), mesmo presente em algumas amostras de DAG, foram identificados em todas as amostras de DBG. Por outro lado, alterações nos genes *ACTN2*, *TP53I11*, *GLYCTK* e *CELA1* (Grupo de Genes 2) estavam presentes em todas as amostras de DAG. De acordo aos dois últimos resultados, sugere-se os grupos de genes 1 e 2, para serem usados em futuros estudos, como possíveis marcadores de DEO.

Em relação ao gene *CELA1*, é importante considerar que neste estudo foram observadas três variantes que provocam alto impacto neste gene, e que estas foram identificadas nas mesmas amostras (2,6,7,8,9 e 10), com o mesmo tipo de herança para cada uma destas (homozigota: 6,7 e 10; heterozigota: 2,8 e 9). O gene *CELA1*, também conhecido como *ELAI*, está localizado na região cromossômica 12q13, próxima a um locus para da doença ceratoderma palmoplantar não epidermolítica difusa, teve

expressão detectada em queratinócitos primários humanos cultivados (TALAS et al., 2000). No pâncreas humano a elastase-1 é funcionalmente silenciosa, mas foi encontrado que altos níveis desta proteína pré-operatória podem ser um fator de risco para recorrência pós-operatória em pacientes com neoplasias neuroendócrinas pancreáticas bem diferenciadas (NANNO et al., 2018).

Há muito tempo se conhece que a distribuição de mutações no genoma não é completamente aleatória. No presente estudo, a observação de variantes que afetam o mesmo gene (*CELA1*) de um mesmo grupo de amostras pode ser explicado pelo fenômeno, que ainda não é totalmente compreendido, e é denominado *mutation showers*. Este está associado há presença de múltiplas mutações simultâneas ocorrendo em um mesmo gene ou em pequenas regiões dentro dos cromossomos (CHAN; GORDENIN, 2015). Ainda não existem muitos estudos que explicam ou associam estas alterações com o câncer (CHAN; GORDENIN, 2015; ZOU et al., 2017), no entanto a partir da análise de catálogos de mutações disponíveis, foi identificado mutagênese agrupada em mieloma múltiplo, tumores de próstata e cabeça e pescoço (CHAN; GORDENIN, 2015).

Para o gene *TP53III1* foram identificadas duas variantes, que causam alto impacto neste gene, em diferentes amostras (amostras: 2,3,4,5,6 e 7; 1,3,7,8,9 e 10), sendo que as amostras 3 e 7 apresentaram as duas variantes. O gene *TP53III1* (tumor protein 53-inducible protein 11) é candidato a ser gene supressor de tumor, pertencente a família do gene da proteína induzida por p53 (WU et al., 2009). Até agora, os estudos sobre *TP53III1* mostraram no efeito positivo na apoptose, atuando um supressor tumoral em células cancerígenas hepáticas e epiteliales mamárias (RAMACHANDRAN et al., 2005; WU et al., 2009; WANG et al., 2018).

Comparando os genes mais severamente afetados identificados no presente estudo com os genes mutados reportados no estudo de Wood et al. (WOOD et al., 2017), ainda que sejam variantes diferentes, foi encontrado coincidência com o gene *WNKI* mutado somente em amostras de DEO, gene *MCF2L* mutado em amostras de DEO e COCE, e os genes *LAMA5*, *FARPI* e *SHANK2* mutados exclusivamente em amostras de COCE (WOOD et al., 2017). Esta coincidência observada somente em dois genes mutados em DEO pode ser explicada porque, diferente do presente estudo, que utilizou amostras representativas do ponto de vista clínico e histopatológico, o estudo de Wood et al. (WOOD et al., 2017) extraiu áreas de DEO a partir de amostras de COCE, o que pode levar a uma maior probabilidade de apresentar diferenças do ponto de vista molecular.

De fato o mesmo estudo, apesar de reportar que existiram menos mutações nas amostras de DEO que as de COCE, a grande maioria das mutações encontradas em DEO foram observadas também em COCE (WOOD et al., 2017).

Assim, como está estabelecido o *Cancer Genome Atlas* (TCGA) para consultas sobre a diversidade genômica de varios tipos de cânceres, foi iniciado em 2016 o projeto de criação do *Pre-Cancer Genome Atlas* (PCGA), no entanto, até o momento não está disponível em plataformas os genes mutados para nenhuma das lesões que antecedem os diversos tipos de cânceres, incluindo as DEO (CAMPBELL et al., 2016; SPIRA et al., 2017; SRIVASTAVA et al., 2018). Baseando-se no TCGA (AGRAWAL et al., 2011; STRANSKY et al., 2011; PICKERING et al., 2013; CANCER GENOME ATLAS NETWORK, 2015) foi observado que 64% dos genes identificados no presente estudo com variantes de alto impacto, também foram encontrados em carcinomas de células escamosas de cabeça e pescoço. Deste total de genes comuns aos do presente estudo, 8 não foram identificados em COCE.

Importante ressaltar que dos 11 genes alterados, identificados em mais da metade das amostras deste estudo, 9 também estão mutados em carcinomas de células escamosas de cabeça e pescoço (*ACTN2*, *GLYCTK*, *TP53I11*, *WNK1*, *SIX1*, *GABRG3*, *KRT24*, *MRPL27* e *LAMA5*) e destes genes, 6 (*TP53I11*, *WNK1*, *GABRG3*, *KRT24*, *MRPL27* e *LAMA5*) também estão mutados em COCE. O estudo de Wood et al (WOOD et al., 2017) adiciona *SHANK2* e *FARP1*, mutados em COCE, e *MFC2L*, mutado em COCE e DEO, que também estão alterados nas amostras do presente estudo e outros trabalhos em carcinoma de células escamosas de cabeça e pescoço (AGRAWAL et al., 2011; STRANSKY et al., 2011; CANCER GENOME ATLAS NETWORK, 2015) e COCE (AGRAWAL et al., 2011; STRANSKY et al., 2011; ; PICKERING et al., 2013; CANCER GENOME ATLAS NETWORK, 2015). Semelhanças no perfil genômico de lesões com displasia epitelial e câncer já foram descritas em epitélio intestinal, de mama, cérebro, rim, pulmão e pele, mostrando que o processo mutacional pode causar uma evolução clonal de células normais a neoplásicas (CAMPBELL et al., 2016; SPIRA et al., 2017; HU et al., 2018). No entanto, Wood et al., 2015 (WOOD et al., 2015) observaram que cinco amostras de DEO tinham heterogeneidade subclonal com o COCE e sugerem que mudanças mutacionais em etapas prévias ao câncer não preveem o início da invasão (WOOD et al., 2015).

Um achado importante do presente estudo foi em relação a alteração do gene *SIX1*,

esta variante foi identificada em todas amostras do estudo, com genótipo homozigoto associado a todas amostras de DBG. Por outro lado, em DAG foi identificado em três dos quatro pacientes um genótipo heterozigoto. Este gene, localizado na região cromossômica 14q23.1, foi considerado um oncogene em cânceres de mama, ovário, colo do útero, carcinoma hepatocelular, rabdomiossarcoma, tumor de *Wilms* e câncer colorretal (COLETTA et al., 2008; WU et al., 2015; XIE et al., 2018). Além disso, *SIX1* foi previamente encontrado em carcinoma de células escamosas de cabeça e pescoço (AGRAWAL et al., 2011; STRANSKY et al., 2011; CANCER GENOME ATLAS NETWORK, 2015). De acordo a la relevância deste gene em um grupo extenso de neoplasias, pode-se sugerir o estudo deste gene como possível marcador de DEO.

O gene *MCF2L* considerado um oncogene (UNIPROT CONSORTIUM, 2019), além de estar alterado em amostras de DBG do presente estudo, foi também reportado mutado em DEO (WOOD et al., 2017), COCE (AGRAWAL et al., 2011; STRANSKY et al., 2011; PICKERING et al., 2013; CANCER GENOME ATLAS NETWORK, 2015; WOOD et al., 2017) e outros carcinomas de cabeça e pescoço (AGRAWAL et al., 2011; STRANSKY et al., 2011; CANCER GENOME ATLAS NETWORK, 2015). *MCF2L*, localizado na região cromossômica 13q34, foi previamente relacionado com câncer de ovário e endométrio (UNIPROT CONSORTIUM, 2019). Apesar de mutações em *MCF2L* terem sido reportadas em varias neoplasias, inclusive em DEO, no presente estudo só foi observado a variante em duas das seis amostras de DBG. No entanto, como estas amostras tinham diferente origem (Brasil e Chile) e foram homozigotas para a variante, sugere-se considerar este gene para avaliação, em relação a presença de alterações e como marcador diagnóstico em futuras amostras de DEO.

O gen *KISS1*, localizado na região cromossômica 1q32.1, foi reportado como gene supressor de metástase em câncer de mama e melanoma (MITCHELL et al., 2007; TIAN et al., 2018). Este gene que desenvolve possivelmente função na regulação de da adesão célula-matriz (MITCHELL et al., 2007) , foi encontrado em três das quatro amostras de DAG e uma de DBG, mas não foi reportado mutado na literatura em DEO, COCE ou outros carcinomas de cabeça e pescoço. Estando presente em quase todas amostras de DAG e sendo reportado como supressor de metástase, é possível sugerir este gene como marcador de progressão em DEO.

Desde 2009 um estudo já mostrava que DBGs que avançavam a COCE, apesar das semelhanças histológicas com outras DBG que não progrediam a câncer, apresentavam

perfis genômicos completamente diferentes (GARNIS et al., 2009). No entanto, dez anos depois destas descobertas, o diagnóstico histopatológico, com presença de DEO nos distintos graus, segue sendo o exame complementar padrão no delineamento do risco de progressão e decisões de tratamentos para DOPMs. Porém, este método continua sendo subjetivo e com uma baixa concordância de diagnóstico entre os patologistas. Além disso, nem todas DEO, independente do grau, avançam a COCE e esta informação não pode ser identificada em uma análise histopatológica. Diante deste panorama atual, esta pesquisa fornece informações relevantes na caracterização genômica de DEO. Apesar do número reduzido de amostras, ao contar com uma amostra de origem populacional heterogênea e utilizar uma metodologia de avaliação genômica em profundidade, que atualmente tem uma taxa de erro extremamente reduzida, foi possível estabelecer resultados de alta confiabilidade. Por outro lado, é importante complementar as principais informações encontradas em um maior número de amostras, através de estudos em PCR e imunohistoquímica.

7 CONCLUSÃO

A partir da avaliação do exoma completo de 10 amostras de DEO de indivíduos brasileiros e chilenos, um alto número de variantes foram encontradas, de forma mais frequente em DAG e sendo a maioria do tipo SNV. Destas variantes, menos de 1% foram previstas com possibilidades de produzir alterações de baixo, moderado e alto impacto nas estruturas aminoacídicas.

Identificou-se 42 variantes capazes de produzir alto impacto deletério em 39 genes. O número maior de variantes se deu, porque os genes *TP53I11* e *CELA1* foram associados a duas e três variantes, respectivamente. Estas variantes foram do tipo *frameshift* e SNV em quase igual frequência, a maioria era exônica, mas aquelas intrônicas tinham consequências do tipo *splice acceptor variant* ou *splice donor variant*. As amostras de DAG mostraram uma tendência a acumular maior quantidade das variantes de alto impacto deletério, mas pelo baixo número de amostras, não foi possível estabelecer uma associação das características clínicas consideradas com estas variantes.

Seis novas variantes SNV e de deleção, consideradas de alto impacto deletério, foram identificadas em relação aos genes: *FAM198B*, *CBWD5*, *PKD1L3*, *LRRC37A2*, *GAREM1* e *GIPC1*.

As variantes encontradas nos genes: *DRAM2*, *OR10J1*, *C4orf36*, *OPRM1*, *SEPT14P1*, *TRPA1*, *CBWD5*, *SHANK2*, *VSIG10L2*, *PKD1L3*, *AC009120.3*, *LRRC37A2*, *GAREM1*, *GIPC1*, *MTATP6P27*, *AP001062.3*, *Z82186.1*, foram observadas exclusivamente em amostras do grupo DAG. Já em relação ao grupo DBG, as variantes exclusivas de amostras deste grupo foram identificadas nos genes: *FAM198B*, *OR5M1*, *AC092821.1*, *FARP1*, *MCF2L*, *PCSK4*, *ZNF83*, *AIFM1*. Variantes dos genes *OR6C1*, *GABRG3*, *KRT24*, *MRPL27* e *LAMA5* foram identificadas em todas as amostras de DBG e variantes dos genes: *ACTN2*, *TP53I11*, *GLYCTK* e *CELA1* estavam presentes em todas as amostras de DAG. Sugere-se o estudo dos nove últimos genes como possíveis marcadores de DEO.

Dos genes identificados com variantes de alto impacto no presente estudo, os genes *SIX1* e *MCF2L* são considerados oncogenes e o *KISS1* como supressor de metástase. A variante que produz alto impacto no gene *SIX1* foi identificada em todas as amostras do estudo. Variantes nos genes *WNK1* e *MCF2L* foram previamente reportados na literatura

em DEO e variantes em *TP53I11*, *WNK1*, *GABRG3*, *KRT24*, *MRPL27*, *LAMA5*, *SHANK2*, *MCF2L* e *FARPI* foram descritas em COCE.

REFERÊNCIAS

- 1000 GENOMES PROJECT CONSORTIUM, ABECASIS, G. et al. An integrated map of genetic variation from 1,092 human genomes. **Nature**, v. 491, n. 7422, p. 56–65, 2012.
- ABRAMOWICZ, A.; GOS, M. Splicing mutations in human genetic disorders: examples, detection, and confirmation. **J Appl Genet**, v. 59, n. 3, p. 253–268, 2018.
- ADZHUBEI, I. et al. A method and server for predicting damaging missense mutations. **Nat Methods**, v. 7, n. 4, p. 248–9, 2010.
- AGRAWAL, N. et al. Exome sequencing of head and neck squamous cell carcinoma reveals inactivating mutations in NOTCH1. **Science**, v. 333, n. 6046, p. 1154–7, 2011.
- AKEN, B. et al. The ensembl gene annotation system. **Database (Oxford)**, v. 2016. pii:, 2016.
- AL-HEBSHI, N. N. et al. Exome sequencing of oral squamous cell carcinoma in users of Arabian snuff reveals novel candidates for driver genes Nezar. **Int J Cancer**, v. 139, n. 2, p. 363–372, 2016.
- ALAIZARI, N. et al. Meta-analysis of the predictive value of DNA aneuploidy in malignant transformation of oral potentially malignant disorders. **J Oral Pathol Med**, v. 47, n. 2, p. 97–103, 2018.
- ALBERTS, B. et al. **Biologia Molecular da Célula**. 6^a ed. Porto Alegre: ARTMED EDITORA S.A, 2017.
- ANDREWS, S. A Quality Control tool for High Throughput Sequence Data. <http://www.bioinformatics.babraham.ac.uk/projects/fastqc/>, 2010.
- AWADALLAH, M. et al. Management update of potentially premalignant oral epithelial lesions. **Oral Surg Oral Med Oral Pathol Oral Radiol**, v. 125, n. 6, p. 628–636, 2018.
- AXÉLL, T. et al. Oral white lesions with special reference to precancerous and tobacco-related lesions: conclusions of an international symposium held in Uppsala, Sweden, May 18-21, 1994. International Collaborative Group on Oral White Lesions. **J Oral Pathol Med**, v. 25, n. 2, p. 49–54, 1996.
- BHATIA, A.; BURTNESS, B. Novel molecular targets for chemoprevention in malignancies of the head and neck. **Cancers**, v. 9, n. 9, 2017.
- BHATTACHARYA, A. et al. Two distinct routes to oral cancer differing in genome instability and risk for cervical node metastasis. **Clin Cancer Res**, v. 17, n. 22, p. 7024–34, 2011.
- BRADLEY, G. et al. Abnormal DNA content in oral epithelial dysplasia is associated with increased risk of progression to carcinoma. **Br J Cancer**, v. 6, n. 8, p. 822–31, 2010.
- BRAY, F. et al. Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. **CA: A Cancer Journal for Clinicians**, v. 00, n. 00, p. 1–31, 2018.
- CALIFANO, J. et al. Genetic progression model for head and neck cancer: Implications for field cancerization. **Cancer Res**, v. 56, n. 11, p. 2488–92, 1996.
- CAMPBELL, J. D. et al. The Case for a Pre-Cancer Genome Atlas (PCGA). **Cancer Prev Res**, v. 9, n. 2, p. 119–124, 2016.

- CANCER GENOME ATLAS NETWORK, C. Comprehensive genomic characterization of head and neck squamous cell carcinomas. **Nature**, v. 517, n. 7536, p. 576–82, 2015.
- CHABOT, B.; SHKRETA, L. Defective control of pre-messenger RNA splicing in human disease. **J Cell Biol**, v. 212, n. 2, p. 13–27, 2016.
- CHAN, K.; GORDENIN, D. Clusters of Multiple Mutations: Incidence and Molecular Mechanisms. **Annu Rev Genet**, v. 49, p. 243–267, 2015.
- CHAU, L. et al. Topical agents for oral cancer chemoprevention: A systematic review of the literature. **Oral Oncology**, v. 67, p. 153–159, 2017.
- CINGOLANI, P. et al. Using drosophila melanogaster as a model for genotoxic chemical mutational studies with a new program, snpsift. **Front Genet**, v. 3, p. 35, 2012.
- COLETTA, R. et al. Six1 overexpression in mammary cells induces genomic instability and is sufficient for malignant transformation. **Cancer Res**, v. 68, n. 7, p. 2204–13, 2008.
- CONSORTIUM., I. H. The International HapMap Project. **Nature**, v. 426, n. 6968, p. 789–96, 2003.
- DATTA, S. et al. Sequence and expression variations in 23 genes involved in mitochondrial and non-mitochondrial apoptotic pathways and risk of oral leukoplakia and cancer. **Mitochondrion**, v. 25, p. 28–33, 2015.
- DIONNE, K. R. et al. Potentially malignant disorders of the oral cavity: Current practice and future directions in the clinic and laboratory. **Int J Cancer**, v. 136, n. 3, p. 503–515, 2015.
- DOST, F. et al. Malignant transformation of oral epithelial dysplasia: a real-world evaluation of histopathologic grading. **Oral Surg Oral Med Oral Pathol Oral Radiol**, v. 117, n. 3, p. 343–352, 2014.
- EGGER, G. et al. Epigenetics in human disease and prospects for epigenetic therapy. **Nature**, v. 429, n. 6990, p. 457–463, 2004.
- EL-NAGGAR, A. K. et al. **WHO Classification of Head and Neck Tumours**. 4rd. ed. Lyon, France: IARC, 2017.
- ER, T. K. et al. Molecular characterization of oral squamous cell carcinoma using targeted next-generation sequencing. **Oral Dis**, v. 21, n. 7, p. 872–878, 2015.
- Estimativa 2018: incidência de câncer no Brasil / Instituto Nacional de Câncer José Alencar Gomes da Silva. **Coordenação de Prevenção e Vigilância. – Rio de Janeiro: INCA**, 2018.
- FERLAY, J. et al. Cancer incidence and mortality worldwide: Sources, methods and major patterns in GLOBOCAN 2012. **Int J Cancer**, v. 136, n. 5, p. E359–E386, 2015.
- FERLAY, J. et al. Estimating the global cancer incidence and mortality in 2018: GLOBOCAN sources and methods. **Int J Cancer**, v. [Epub ahead of print], 2018a.
- FERLAY, J. et al. **Global Cancer Observatory: Cancer Today**.
- FLESKENS, S.; SLOOTWEG, P. Grading systems in head and neck dysplasia: their prognostic value, weaknesses and utility. **Head Neck Oncol**, n. 1, p. 11, 2009.
- FORBES, S. et al. Cosmic: somatic cancer genetics at high-resolution. **Nucleic Acids Res.**, v. 45, n. D1, p. D777–D783, 2017.
- GARNIS, C. et al. Genomic imbalances in precancerous tissues signal oral cancer risk.

Mol Cancer, v. 8, p. 1–7, 2009.

GONZALEZ-MOLES, M. et al. Beta-catenin in oral cancer: an update on current knowledge. **Oral Oncol**, v. 50, n. 9, p. 818–824, 2014.

GONZÁLEZ-PÉREZ, A.; LÓPEZ-BIGAS, N. Improving the Assessment of the Outcome of Nonsynonymous SNVs with a Consensus Deleteriousness Score, *Condel*. **Am J Hum Genet**, v. 88, n. 4, p. 440–9, 2011.

GOODWIN, S.; MCPHERSON, J. D.; MCCOMBIE, W. R. Coming of age: Ten years of next-generation sequencing technologies. **Nat Rev Genetics**, v. 17, n. 6, p. 333–351, 2016.

GRAVELAND, A. P. et al. Molecular screening of oral precancer. **Oral Oncol**, v. 49, n. 12, p. 1129–1135, 2013.

HA, P. K. et al. A transcriptional progression model for head and neck cancer. **Clinical Cancer Res**, v. 9, n. 8, p. 3058–3064, 2003.

HAMIDI, H.; IVASKA, J. Every step of the way: integrins in cancer progression and metastasis. **Nat Rev Cancer**, v. 18, n. 9, p. 533–548, 2018.

HANAHAN, D.; WEINBERG, R. Hallmarks of cancer: the next generation. **Cell**, v. 144, n. 5, p. 646–674, 2011.

HARROW, J. et al. GENCODE: the reference human genome annotation for The ENCODE Project. **Genome Res**, v. 22, n. 9, p. 1760–74, 2012.

HOLMSTRUP, P. Oral erythroplakia—What is it? **Oral Dis**, v. 24, n. 1–2, p. 138–143, 2018.

HU, T. et al. Forward and reverse mutations in stages of cancer development. **Human Genomics**, v. 12, n. 1, p. 40, 2018.

HUNG, K. F. et al. MicroRNA-31 upregulation predicts increased risk of progression of oral potentially malignant disorder. **Oral Oncol**, v. 53, p. 42–47, 2016.

KRAMER, I. et al. Definition of leukoplakia and related lesions: an aid to studies on oral precancer. **Oral Surg Oral Med Oral Pathol**, v. 46, n. 4, p. 518–39, 1978.

KUJAN, O. et al. Evaluation of a new binary system of grading oral epithelial dysplasia for prediction of malignant transformation. **Oral Oncol**, v. 42, n. 10, p. 987–993, 2006.

KUJAN, O. et al. Why oral histopathology suffers inter-observer variability on grading oral epithelial dysplasia: an attempt to understand the sources of variation. **Oral Oncol**, v. 43, n. 3, p. 224–31, 2007.

LANDRUM, M. et al. Clinvar: public archive of interpretations of clinically relevant variants. **Nucleic Acids Res.**, v. 44, n. D1, p. D862–8, 2016.

LAWRENCE, M. S. et al. Comprehensive genomic characterization of head and neck squamous cell carcinomas. **Nature**, v. 517, n. 7536, p. 576–582, 2015.

LEK, M. et al. Analysis of protein-coding genetic variation in 60,706 humans. **Nature**, v. 536, n. 7616, p. 285–91, 2016.

LI, H. Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. **ArXiv e-prints**, 2013.

LINGEN, M. et al. Genetics/epigenetics of oral premalignancy: Current status and future research. **Oral Dis**, v. 17, n. SUPPL. 1, p. 7–22, 2011.

- LÓPEZ-FERRANDO, V. et al. PMut: a web-based tool for the annotation of pathological variants on proteins, 2017 update. **Nucleic Acids Res.**, v. 45, n. W1, p. W222–W228, 2017.
- MCKENNA, A. et al. The Genome Analysis Toolkit: A MapReduce framework for analyzing next-generation DNA sequencing data. **Proceedings of the International Conference on Intellectual Capital, Knowledge Management & Organizational Learning**, v. 20, n. 9, p. 1297–1303, 2010.
- MCLAREN, W. et al. The ensembl variant effect predictor. **Genome Biol**, v. 17, n. 1, p. 122, 2016.
- MELLO, F. W. et al. Prevalence of oral potentially malignant disorders: A systematic review and meta-analysis. **J Oral Pathol Med**, v. 47, n. 7, p. 633–640, 2018.
- MEMORIAL SLOAN KETTERING CANCER CENTER, M. **cbioportal**. Disponible en: <<https://www.cbioportal.org/about>>.
- MI, H. et al. Protocol Update for large-scale genome and gene function analysis with the PANTHER classification system (v.14.0). **Nat Protocols**, v. 14, p. 703–721, 2019.
- MITCHELL, D. et al. Transcriptional regulation of KiSS-1 gene expression in metastatic melanoma by specificity protein-1 and its coactivator DRIP-130. **Oncogene**, v. 26, n. 12, p. 1739–1747, 2007.
- NANNO, Y. et al. Serum Elastase 1 Level as a Risk Factor for Postoperative Recurrence in Patients with Well-Differentiated Pancreatic Neuroendocrine Neoplasms. **Annals of Surgical Oncology**, v. 25, n. 11, p. 3358–3364, 2018.
- NG, P.; HENIKOFF, S. Predicting deleterious amino acid substitutions. **Genome Res**, v. 11, n. 5, p. 863–74, 2001.
- NHLBI EXOME SEQUENCING PROJECT (ESP). **Exome Variant Server**.
- NIKITAKIS, N. G. et al. Molecular markers associated with development and progression of potentially premalignant oral epithelial lesions: Current knowledge and future implications. **Oral Surg Oral Med Oral Pathol Oral Radiol**, v. 125, n. 6, p. 650–669, 2018a.
- NIKITAKIS, N. G. et al. Alterations in the expression of DNA damage response-related molecules in potentially preneoplastic oral epithelial lesions. **Oral Surg Oral Med Oral Pathol Oral Radiol**, v. 125, n. 6, p. 637–649, 2018b.
- PEREZ-ORDONEZ, B.; BEAUCHEMIN, M.; JORDAN, R. Molecular biology of squamous cell carcinoma of the head and neck. **J Clin Pathol**, v. 59, n. 5, p. 445–53, 2006.
- PETTI, S. Pooled estimate of world leukoplakia prevalence: a systematic review. **Oral Oncol**, v. 39, p. 770–780, 2003.
- PICKERING, C. et al. Integrative genomic characterization of oral squamous cell carcinoma identifies frequent somatic drivers. **Cancer Discov**, v. 3, n. 7, p. 770–81, 2013.
- PINDBORG, J. et al. World Health Organization: histological typing of cancer and precancer of the oral mucosa. **Berlin: Springer-Verlag**, 1997.
- RAMACHANDRAN, C. et al. Expression profiles of apoptotic genes induced by curcumin in human 312 breast cancer and mammary epithelial cell lines. **Anticancer research**, v. 25, p. 3293–33, 2005.

- REIBEL, J. Prognosis of oral pre-malignant lesions: significance of clinical, histopathological, and molecular biological characteristics. **Crit Rev Oral Biol Med**, v. 14, n. 1, p. 47–62, 2003.
- REICHART, P.; PHILIPSEN, H. Oral erythroplakia—a review. **Oral Oncol**, v. 41, n. 6, p. 551–561, 2005.
- RIERA, P.; MARTÍNEZ, B. Morbilidad y mortalidad por cáncer oral y faríngeo en Chile. **Rev Méd Chile**, v. 133, p. 555–563, 2005.
- ROSIN, M. et al. Use of allelic loss to predict malignant risk for low-grade oral epithelial dysplasia. **Clin Cancer Res**, v. 6, n. 2, p. 357–362, 2000.
- SALAHSHOURIFAR, I. et al. Genomic DNA copy number alterations from precursor oral lesions to oral squamous cell carcinoma. **Oral Oncol**, v. 50, n. 5, p. 404–412, 2014.
- SCULLY, C. Challenges in predicting which oral mucosal potentially malignant disease will progress to neoplasia. **Oral Dis**, v. 20, n. 1, p. 1–5, 2014.
- SCULLY, C.; BAGAN, J. Oral squamous cell carcinoma overview. **Oral Oncol**, v. 45, n. 4–5, p. 301–8, 2009.
- SEHN, J. **Insertions and Deletions (Indels)**. ELSEVIER ed. [s.l: s.n.].
- SHERRY, S. et al. dbsnp: the ncbi database of genetic variation. **Nucleic Acids Res.**, v. 29, n. 1, p. 308–11, 2001.
- SHIAH, S.; SHIEH, Y.; CHANG, J. The Role of Wnt Signaling in Squamous Cell Carcinoma. **J Dental Res**, v. 95, n. 2, p. 129–34, 2016.
- SHIGEMIZU, D. et al. A practical method to detect SNVs and indels from whole genome and exome sequencing data. **Scientific Reports**, v. 3, p. 1–6, 2013.
- SHRIDHAR, K. et al. Single nucleotide polymorphisms as markers of genetic susceptibility for oral potentially malignant disorders risk: Review of evidence to date. **Oral Oncol**, v. 61, p. 146–151, 2016.
- SMITH, J. et al. Biomarkers in dysplasia of the oral cavity: a systematic review. **Oral Oncol**, v. 45, n. 8, p. 647–653, 2009.
- SPEIGHT, P. M.; KHURRAM, S. A.; KUJAN, O. Oral potentially malignant disorders: risk of progression to malignancy. **Oral Surg Oral Med Oral Pathol Oral Radiol**, v. 125, n. 6, p. 612–627, 2018.
- SPERANDIO, M. et al. Predictive value of dysplasia grading and DNA ploidy in malignant transformation of oral potentially malignant disorders. **Cancer Prev Res (Phila)**, v. 6, n. 8, p. 822–31, 2013.
- SPIRA, A. et al. Precancer Atlas to Drive Precision Prevention Trials. **Cancer Res**, v. 77, n. 7, p. 1510–1542, 2017.
- SRIVASTAVA, S. et al. The Making of a PreCancer Atlas : Promises, Challenges , and Opportunities. **Trends cancer**, v. 4, n. 8, p. 523–536, 2018.
- STENSON, P. et al. Human Gene Mutation Database (HGMD): 2003 update. **Hum Mutat**, v. 21, n. 6, p. 577–81, 2003.
- STEWART, B. W.; WILD, C. P. World Cancer Report 2014. **IARC. Lyon: WHO Press**, p. 1–632, 2014.
- STRANGER, B. E. et al. Relative impact of nucleotide and copy number variation on

- gene expression phenotypes. **Science**, v. 315, n. 5813, p. 848–853, 2007.
- STRANSKY, N. et al. The mutational landscape of head and neck squamous cell carcinoma. **Science**, v. 333, n. 6046, p. 1157–60, 2011.
- STRANSKY, N. et al. The Mutational Landscape of Head and Neck Squamous Cell Carcinoma. **Science**, v. 333, n. 6046, p. 1157–1160, 2012.
- SUDMANT, P. et al. An integrated map of structural variation in 2,504 human genomes. **Nature**, v. 526, n. 7571, p. 75–81, 2015.
- SZKLARCZYK, D. et al. STITCH 5: augmenting protein-chemical interaction networks with tissue and affinity data. **Nucleic Acids Res**, v. 44, n. D1, p. D380-4, 2016.
- TALAS, U. et al. Human elastase 1: Evidence for expression in the skin and the identification of a frequent frameshift polymorphism. **J Inv Dermatol**, v. 114, n. 1, p. 165–170, 2000.
- THOMSON, P. J. et al. Interventional laser surgery for oral potentially malignant disorders: a longitudinal patient cohort study. **Int J Oral and Maxillofac Surg**, v. 46, n. 3, p. 337–342, 2017.
- TIAN, J. et al. KiSS1 gene as a novel mediator of TGF β -mediated cell invasion in triple negative breast cancer. **Cell Signal**, v. 42, p. 1–10, 2018.
- UNIPROT CONSORTIUM, U. C. UniProt: a worldwide hub of protein knowledge. **Nucleic Acids Res**, v. 47, n. D1, p. D506–D515, 2019.
- VAN DER AUWERA, G. et al. From FastQ data to high confidence variant calls: the Genome Analysis Toolkit bestpractices pipeline. **Curr Protoc Bioinformatics**, v. 43, n. 11.10, p. 1–33, 2013.
- VAN DER WAAL, I. Potentially malignant disorders of the oral and oropharyngeal mucosa; terminology, classification and present concepts of management. **Oral Oncol**, v. 45, n. 4–5, p. 317–323, 2009.
- VAN DER WAAL, I. Oral leukoplakia, the ongoing discussion on definition and terminology. **Med Oral Patol Oral Cir Bucal**, v. 20, n. 6, p. e685–e692, 2015.
- WANG, Y. et al. PIG11 over-expression predicts good prognosis and induces HepG2 cell apoptosis via reactive oxygen species-dependent mitochondrial pathway. **Biomedicine and Pharmacotherapy**, v. 108, n. September, p. 435–442, 2018.
- WARNAKULASURIYA, S. et al. Oral epithelial dysplasia classification systems: Predictive value, utility, weaknesses and scope for improvement. **J Oral Pathol Med**, v. 37, n. 3, p. 127–133, 2008.
- WARNAKULASURIYA, S. Global epidemiology of oral and oropharyngeal cancer. **Oral Oncol**, v. 45, n. 4–5, p. 309–316, 2009.
- WARNAKULASURIYA, S. et al. Factors predicting malignant transformation in oral potentially malignant disorders among patients accrued over a 10-year period in South East England. **J Oral Pathol Med**, v. 40, n. 9, p. 677–683, 2011.
- WARNAKULASURIYA, S. Prognostic and predictive markers for oral squamous cell carcinoma: The importance of clinical, pathological and molecular markers. **Saudi J Med Sci**, v. 2, n. 1, p. 12–16, 2014.
- WARNAKULASURIYA, S. Clinical features and presentation of oral potentially malignant disorders. **Oral Surg Oral Med Oral Pathol Oral Radiol**, v. 125, n. 6, p.

582–590, 2018.

WARNAKULASURIYA, S.; JOHNSON, N. W.; VAN DER WAAL, I. Nomenclature and classification of potentially malignant disorders of the oral mucosa. **J Oral Pathol Med**, v. 36, n. 10, p. 575–580, 2007.

WOOD, H. M. et al. The clonal relationships between pre-cancer and cancer revealed by ultra-deep sequencing. **J Pathol**, v. 237, n. 3, p. 296–306, 2015.

WOOD, H. M. et al. The genomic road to invasion-examining the similarities and differences in the genomes of associated oral pre-cancer and cancer samples. **Genome Med**, v. 9, n. 1, p. 1–14, 2017.

WOOLGAR, J. Histopathological prognosticators in oral and oropharyngeal squamous cell carcinoma. **Oral Oncol**, v. 42, n. 3, p. 229–239, 2006.

WU, W. et al. Six1: A critical transcription factor in tumorigenesis. **Int J Cancer**, v. 136, n. 6, p. 1245–1253, 2015.

WU, Y. et al. PIG11 is involved in hepatocellular carcinogenesis and its over-expression promotes Hepg2 cell apoptosis. **Pathol Oncol Res**, v. 15, n. 3, p. 411–416, 2009.

XIE, Y. et al. SIX1 is upregulated in gastric cancer and regulates proliferation and invasion by targeting the ERK pathway and promoting epithelial-mesenchymal transition. **Cell Biochem Funct.**, v. 36, n. 8, p. 413–419, 2018.

ZERBINO, D. R. et al. Ensembl 2018. **Nucleic Acids Research**, v. 46, n. D1, p. D754–D761, 2018.

ZHANG, L. et al. Loss of Heterozygosity (LOH) Profiles – Validated Risk Predictors for Progression to Oral Cancer. **Cancer Prev Res (Phila)**, v. 5, n. 9, p. 1081–1089, 2012.

ZHANG, L. et al. Should severe epithelial dysplasia be treated? **Oral Oncol**, v. 60, p. 125–129, 2016.

ZHANG, X. et al. Nomogram for risk prediction of malignant transformation in oral leukoplakia patients using combined biomarkers. **Oral Oncol**, v. 72, p. 132–139, 2017.

ZHU, M. et al. Expression of DNA doublestrand repair proteins in oral leukoplakia and the risk of malignant transformation. **Oncology Letters**, v. 15, n. 6, p. 9827–9835, 2018.

ZOU, J. et al. APOBEC3B, a molecular driver of mutagenesis in human cancers. **Cell Biosci**, v. 7, p. 29, 2017.

ANEXO A



Ed 22/12/2014

ACTA DE APROBACION DE PROTOCOLO DE INVESTIGACIÓN

Dra. MA.TorresPdte./ Dr. E.Rodriguez/ Srta. A.Herrera/Dra. B. Urzúa O./ Dra. X. Lee M./ Srta. K. Lagos B./ Sra. I.Comejo P./ Sra. V. Rodriguez D./

ACTA N°:14

1. Acta De Aprobación De Protocolo De Estudio N° 2014/29
2. Miembros del Comité Ético-Científico de la Facultad de Odontología de la Universidad de Chile participantes en la aprobación del Proyecto:

| | | |
|---|---|---|
| Dra. M ^a Angélica Torres V Presidente CEC | Srta. Karín Lagos Secretaría CEC | Dr. Eduardo Rodriguez Miembro permanente del CEC |
| Dra. Blanca Urzúa Miembro permanente del CEC | Srta. Ma. Isabel Comejo Miembro permanente del CEC | Dra. Ximena Lee Miembro permanente del CEC |
| Srta. Andrea Herrera Miembro permanente del CEC | | |
3. Fecha d Aprobación: 17 de Diciembre 2014.
4. Título completo del proyecto: en español: "DIAGNÓSTICO, MANEJO Y TRATAMIENTO DE LESIONES POTENCIALMENTE MALIGNAS Y MALIGNAS DEL TERRITORIO BUCOMAXILOFACIAL/ IDENTIFYING GENETIC ALTERATIONS DURING ORAL TUMOR DEVELOPMENT". En Inglés: IDENTIFYING GENETIC ALTERATIONS DURING ORAL TUMOR DEVELOPMENT" Proyecto FONDECYT versión 25/05/2014
5. Investigador responsable: Dr. Ricardo Fernández Ramírez, Genetista
6. Institución Patrocinante: FONDECYT, Facultad de Odontología de la Universidad de Chile
7. Documentación Revisada:
 - Propuesta Fondecyt versión 25-05-2014
 - Consentimiento Informado (CI) versión 03-11-2014
 - Currículo del Investigador responsable
 - Nómina de los coinvestigadores y colaboradores directos de la Investigación.
 - Material que se utilizará: muestras, secuenciadores, información que se entregará a los participantes, entre otros.
 - Cartas de aceptación:
 - Carta de Director del Instituto Nacional del Cáncer Mauro Salinas Cortés, y del Dr. Fabio Valdés, Investigador del Instituto Nacional del Cáncer de aceptación del estudio.
 - Carta del Comité Bioético de la Facultad de Odontología de la Universidad de Chile comprometiendo el uso de muestras.

Ed 22/12/2014

- Carta de la Dra. Sandra Beatriz Chaves de la Facultad de Odontología de la Universidad de Pelotas, Brasil, aprobando el uso de muestras.

8. **Carácter de la población:** Se estudia población con lesiones potencialmente malignas y malignas del territorio bucomaxilofacial para determinar qué tipo de mutaciones genéticas caracterizan el desarrollo de cáncer y así mejorar el diagnóstico pronóstico y en el futuro poder diseñar estrategias de tratamiento de estos pacientes.

9. Fundamentación de la aprobación

El comité ético científico considera que el proyecto del Dr. Ricardo Fernández tiene mucha relevancia en investigación en salud bucal, y para la sociedad, contribuyendo a mejorar las herramientas para un diagnóstico precoz, de esta terrible enfermedad como es el cáncer oral; mejorar los protocolos de tratamiento y obtener un mejor pronóstico para los pacientes. El proyecto del Dr. Ricardo Fernández respeta la confidencialidad de los datos personales y muestras de los pacientes de su investigación. Las muestras donadas solo se usarán para propósito de investigación de mutaciones genéticas relacionadas con la progresión de cáncer bucomaxilofacial. El proyecto se ajusta a las pautas éticas de investigación con seres humanos (Declaración de Helsinki, CIOMS, Ley 20.124) y la relación riesgo/beneficio es aceptable. El formulario de consentimiento informado cumple con los requisitos éticos exigidos. Los antecedentes curriculares del Investigador Principal garantizan la ejecución del proyecto dentro de los marcos éticamente aceptables.

En consecuencia, el Comité Ético Científico de la Facultad de Odontología de la Universidad de Chile, Aprueba por unanimidad de sus miembros el proyecto Fondecyt versión 25-05-2014: **en español:** "DIAGNÓSTICO, MANEJO Y TRATAMIENTO DE LESIONES POTENCIALMENTE MALIGNAS Y MALIGNAS DEL TERRITORIO BUCOMAXILOFACIAL/ IDENTIFYING GENETIC ALTERATIONS DURING ORAL TUMOR DEVELOPMENT". En Inglés: IDENTIFYING GENETIC ALTERATIONS DURING ORAL TUMOR DEVELOPMENT" bajo la conducción del Dr. Ricardo Fernández, profesor de la Facultad de Odontología de la U. de Chile.

El Dr. Ricardo Fernández asume el compromiso de enviar a este Comité cualquier enmienda realizada durante la ejecución del protocolo y una copia del Informe final de resultados. Este Comité se reserva el derecho de monitorear este proyecto si lo considera necesario y el Investigador deberá, bajo mutuo acuerdo, presentar los antecedentes solicitados.


Dra. María Angélica Torres V.
Presidente C.E.C

c/c.: Investigador Principal y Secretaria C.E.C.

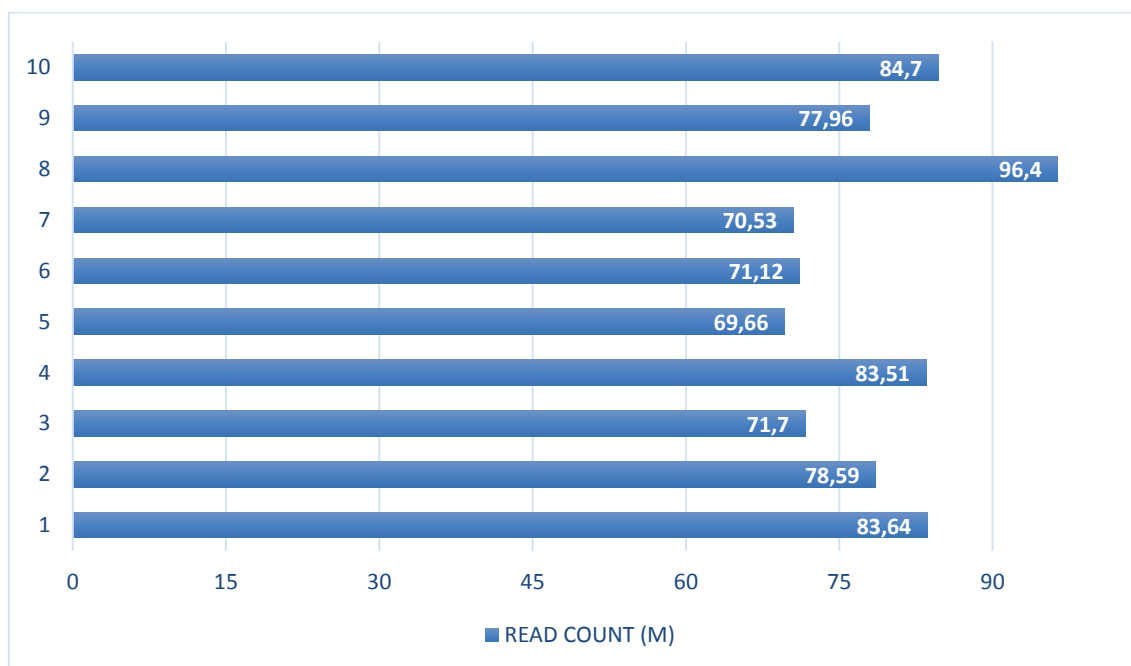


ANEXO B

Tabela S1 Estatística de dados *raw*

| Amostra | Total de bases <i>read</i> (bp) | Total de <i>reads</i> | GC (%) | AT (%) | Q20 (%) | Q30 (%) |
|---------|------------------------------------|--------------------------|--------|--------|---------|---------|
| 1 | 8,447,176,006 | 83,635,406 | 49.48 | 50.52 | 98.13 | 95.33 |
| 2 | 7,937,180,142 | 78,585,942 | 48.84 | 51.16 | 98.18 | 95.43 |
| 3 | 7,242,433,866 | 71,707,266 | 48.3 | 51.7 | 98.4 | 95.9 |
| 4 | 8,434,901,476 | 83,513,876 | 48.54 | 51.46 | 98.27 | 95.62 |
| 5 | 7,035,847,254 | 69,661,854 | 48.65 | 51.35 | 98.29 | 95.62 |
| 6 | 7,184,115,860 | 71,129,860 | 48.78 | 51.22 | 98.25 | 95.59 |
| 7 | 7,123,895,216 | 70,533,616 | 48.68 | 51.32 | 98.21 | 95.49 |
| 8 | 9,736,227,088 | 96,398,288 | 49.06 | 50.94 | 98.21 | 95.5 |
| 9 | 7,874,290,472 | 77,963,272 | 48.38 | 51.62 | 98.36 | 95.83 |
| 10 | 8,555,191,668 | 84,704,868 | 48.34 | 51.66 | 98.41 | 95.98 |

- Total de bases *read*: Número de bases sequenciadas.
- Total de *reads*: Número total de *reads*. Para o sequenciamento *paired-end* do Illumina, esse valor refere-se à soma de *read* 1 e *read* 2.
- GC (%): Conteúdo de GC.
- AT (%): Conteúdo de AT.
- Q20 (%): Proporção de *reads* que possuem pontuação de qualidade do *phred* acima de 20.
- Q30 (%): Proporção de *reads* que possuem pontuação de qualidade do *phred* acima de 30.

Figura S1 Total de *reads*

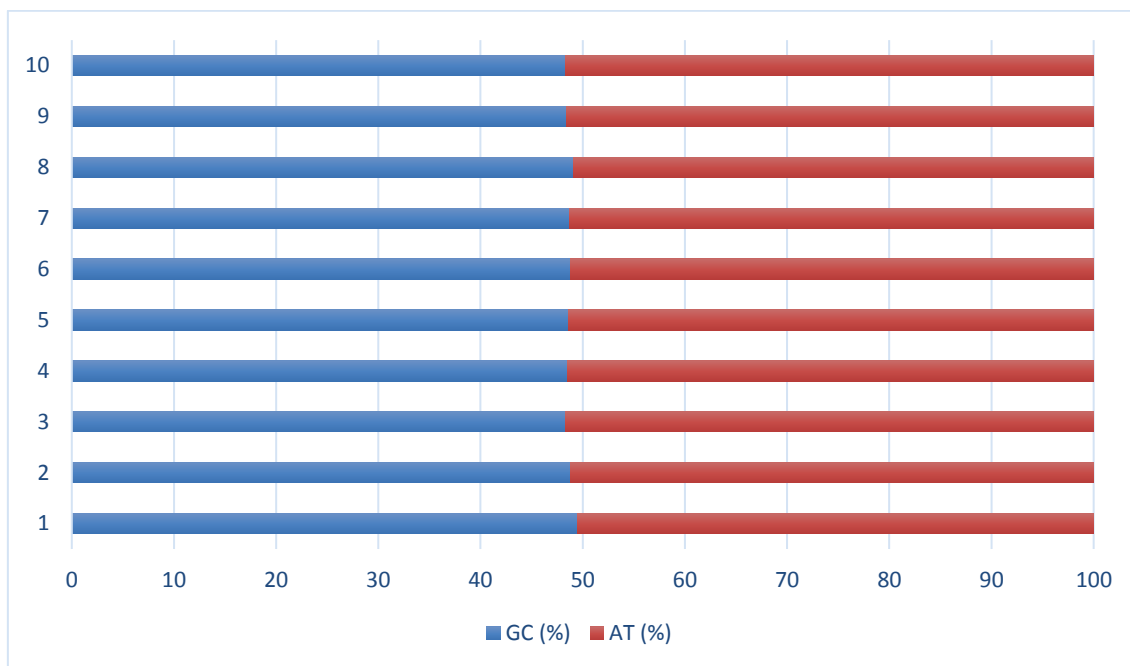


Figura S2 Conteúdo de GC/AT

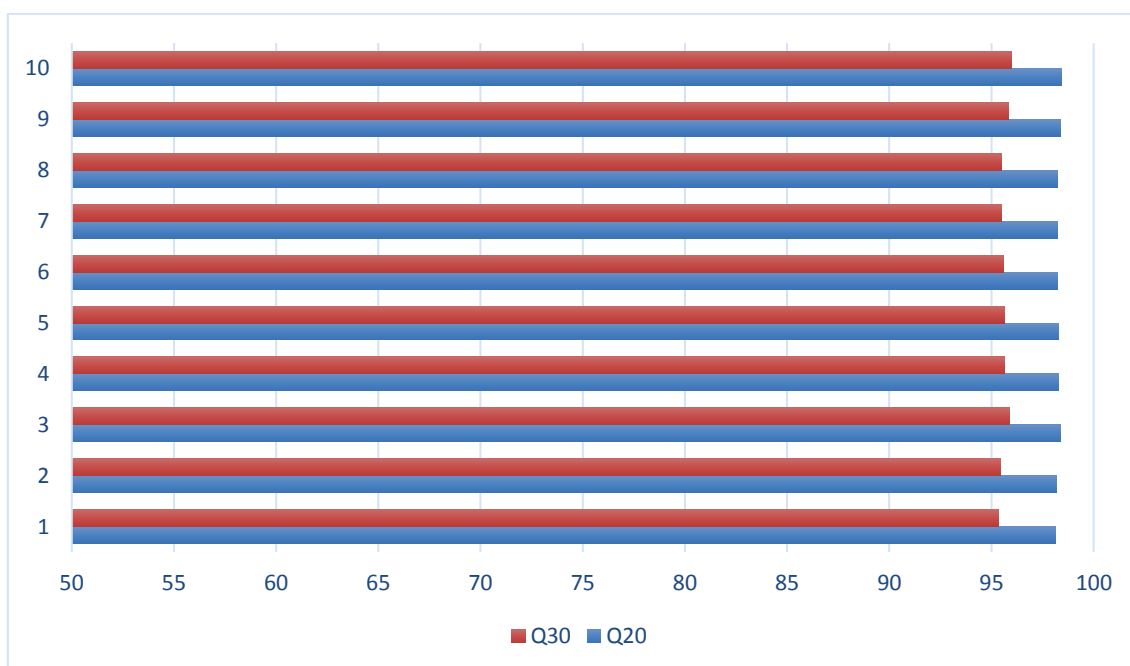


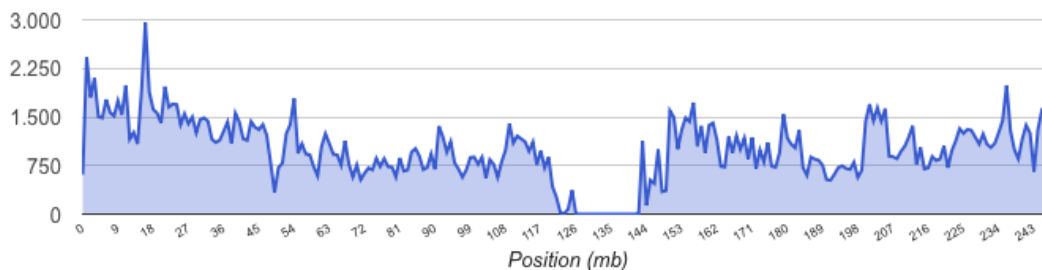
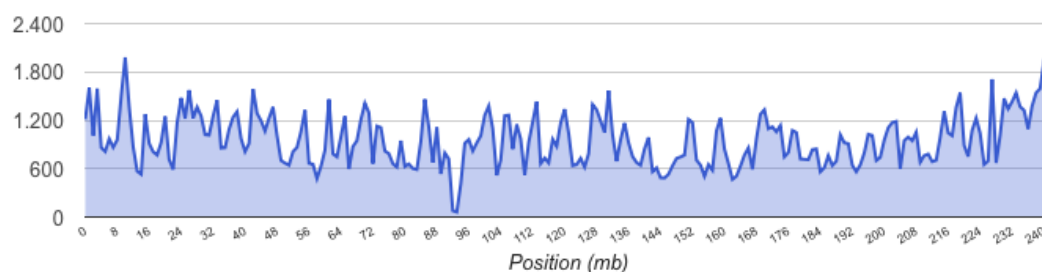
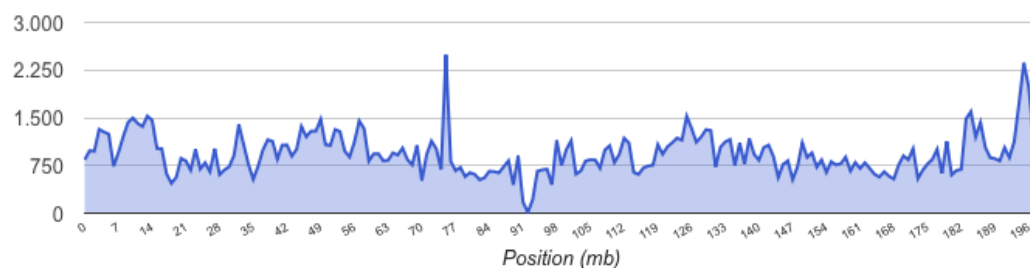
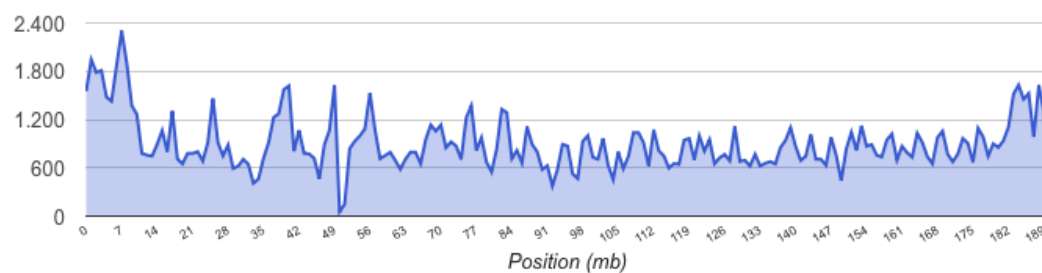
Figura S3 Q20/Q30 (%)

Tabela S2 Localização genômica das consequências calculadas das variantes

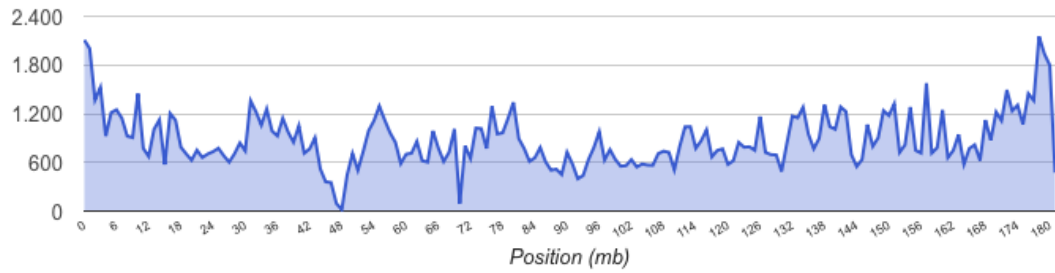
| Tipo de consequência | Total Variantes | % |
|---|------------------------|-------------|
| <i>splice_acceptor_variant</i> | 401 | 0,013123226 |
| <i>splice_donor_variant</i> | 525 | 0,017181282 |
| <i>stop_gained</i> | 553 | 0,018097617 |
| <i>frameshift_variant</i> | 1.139 | 0,037275199 |
| <i>stop_lost</i> | 125 | 0,004090781 |
| <i>start_lost</i> | 113 | 0,003698066 |
| <i>inframe_insertion</i> | 465 | 0,015217706 |
| <i>inframe_deletion</i> | 596 | 0,019504845 |
| <i>missense_variant</i> | 33.943 | 1,11082712 |
| <i>protein_altering_variant</i> | 6 | 0,000196358 |
| <i>splice_region_variant</i> | 8.875 | 0,290445473 |
| <i>incomplete_terminal_codon_variant</i> | 4 | 0,000130905 |
| <i>start_retained_variant</i> | 2 | 6,54525E-05 |
| <i>stop_retained_variant</i> | 51 | 0,001669039 |
| <i>synonymous_variant</i> | 29.786 | 0,974784097 |
| <i>coding_sequence_variant</i> | 12 | 0,000392715 |
| <i>mature_miRNA_variant</i> | 143 | 0,004679854 |
| <i>5_prime_UTR_variant</i> | 17.893 | 0,5855708 |
| <i>3_prime_UTR_variant</i> | 44.636 | 1,460768916 |
| <i>non_coding_transcript_exon_variant</i> | 88.195 | 2,886291661 |
| <i>intron_variant</i> | 1.587.688 | 51,95907517 |
| <i>non_coding_transcript_variant</i> | 1 | 3,27263E-05 |
| <i>upstream_gene_variant</i> | 162.658 | 5,32318645 |
| <i>downstream_gene_variant</i> | 123.299 | 4,035113958 |
| <i>TFBS_ablation</i> | 9 | 0,000294536 |
| <i>TF_binding_site_variant</i> | 1.350 | 0,044180438 |
| <i>regulatory_region_variant</i> | 82.396 | 2,696512134 |
| <i>intergenic_variant</i> | 870.787 | 28,49759348 |
| Total | 3.055.651 | 100 |

Tabela S3 Total de variantes por cromossomo

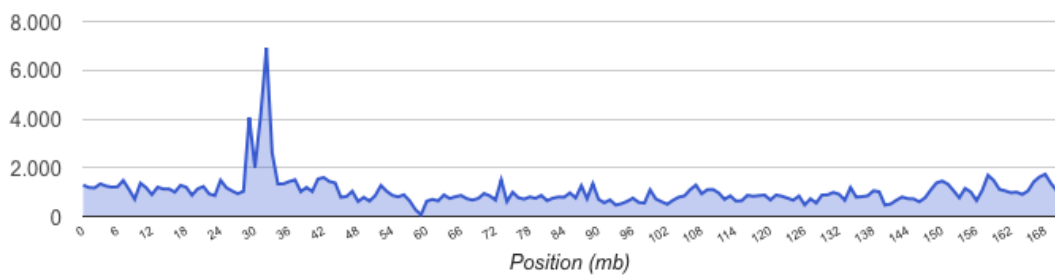
| Cromossoma | Total variantes |
|-------------------|------------------------|
| 1 | 248.166 |
| 2 | 232.138 |
| 3 | 184.824 |
| 4 | 173.144 |
| 5 | 160.562 |
| 6 | 176.016 |
| 7 | 169.504 |
| 8 | 142.986 |
| 9 | 126.434 |
| 10 | 149.485 |
| 11 | 145.230 |
| 12 | 143.747 |
| 13 | 92.707 |
| 14 | 96.117 |
| 15 | 95.063 |
| 16 | 107.907 |
| 17 | 115.687 |
| 18 | 73.903 |
| 19 | 109.384 |
| 20 | 82.480 |
| 21 | 45.190 |
| 22 | 61.210 |
| X | 87.451 |
| Y | 6.749 |

Figura S4 Distribuição das variantes em cada cromossoma**Distribution of variants on chromosome 1****Distribution of variants on chromosome 2****Distribution of variants on chromosome 3****Distribution of variants on chromosome 4**

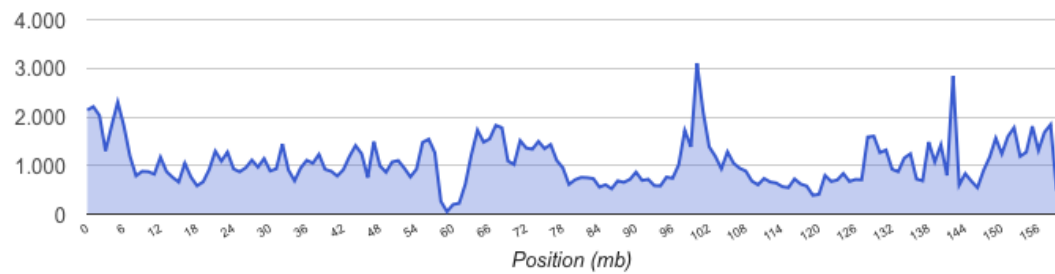
Distribution of variants on chromosome 5



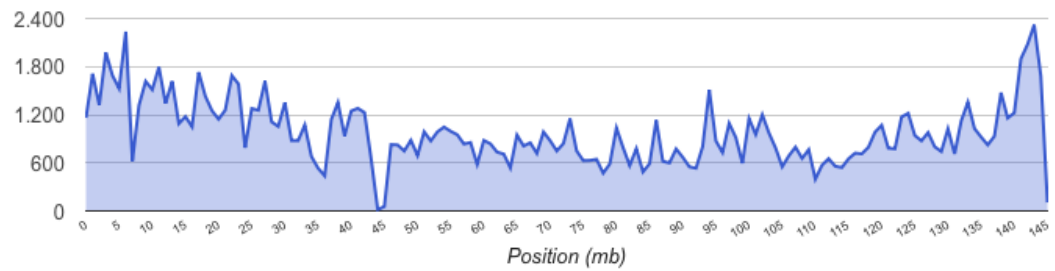
Distribution of variants on chromosome 6

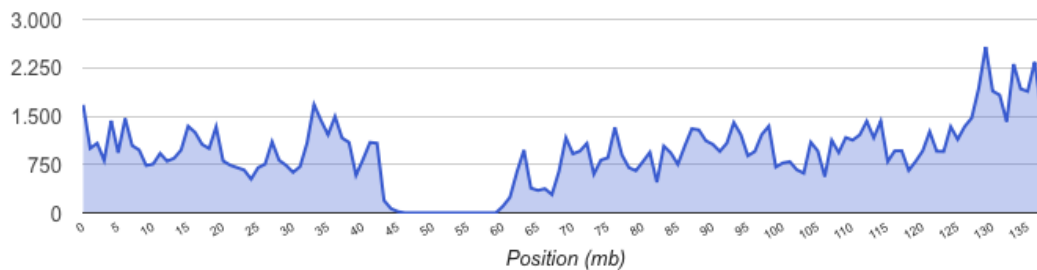
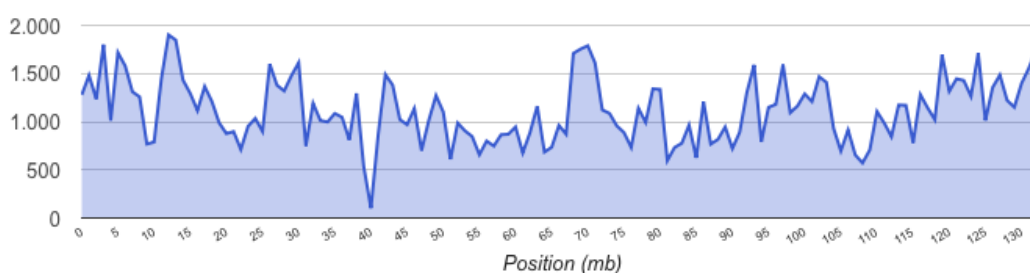
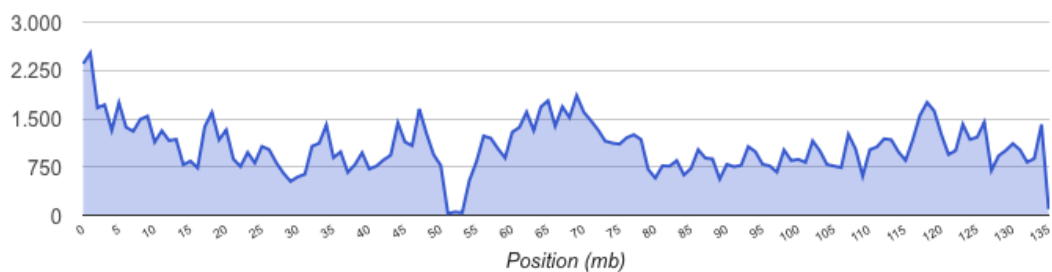
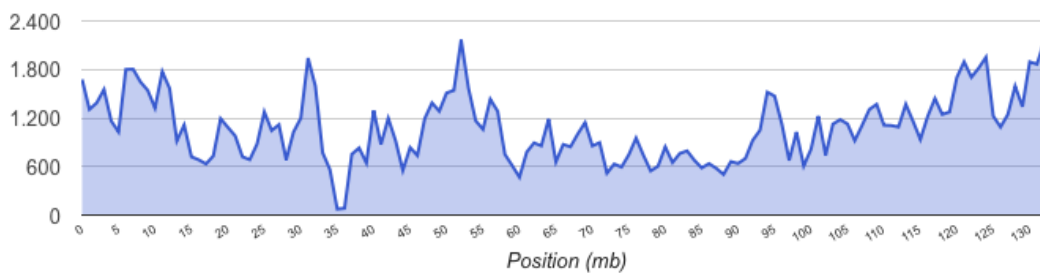


Distribution of variants on chromosome 7

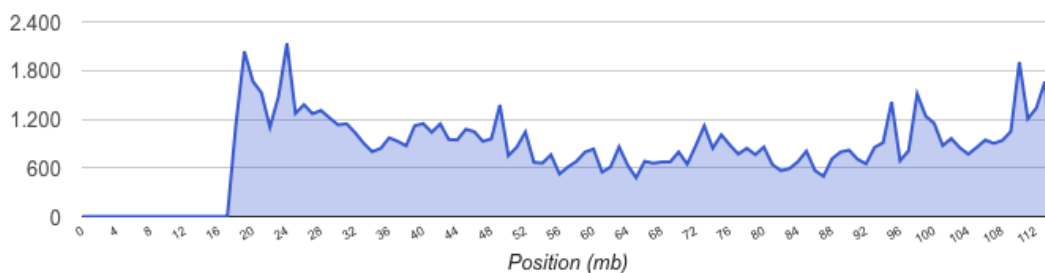


Distribution of variants on chromosome 8

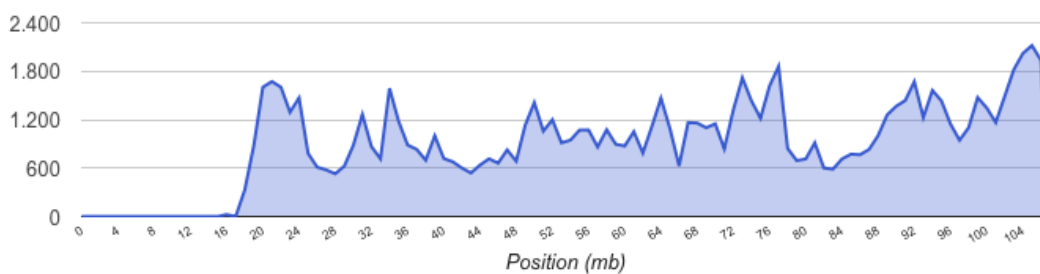


Distribution of variants on chromosome 9**Distribution of variants on chromosome 10****Distribution of variants on chromosome 11****Distribution of variants on chromosome 12**

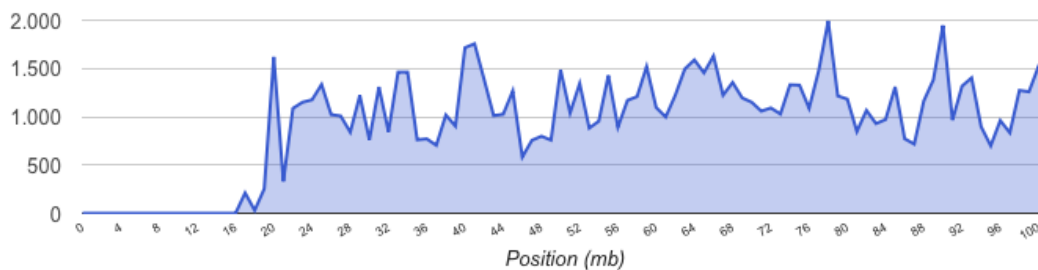
Distribution of variants on chromosome 13



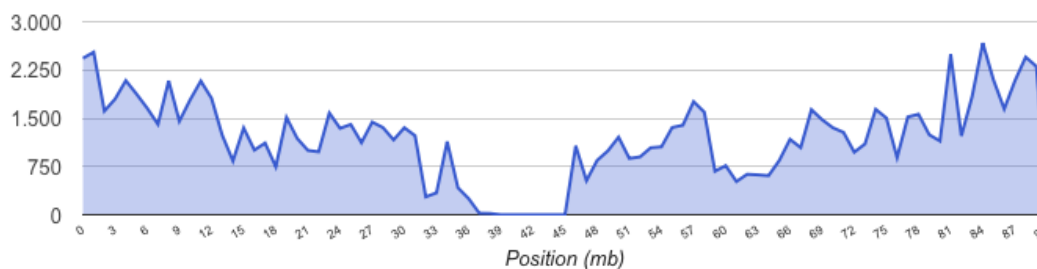
Distribution of variants on chromosome 14



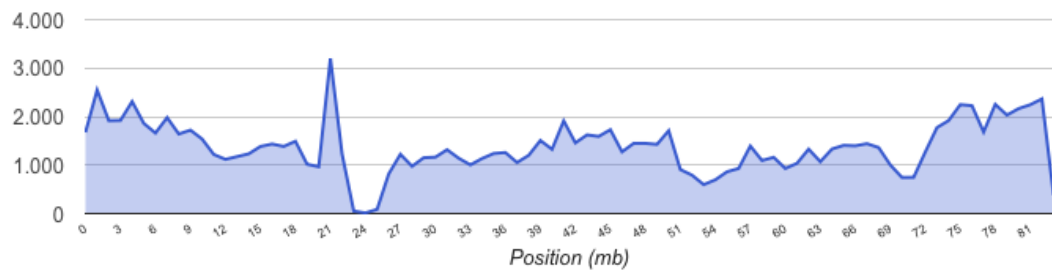
Distribution of variants on chromosome 15



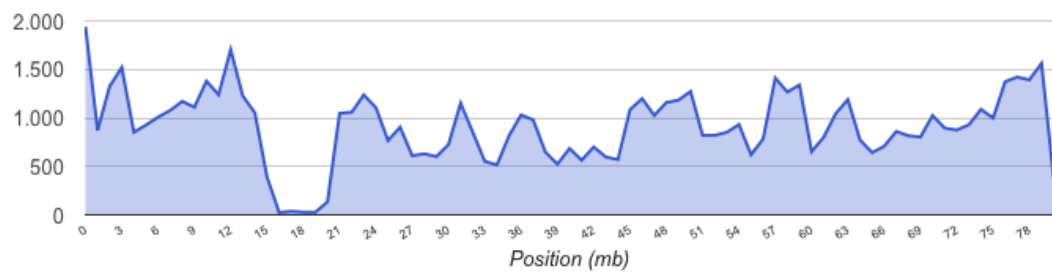
Distribution of variants on chromosome 16



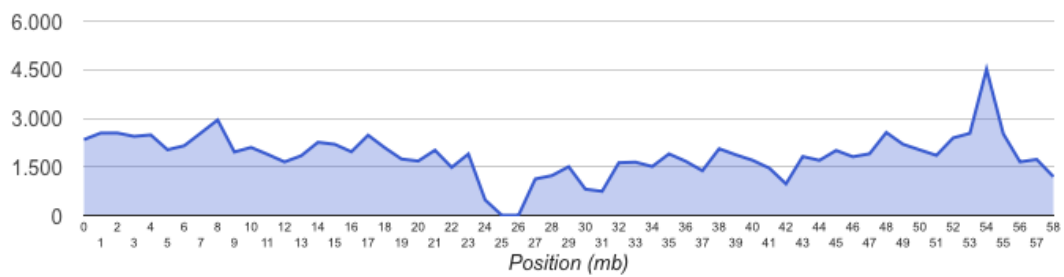
Distribution of variants on chromosome 17



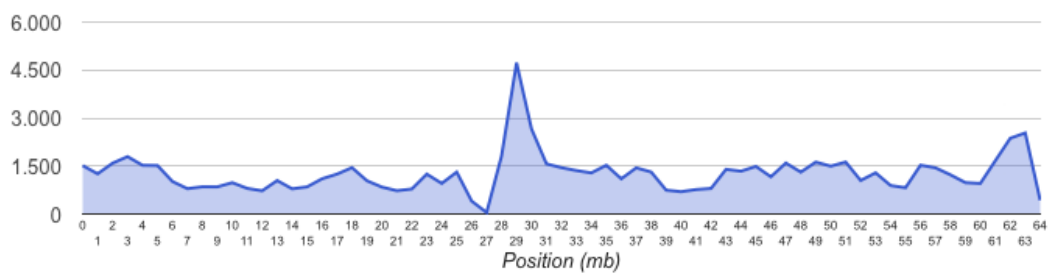
Distribution of variants on chromosome 18



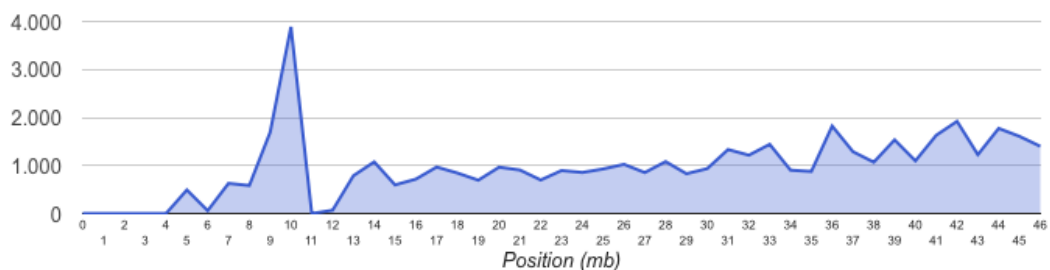
Distribution of variants on chromosome 19



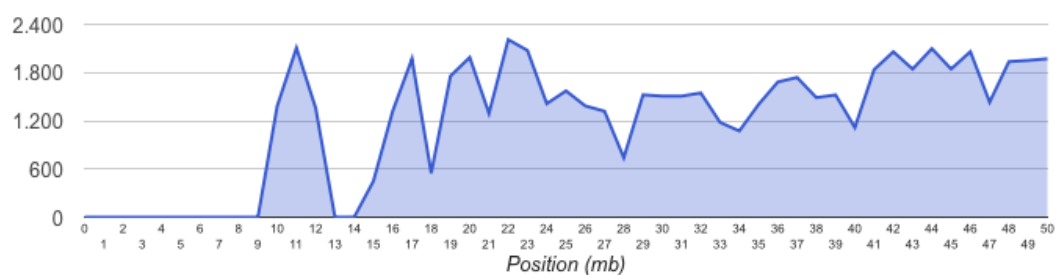
Distribution of variants on chromosome 20



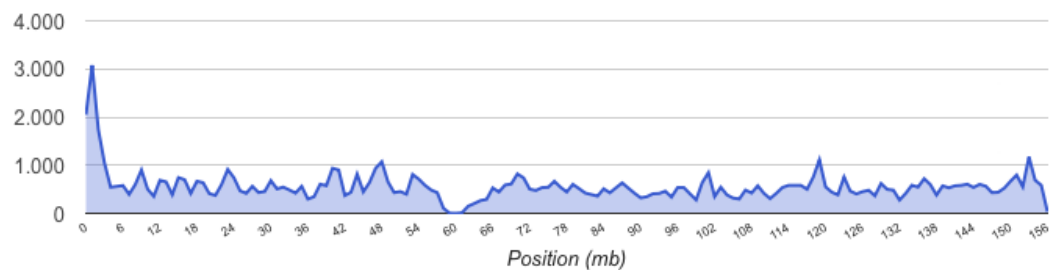
Distribution of variants on chromosome 21



Distribution of variants on chromosome 22



Distribution of variants on chromosome X



Distribution of variants on chromosome Y

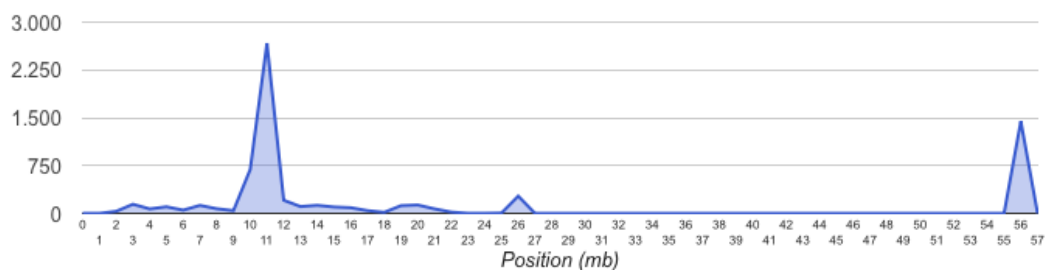


Tabela S4 Resumo das predições pelos programas *SIFT* e *PolyPhen*

| Total variantes | |
|-----------------------------------|--------|
| Predição <i>SIFT</i> | |
| <i>deleterious_low_confidence</i> | 4.699 |
| <i>tolerated_low_confidence</i> | 10.845 |
| <i>deleterious</i> | 17.966 |
| <i>tolerated</i> | 51.460 |
| Predição <i>PolyPhen</i> | |
| <i>unknown</i> | 2.693 |
| <i>possibly_damaging</i> | 9.385 |
| <i>probably_damaging</i> | 9.768 |
| <i>benign</i> | 65.329 |

Tabela S5 Genes associados a variantes, observadas nas amostras de DEO do presente estudo, que causam baixo (cor branco), moderado (cor azul) e alto impacto (cor rosa) em estruturas aminoacídicas.

| SÍMBOLO | GENE | SÍMBOLO | GENE |
|--------------|-----------------|------------|-----------------|
| A2M | ENSG00000175899 | AC073488.1 | ENSG00000232111 |
| A2MP1 | ENSG00000256069 | AC073869.4 | ENSG00000276460 |
| AADACL3 | ENSG00000188984 | AC087273.2 | ENSG00000254054 |
| AARS2 | ENSG00000124608 | AC090707.1 | ENSG00000254906 |
| AATF | ENSG00000275700 | AC092821.1 | ENSG00000214776 |
| ABCA10 | ENSG00000154263 | AC093762.1 | ENSG00000284635 |
| ABCA13 | ENSG00000179869 | AC095038.1 | ENSG00000228303 |
| ABCB11 | ENSG00000073734 | AC097467.3 | ENSG00000250910 |
| ABCC4 | ENSG00000125257 | AC097637.1 | ENSG00000173366 |
| ABCC6 | ENSG00000091262 | AC103739.2 | ENSG00000259314 |
| ABHD14A | ENSG00000248487 | AC104393.1 | ENSG00000253509 |
| ABHD14A-ACY1 | ENSG00000114786 | AC107990.1 | ENSG00000267066 |
| ABHD5 | ENSG00000011198 | AC112484.1 | ENSG00000187695 |
| AC006974.1 | ENSG00000283504 | AC115220.1 | ENSG00000241149 |
| AC008440.2 | ENSG00000232220 | ACAT2 | ENSG00000120437 |
| AC008758.6 | ENSG00000269755 | ACOT11 | ENSG00000162390 |
| AC009120.3 | ENSG00000260884 | ACOT8 | ENSG00000101473 |
| AC011405.1 | ENSG00000249593 | ACSF3 | ENSG00000176715 |
| AC015909.1 | ENSG00000249406 | ACTN2 | ENSG00000077522 |
| AC064853.5 | ENSG00000284667 | ACTR1B | ENSG00000115073 |
| AC064853.6 | ENSG00000284704 | ADAM30 | ENSG00000134249 |
| AC068987.5 | ENSG00000284730 | ADAMTS17 | ENSG00000140470 |

| | | | |
|------------|-----------------|-----------|-----------------|
| ADAMTS9 | ENSG00000163638 | ARL11 | ENSG00000152213 |
| ADAMTSL2 | ENSG00000197859 | ARSD | ENSG00000006756 |
| ADAT1 | ENSG00000065457 | ARVCF | ENSG00000099889 |
| ADCY10 | ENSG00000143199 | ASPM | ENSG00000066279 |
| ADCYAP1 | ENSG00000141433 | ATP10B | ENSG00000118322 |
| ADGB | ENSG00000118492 | ATP12A | ENSG00000075673 |
| ADGRE1 | ENSG00000174837 | ATP7A | ENSG00000165240 |
| ADGRG4 | ENSG00000156920 | ATP8A2 | ENSG00000132932 |
| ADSSL1 | ENSG00000185100 | ATXN7 | ENSG00000163635 |
| AFF3 | ENSG00000144218 | AVPR2 | ENSG00000126895 |
| AGBL2 | ENSG00000165923 | AXIN2 | ENSG00000168646 |
| AHSG | ENSG00000145192 | BAALC | ENSG00000164929 |
| AIFM1 | ENSG00000156709 | BAK1 | ENSG00000030110 |
| AKAP3 | ENSG00000111254 | BDP1 | ENSG00000145734 |
| AKR1C3 | ENSG00000196139 | BICD1 | ENSG00000151746 |
| AKR1C5P | ENSG00000225418 | BPIFB1 | ENSG00000125999 |
| AL117338.1 | ENSG00000223470 | BRD3 | ENSG00000169925 |
| AL356585.2 | ENSG00000279124 | BRINP1 | ENSG00000078725 |
| AL671883.1 | ENSG00000229836 | BTBD10 | ENSG00000148925 |
| ALAS1 | ENSG00000023330 | BTBD11 | ENSG00000151136 |
| ALG1L2 | ENSG00000251287 | C10orf120 | ENSG00000183559 |
| ALK | ENSG00000171094 | C11orf24 | ENSG00000171067 |
| ALKBH1 | ENSG00000100601 | C11orf52 | ENSG00000149300 |
| ALKBH2 | ENSG00000189046 | C12orf42 | ENSG00000179088 |
| ALPK2 | ENSG00000198796 | C14orf178 | ENSG00000197734 |
| ANAPC1 | ENSG00000153107 | C14orf37 | ENSG00000139971 |
| ANKDD1A | ENSG00000166839 | C17orf47 | ENSG00000181013 |
| ANKRD2 | ENSG00000165887 | C1orf159 | ENSG00000131591 |
| ANKRD36 | ENSG00000135976 | C2orf16 | ENSG00000221843 |
| ANKRD36B | ENSG00000196912 | C4orf36 | ENSG00000163633 |
| ANKRD60 | ENSG00000124227 | C4orf50 | ENSG00000181215 |
| ANKS1B | ENSG00000185046 | C5orf42 | ENSG00000197603 |
| ANLN | ENSG00000011426 | C5orf46 | ENSG00000178776 |
| AP001062.3 | ENSG00000241728 | C6orf222 | ENSG00000189325 |
| AP2M1 | ENSG00000161203 | C8orf34 | ENSG00000165084 |
| AP3B1 | ENSG00000132842 | CACNA1H | ENSG00000196557 |
| APCDD1 | ENSG00000154856 | CALCR | ENSG00000004948 |
| AQP12B | ENSG00000185176 | CALCRL | ENSG00000064989 |
| AR | ENSG00000169083 | CAND1 | ENSG00000111530 |
| ARHGAP26 | ENSG00000145819 | CAPN3 | ENSG00000092529 |
| ARHGAP4 | ENSG00000089820 | CARD14 | ENSG00000141527 |
| ARHGAP45 | ENSG00000180448 | CARS | ENSG00000110619 |
| ARID3A | ENSG00000116017 | CBFA2T2 | ENSG00000078699 |

| | | | |
|----------|-----------------|----------|-----------------|
| CBWD5 | ENSG00000147996 | COL9A3 | ENSG00000092758 |
| CCDC146 | ENSG00000135205 | COLGALT1 | ENSG00000130309 |
| CCDC187 | ENSG00000260220 | CPB2 | ENSG00000080618 |
| CCDC54 | ENSG00000138483 | CPNE3 | ENSG00000085719 |
| CCDC88C | ENSG00000015133 | CPQ | ENSG00000104324 |
| CCIN | ENSG00000185972 | CRABP1 | ENSG00000166426 |
| CCT6P3 | ENSG00000234585 | CRY1 | ENSG00000008405 |
| CD24 | ENSG00000272398 | CRYL1 | ENSG00000165475 |
| CD248 | ENSG00000174807 | CSMD3 | ENSG00000164796 |
| CD44 | ENSG00000026508 | CSNK1A1L | ENSG00000180138 |
| CDC123 | ENSG00000151465 | CT47B1 | ENSG00000236446 |
| CDH16 | ENSG00000166589 | CTU2 | ENSG00000174177 |
| CDH20 | ENSG00000101542 | CXorf58 | ENSG00000165182 |
| CDH4 | ENSG00000179242 | CYBRD1 | ENSG00000071967 |
| CDHR3 | ENSG00000128536 | CYP2D7 | ENSG00000205702 |
| CDK20 | ENSG00000156345 | CYSLTR1 | ENSG00000173198 |
| CDSN | ENSG00000204539 | DAAM1 | ENSG00000100592 |
| CEL | ENSG00000170835 | DACT1 | ENSG00000165617 |
| CELA1 | ENSG00000139610 | DAPK1 | ENSG00000196730 |
| CENPIP1 | ENSG00000224778 | DAXX | ENSG00000204209 |
| CENPQ | ENSG00000031691 | DBNDD2 | ENSG00000244274 |
| CEP128 | ENSG00000100629 | DCAF17 | ENSG00000115827 |
| CES1P1 | ENSG00000228695 | DCHS2 | ENSG00000197410 |
| CFAP157 | ENSG00000160401 | DCP1A | ENSG00000272886 |
| CFHR1 | ENSG00000244414 | DDAH1 | ENSG00000153904 |
| CHRNB1 | ENSG00000170175 | DDO | ENSG00000203797 |
| CLDN14 | ENSG00000159261 | DDX4 | ENSG00000152670 |
| CLEC4D | ENSG00000166527 | DENND1B | ENSG00000213047 |
| CLIP1 | ENSG00000130779 | DEUP1 | ENSG00000165325 |
| CLSTN1 | ENSG00000171603 | DGKK | ENSG00000274588 |
| CLUH | ENSG00000132361 | DHRS3 | ENSG00000162496 |
| CMYA5 | ENSG00000164309 | DHRX | ENSG00000169084 |
| CNDP1 | ENSG00000150656 | DHX37 | ENSG00000150990 |
| CNN2 | ENSG00000064666 | DIDO1 | ENSG00000101191 |
| CNTN4 | ENSG00000144619 | DIMT1 | ENSG00000086189 |
| CNTNAP3B | ENSG00000154529 | DIRC3 | ENSG00000231672 |
| COBLL1 | ENSG00000082438 | DLAT | ENSG00000150768 |
| COL18A1 | ENSG00000182871 | DNAAF4 | ENSG00000256061 |
| COL22A1 | ENSG00000169436 | DNAH1 | ENSG00000114841 |
| COL24A1 | ENSG00000171502 | DNAH14 | ENSG00000185842 |
| COL26A1 | ENSG00000160963 | DNAH9 | ENSG00000007174 |
| COL4A2 | ENSG00000134871 | DNAJC11 | ENSG00000007923 |
| COL9A2 | ENSG00000049089 | DNMT3A | ENSG00000119772 |

| | | | |
|-----------|-----------------|----------|-----------------|
| DOCK2 | ENSG00000134516 | FMN2 | ENSG00000155816 |
| DPT | ENSG00000143196 | FMO2 | ENSG00000094963 |
| DPYSL2 | ENSG00000092964 | FMR1NB | ENSG00000176988 |
| DRAM2 | ENSG00000156171 | FZD6 | ENSG00000164930 |
| DRC3 | ENSG00000171962 | GAA | ENSG00000171298 |
| DROSHA | ENSG00000113360 | GABBR1 | ENSG00000204681 |
| DSCAML1 | ENSG00000177103 | GABRA1 | ENSG00000022355 |
| DSCR4 | ENSG00000184029 | GABRG3 | ENSG00000182256 |
| DSE | ENSG00000111817 | GALNTL5 | ENSG00000106648 |
| DTX2 | ENSG00000091073 | GAREM1 | ENSG00000141441 |
| DZIP1L | ENSG00000158163 | GDA | ENSG00000119125 |
| EFCAB3 | ENSG00000172421 | GFM1 | ENSG00000168827 |
| EIF2AK3 | ENSG00000172071 | GID4 | ENSG00000141034 |
| EIF4ENIF1 | ENSG00000184708 | GIPC1 | ENSG00000123159 |
| EMCN | ENSG00000164035 | GLYCTK | ENSG00000168237 |
| EML1 | ENSG00000066629 | GMPPA | ENSG00000144591 |
| ERICH6B | ENSG00000165837 | GNLY | ENSG00000115523 |
| ERMAP | ENSG00000164010 | GOLGA8B | ENSG00000215252 |
| ESD | ENSG00000139684 | GPC1 | ENSG00000063660 |
| EXD1 | ENSG00000178997 | GPN2 | ENSG00000142751 |
| EXOC3L4 | ENSG00000205436 | GPR161 | ENSG00000143147 |
| FAM129C | ENSG00000167483 | GRID2IP | ENSG00000215045 |
| FAM133A | ENSG00000179083 | GRIK4 | ENSG00000149403 |
| FAM149A | ENSG00000109794 | GRIN2B | ENSG00000273079 |
| FAM151A | ENSG00000162391 | GRIN3A | ENSG00000198785 |
| FAM162B | ENSG00000183807 | GSE1 | ENSG00000131149 |
| FAM189A1 | ENSG00000104059 | GSTM1 | ENSG00000134184 |
| FAM198B | ENSG00000164125 | GTF3C2 | ENSG00000115207 |
| FAM46C | ENSG00000183508 | GUCA2B | ENSG00000044012 |
| FAM53A | ENSG00000174137 | HAUS7 | ENSG00000213397 |
| FAM69C | ENSG00000187773 | HBG1 | ENSG00000213934 |
| FAM71A | ENSG00000162771 | HCFC1 | ENSG00000172534 |
| FANCM | ENSG00000187790 | HDAC4 | ENSG00000068024 |
| FARP1 | ENSG00000152767 | HHLA2 | ENSG00000114455 |
| FAT1 | ENSG00000083857 | HIVEP2 | ENSG0000010818 |
| FBN3 | ENSG00000142449 | HK1 | ENSG00000156515 |
| FBXL12 | ENSG00000127452 | HK3 | ENSG00000160883 |
| FBXO7 | ENSG00000100225 | HKR1 | ENSG00000181666 |
| FCGBP | ENSG00000275395 | HLA-DOA | ENSG00000204252 |
| FECH | ENSG00000066926 | HLA-DQA1 | ENSG00000196735 |
| FKTN | ENSG00000106692 | HMCN2 | ENSG00000148357 |
| FLG2 | ENSG00000143520 | HRG | ENSG00000113905 |
| FLT3LG | ENSG00000090554 | HRH3 | ENSG00000101180 |

| | | | |
|-----------|-----------------|-----------|-----------------|
| HS3ST3A1 | ENSG00000153976 | KLB | ENSG00000134962 |
| HSD17B11 | ENSG00000198189 | KLF2P1 | ENSG00000231240 |
| HTR4 | ENSG00000164270 | KLHL6 | ENSG00000172578 |
| IFNAR2 | ENSG00000159110 | KLK3 | ENSG00000142515 |
| IGFN1 | ENSG00000163395 | KLK7 | ENSG00000169035 |
| IGHG3 | ENSG00000211897 | KLRB1 | ENSG00000111796 |
| IGHV2-70D | ENSG00000211974 | KNTC1 | ENSG00000184445 |
| IGHV3-11 | ENSG00000211941 | KPTN | ENSG00000118162 |
| IGLV1-50 | ENSG00000211645 | KRT14 | ENSG00000186847 |
| IGLV1-51 | ENSG00000211644 | KRT24 | ENSG00000167916 |
| IGLV6-57 | ENSG00000211640 | KRTAP10-6 | ENSG00000188155 |
| IGSF8 | ENSG00000162729 | KRTAP25-1 | ENSG00000232263 |
| IL2RA | ENSG00000134460 | KRTAP4-16 | ENSG00000241241 |
| IL31RA | ENSG00000164509 | KRTAP4-8 | ENSG00000204880 |
| INTS1 | ENSG00000164880 | KRTAP5-5 | ENSG00000185940 |
| INTS6L | ENSG00000165359 | KRTAP5-7 | ENSG00000244411 |
| IPO11 | ENSG00000086200 | KY | ENSG00000174611 |
| IRAK1 | ENSG00000184216 | LAMA2 | ENSG00000196569 |
| IRF7 | ENSG00000185507 | LAMA3 | ENSG00000053747 |
| ISG15 | ENSG00000187608 | LAMA5 | ENSG00000130702 |
| ITGA7 | ENSG00000135424 | LAMB4 | ENSG00000091128 |
| ITGAE | ENSG00000083457 | LCA5L | ENSG00000157578 |
| ITGAM | ENSG00000169896 | LGALS8 | ENSG00000116977 |
| ITGAX | ENSG00000140678 | LHX4 | ENSG00000121454 |
| ITIH3 | ENSG00000162267 | LILRA4 | ENSG00000239961 |
| ITIH4 | ENSG00000055955 | LIMCH1 | ENSG00000064042 |
| JAG2 | ENSG00000184916 | LINC00969 | ENSG00000242086 |
| JARID2 | ENSG00000008083 | LINC01098 | ENSG00000231171 |
| JCAD | ENSG00000165757 | LINC01127 | ENSG00000281162 |
| KBTBD13 | ENSG00000234438 | LINC01226 | ENSG00000284543 |
| KBTBD2 | ENSG00000170852 | LINC02416 | ENSG00000257924 |
| KCNH2 | ENSG00000055118 | LINC02440 | ENSG00000255618 |
| KCNH3 | ENSG00000135519 | LINGO1 | ENSG00000169783 |
| KCNJ6 | ENSG00000157542 | LMO7 | ENSG00000136153 |
| KCNN3 | ENSG00000143603 | LONRF2 | ENSG00000170500 |
| KCNQ2 | ENSG00000075043 | LONRF3 | ENSG00000175556 |
| KDM4B | ENSG00000127663 | LOR | ENSG00000203782 |
| KIAA1324 | ENSG00000116299 | LOX | ENSG00000113083 |
| KIAA1755 | ENSG00000149633 | LPCAT2 | ENSG00000087253 |
| KIF13A | ENSG00000137177 | LRIG1 | ENSG00000144749 |
| KIF3C | ENSG00000084731 | LRP2 | ENSG00000081479 |
| KISS1 | ENSG00000170498 | LRPAP1 | ENSG00000163956 |
| KISS1R | ENSG00000116014 | LRRC32 | ENSG00000137507 |

| | | | |
|------------|-----------------|---------|-----------------|
| LRRRC37A2 | ENSG00000238083 | MUC4 | ENSG00000145113 |
| LRRK2 | ENSG00000188906 | MUC5B | ENSG00000117983 |
| LRSAM1 | ENSG00000148356 | MUC6 | ENSG00000184956 |
| LSS | ENSG00000160285 | MUT | ENSG00000146085 |
| LXN | ENSG00000079257 | MVD | ENSG00000167508 |
| LY75 | ENSG00000054219 | MYADML2 | ENSG00000185105 |
| LY75-CD302 | ENSG00000248672 | MYCT1 | ENSG00000120279 |
| MAGEA11 | ENSG00000185247 | MYL10 | ENSG00000106436 |
| MAGEA4 | ENSG00000147381 | MYO15A | ENSG00000091536 |
| MALRD1 | ENSG00000204740 | MYO15B | ENSG00000266714 |
| MAML2 | ENSG00000184384 | MYO18B | ENSG00000133454 |
| MAP3K10 | ENSG00000130758 | MYOM2 | ENSG00000036448 |
| MARCH10 | ENSG00000173838 | MYSM1 | ENSG00000162601 |
| MATN2 | ENSG00000132561 | NAP1L4 | ENSG00000205531 |
| MCF2L | ENSG00000126217 | NAT8B | ENSG00000204872 |
| MCM3AP | ENSG00000160294 | NAV2 | ENSG00000166833 |
| METRNL | ENSG00000176845 | NBPF12 | ENSG00000268043 |
| MFSD13A | ENSG00000138111 | NCOA3 | ENSG00000124151 |
| MGAT5B | ENSG00000167889 | NDNF | ENSG00000173376 |
| MIA3 | ENSG00000154305 | NDOR1 | ENSG00000188566 |
| MIB2 | ENSG00000197530 | NDUFAF1 | ENSG00000137806 |
| MIR646HG | ENSG00000228340 | NEDD9 | ENSG00000111859 |
| MIS18BP1 | ENSG00000129534 | NFRKB | ENSG00000170322 |
| MLF1 | ENSG00000178053 | NID1 | ENSG00000116962 |
| MMD2 | ENSG00000136297 | NID2 | ENSG00000087303 |
| MMP10 | ENSG00000166670 | NKX3-1 | ENSG00000167034 |
| MMP9 | ENSG00000100985 | NLRP1 | ENSG00000091592 |
| MNX1 | ENSG00000130675 | NLRP13 | ENSG00000173572 |
| MORC1 | ENSG00000114487 | NLRP4 | ENSG00000160505 |
| MORC4 | ENSG00000133131 | NLRP5 | ENSG00000171487 |
| MRGPRES | ENSG00000184350 | NLRP8 | ENSG00000179709 |
| MROH2B | ENSG00000171495 | NMB | ENSG00000197696 |
| MRPL14 | ENSG00000180992 | NOMO2 | ENSG00000185164 |
| MRPL18 | ENSG00000112110 | NOTCH2 | ENSG00000134250 |
| MRPL27 | ENSG00000108826 | NOXA1 | ENSG00000188747 |
| MRPL54 | ENSG00000183617 | NPAS2 | ENSG00000170485 |
| MT-CO2 | ENSG00000198712 | NPEPL1 | ENSG00000215440 |
| MTATP6P27 | ENSG00000234003 | NR2C2 | ENSG00000177463 |
| MTHFD1L | ENSG00000120254 | NRP2 | ENSG00000118257 |
| MTMR6 | ENSG00000139505 | NSG2 | ENSG00000170091 |
| MUC12 | ENSG00000205277 | NUAK1 | ENSG00000074590 |
| MUC16 | ENSG00000181143 | NUCB1 | ENSG00000104805 |
| MUC3A | ENSG00000169894 | NUDT6 | ENSG00000170917 |

| | | | |
|---------|-----------------|-----------|-----------------|
| NUP153 | ENSG00000124789 | PLXNA3 | ENSG00000130827 |
| NUP214 | ENSG00000126883 | PLXNB3 | ENSG00000198753 |
| NUSAP1 | ENSG00000137804 | PODXL2 | ENSG00000114631 |
| NUTM2F | ENSG00000130950 | POMZP3 | ENSG00000146707 |
| NWD2 | ENSG00000174145 | POTEE | ENSG00000188219 |
| OAF | ENSG00000184232 | PPIE | ENSG00000084072 |
| ODAPH | ENSG00000174792 | PPP1R15A | ENSG00000087074 |
| OGFR | ENSG00000060491 | PPP4R1 | ENSG00000154845 |
| OPRM1 | ENSG00000112038 | PPWD1 | ENSG00000113593 |
| OR10J1 | ENSG00000196184 | PRAMEF2 | ENSG00000120952 |
| OR1M1 | ENSG00000170929 | PRAMEF6 | ENSG00000232423 |
| OR2B6 | ENSG00000124657 | PRDM7 | ENSG00000126856 |
| OR52E1 | ENSG00000273085 | PRF1 | ENSG00000180644 |
| OR51B4 | ENSG00000183251 | PRKCD | ENSG00000163932 |
| OR56A4 | ENSG00000183389 | PRKCZ | ENSG00000067606 |
| OR5M1 | ENSG00000255012 | PROK1 | ENSG00000143125 |
| OR6C1 | ENSG00000205330 | PROP1 | ENSG00000175325 |
| OR7E24 | ENSG00000237521 | PRRG3 | ENSG00000130032 |
| OR9G4 | ENSG00000172457 | PSG3 | ENSG00000221826 |
| OSBPL2 | ENSG00000130703 | PSMD6 | ENSG00000163636 |
| OSER1 | ENSG00000132823 | PTCHD1 | ENSG00000165186 |
| P3H2 | ENSG00000090530 | PTGER1 | ENSG00000160951 |
| PARD3B | ENSG00000116117 | PTOV1 | ENSG00000104960 |
| PCSK4 | ENSG00000115257 | PTPN5 | ENSG00000110786 |
| PCSK5 | ENSG00000099139 | PTPRD | ENSG00000153707 |
| PCYT1A | ENSG00000161217 | PTRH1 | ENSG00000187024 |
| PDCD1 | ENSG00000188389 | PXMP4 | ENSG00000101417 |
| PDE4A | ENSG00000065989 | PXT1 | ENSG00000179165 |
| PEBP4 | ENSG00000134020 | PYY | ENSG00000131096 |
| PHRF1 | ENSG00000070047 | RAB11FIP5 | ENSG00000135631 |
| PIGM | ENSG00000143315 | RAB38 | ENSG00000123892 |
| PIGZ | ENSG00000119227 | RASA3 | ENSG00000185989 |
| PIP4K2A | ENSG00000150867 | RBM33 | ENSG00000184863 |
| PKD1L3 | ENSG00000277481 | RBP1 | ENSG00000114115 |
| PKDCC | ENSG00000162878 | RBPJ | ENSG00000168214 |
| PKP3 | ENSG00000184363 | RCAN1 | ENSG00000159200 |
| PLA2R1 | ENSG00000153246 | REV3L | ENSG00000009413 |
| PLEKHH2 | ENSG00000152527 | RGS10 | ENSG00000148908 |
| PLEKHH3 | ENSG00000068137 | RGS6 | ENSG00000182732 |
| PLEKHO2 | ENSG00000241839 | RGSL1 | ENSG00000121446 |
| PLET1 | ENSG00000188771 | RHPN2 | ENSG00000131941 |
| PLK5 | ENSG00000185988 | RIMKLA | ENSG00000177181 |
| PLXDC1 | ENSG00000161381 | RIOK2 | ENSG00000058729 |

| | | | |
|----------|-----------------|------------|-----------------|
| RMDN1 | ENSG00000176623 | SLC44A4 | ENSG00000204385 |
| RNASEL | ENSG00000135828 | SLC5A4 | ENSG00000100191 |
| RNF207 | ENSG00000158286 | SLC6A15 | ENSG00000072041 |
| RNF217 | ENSG00000146373 | SLC7A13 | ENSG00000164893 |
| RNF43 | ENSG00000108375 | SLC7A2 | ENSG00000003989 |
| RPL13A | ENSG00000142541 | SLC9C1 | ENSG00000172139 |
| RREB1 | ENSG00000124782 | SLCO1B1 | ENSG00000134538 |
| RSPH10B2 | ENSG00000169402 | SMCHD1 | ENSG00000101596 |
| RSRC2 | ENSG00000111011 | SMO | ENSG00000128602 |
| RTL9 | ENSG00000243978 | SNRPD2 | ENSG00000125743 |
| RUNX1 | ENSG00000159216 | SNTG1 | ENSG00000147481 |
| RYR2 | ENSG00000198626 | SNW1 | ENSG00000100603 |
| SAAL1 | ENSG00000166788 | SNX2 | ENSG00000205302 |
| SBSPON | ENSG00000164764 | SOD2 | ENSG00000112096 |
| SCML4 | ENSG00000146285 | SOWAHA | ENSG00000198944 |
| SCN9A | ENSG00000169432 | SPAG16 | ENSG00000144451 |
| SDHAP1 | ENSG00000185485 | SPANXB1 | ENSG00000227234 |
| SDK1 | ENSG00000146555 | SPATA13 | ENSG00000182957 |
| SELE | ENSG00000007908 | SPATA19 | ENSG00000166118 |
| SEMA3E | ENSG00000170381 | SPATA3 | ENSG00000173699 |
| SEMA6B | ENSG00000167680 | SPHKAP | ENSG00000153820 |
| SEPT14P1 | ENSG00000231523 | SPINT3 | ENSG00000101446 |
| SETD9 | ENSG00000155542 | SPRY2 | ENSG00000136158 |
| SETMAR | ENSG00000170364 | SPSB1 | ENSG00000171621 |
| SFR1 | ENSG00000156384 | SQSTM1 | ENSG00000161011 |
| SFTPFB | ENSG00000168878 | SRMS | ENSG00000125508 |
| SFTPC | ENSG00000168484 | SRPX | ENSG00000101955 |
| SGSM2 | ENSG00000141258 | SRRM4 | ENSG00000139767 |
| SH3BGR | ENSG00000185437 | SSC5D | ENSG00000179954 |
| SH3RF3 | ENSG00000172985 | ST3GAL1 | ENSG00000008513 |
| SHANK2 | ENSG00000162105 | ST6GALNAC6 | ENSG00000160408 |
| SIPA1 | ENSG00000213445 | STEAP4 | ENSG00000127954 |
| SIX1 | ENSG00000126778 | STK17B | ENSG00000081320 |
| SLC17A4 | ENSG00000146039 | STK40 | ENSG00000196182 |
| SLC20A1 | ENSG00000144136 | SUMO3 | ENSG00000184900 |
| SLC22A16 | ENSG00000004809 | SVIL | ENSG00000197321 |
| SLC25A26 | ENSG00000144741 | SWAP70 | ENSG00000133789 |
| SLC2A11 | ENSG00000133460 | SYNE1 | ENSG00000131018 |
| SLC35G4 | ENSG00000236396 | SYNM | ENSG00000182253 |
| SLC38A10 | ENSG00000157637 | SYNPO2 | ENSG00000172403 |
| SLC38A7 | ENSG00000103042 | TAAR9 | ENSG00000237110 |
| SLC38A8 | ENSG00000166558 | TAF1B | ENSG00000115750 |
| SLC41A2 | ENSG00000136052 | TAF4 | ENSG00000130699 |

| | | | |
|------------|-----------------|----------|-----------------|
| TAS2R46 | ENSG00000226761 | TRBV5-3 | ENSG00000211715 |
| TBC1D27P | ENSG00000128438 | TRBV7-9 | ENSG00000278030 |
| TBC1D4 | ENSG00000136111 | TRERF1 | ENSG00000124496 |
| TBC1D7 | ENSG00000145979 | TRGC2 | ENSG0000022719 |
| TCN2 | ENSG00000185339 | TRIM23 | ENSG00000113595 |
| TCTE3 | ENSG00000184786 | TRIM58 | ENSG00000162722 |
| TENM4 | ENSG00000149256 | TRMO | ENSG00000136932 |
| TEX14 | ENSG00000121101 | TRPA1 | ENSG00000104321 |
| TEX261 | ENSG00000144043 | TSHZ2 | ENSG00000182463 |
| THAP3 | ENSG00000041988 | TSPAN10 | ENSG00000182612 |
| TIMELESS | ENSG00000111602 | TSPAN18 | ENSG00000157570 |
| TIMM23B | ENSG00000204152 | TSPOAP1 | ENSG00000005379 |
| TIPRL | ENSG00000143155 | TTC16 | ENSG00000167094 |
| TLN2 | ENSG00000171914 | TTC25 | ENSG00000204815 |
| TLR1 | ENSG00000174125 | TTC30A | ENSG00000197557 |
| TLR7 | ENSG00000196664 | TUBA4B | ENSG00000243910 |
| TLR9 | ENSG00000239732 | TULP2 | ENSG00000104804 |
| TMEM121 | ENSG00000184986 | U52111.1 | ENSG00000232725 |
| TMEM140 | ENSG00000146859 | UGGT2 | ENSG00000102595 |
| TMEM184A | ENSG00000164855 | UGT2B17 | ENSG00000197888 |
| TMEM187 | ENSG00000177854 | UIMC1 | ENSG00000087206 |
| TMEM191B | ENSG00000278558 | UMODL1 | ENSG00000177398 |
| TMEM215 | ENSG00000188133 | UNC13D | ENSG00000092929 |
| TMEM221 | ENSG00000188051 | UNC5C | ENSG00000182168 |
| TMEM238L | ENSG00000263429 | USP17L7 | ENSG00000226430 |
| TMEM67 | ENSG00000164953 | USP34 | ENSG00000115464 |
| TMEM72 | ENSG00000187783 | UVSSA | ENSG00000163945 |
| TMEM9B-AS1 | ENSG00000254860 | VAMP4 | ENSG00000117533 |
| TMSB4X | ENSG00000205542 | VCX2 | ENSG00000177504 |
| TNC | ENSG00000041982 | VMA21 | ENSG00000160131 |
| TNPO3 | ENSG00000064419 | VPS37A | ENSG00000155975 |
| TOM1L1 | ENSG00000141198 | VPS41 | ENSG00000006715 |
| TOR2A | ENSG00000160404 | VPS53 | ENSG00000141252 |
| TP53I11 | ENSG00000175274 | VSIG10L2 | ENSG00000283703 |
| TP53TG5 | ENSG00000124251 | VWF | ENSG00000110799 |
| TPPP | ENSG00000171368 | WDPCP | ENSG00000143951 |
| TPTE2P1 | ENSG00000253771 | WDR11 | ENSG00000120008 |
| TRAF3IP2 | ENSG00000056972 | WDR73 | ENSG00000177082 |
| TRAJ37 | ENSG00000278661 | WFDC10A | ENSG00000180305 |
| TRAK1 | ENSG00000182606 | WFDC10B | ENSG00000182931 |
| TRBV12-5 | ENSG00000275158 | WFDC8 | ENSG00000158901 |
| TRBV13 | ENSG00000276405 | WHRN | ENSG00000095397 |
| TRBV16 | ENSG00000275243 | WNK1 | ENSG00000060237 |

| | |
|----------|------------------|
| WSB2 | ENSG00000176871 |
| WTAP | ENSG00000146457 |
| WWP1 | ENSG00000123124 |
| XYLT2 | ENSG00000015532 |
| YBX2 | ENSG00000006047 |
| YBX3 | ENSG000000060138 |
| YWHAEP7 | ENSG00000276715 |
| Z82186.1 | ENSG00000224715 |
| ZC3HC1 | ENSG00000091732 |
| ZDHHC17 | ENSG00000186908 |
| ZFAT | ENSG00000066827 |
| ZMYND12 | ENSG00000066185 |
| ZNF180 | ENSG00000167384 |
| ZNF195 | ENSG00000005801 |
| ZNF208 | ENSG00000160321 |
| ZNF219 | ENSG00000165804 |
| ZNF229 | ENSG00000278318 |
| ZNF254 | ENSG00000213096 |
| ZNF275 | ENSG00000063587 |
| ZNF276 | ENSG00000158805 |
| ZNF280C | ENSG00000056277 |
| ZNF324B | ENSG00000249471 |
| ZNF335 | ENSG00000198026 |
| ZNF385D | ENSG00000151789 |
| ZNF398 | ENSG00000197024 |
| ZNF415 | ENSG00000170954 |
| ZNF436 | ENSG00000125945 |
| ZNF440 | ENSG00000171295 |
| ZNF561 | ENSG00000171469 |
| ZNF679 | ENSG00000197123 |
| ZNF681 | ENSG00000196172 |
| ZNF697 | ENSG00000143067 |
| ZNF726 | ENSG00000213967 |
| ZNF727 | ENSG00000214652 |
| ZNF761 | ENSG00000160336 |
| ZNF79 | ENSG00000196152 |
| ZNF83 | ENSG00000167766 |
| ZNF831 | ENSG00000124203 |
| ZNF846 | ENSG00000196605 |
| ZSWIM2 | ENSG00000163012 |
| ZSWIM7 | ENSG00000214941 |
| ZZZ3 | ENSG00000036549 |

Tabela S6 Informações das variantes de alto impacto aminoacídico identificadas em amostras de DEO.

| Nº | Genes | Posição | Tipo | REF-ALT | HGVSc | HGVSp | ID | Consequência Funcional Calculada* |
|----|------------|------------------|------|---------|-------------------------------------|--|------------|---|
| 1 | DRAM2 | Chr1:111,119,614 | SNV | C→T | ENST00000477769.1: n.235+1G>A | – | rs6444081 | splice_donor_variant&non_coding_transcript_variant |
| 2 | OR10J1 | Chr1:159,440,550 | SNV | T→A | ENST00000423932.5: c.759T>A | ENSP00000399078.4: p.Cys253Ter | rs12409540 | stop_gained |
| 3 | KISS1 | Chr1:204,190,483 | DEL. | CT→C | ENST00000367194.4: c.417del | ENSP00000356162.4: p.Ter139TrpfsTer8 | rs71745629 | frameshift_variant&stop_lost |
| 4 | ACTN2 | Chr1:236,736,599 | DEL. | TC→T | ENST00000546208.5: c.95del | ENSP00000438384.2: p.Pro32LeufsTer12" | rs11355106 | frameshift_variant |
| 5 | GLYCTK | Chr3:52,291,743 | SNV | A→G | ENST00000486393.5: c.530-2A>G | – | rs2276834 | splice_acceptor_variant&NMD_transcript_variant |
| 6 | C4orf36 | Chr4:86,892,061 | SNV | A→T | ENST00000506308.5:" c.-74+2T>A | – | rs10034336 | splice_donor_variant |
| 7 | FAM198B | Chr4:158,171,600 | SNV | C→A | ENST00000296530.12" :c.-224-1G>T | – | ND | splice_acceptor_variant |
| 8 | OPRM1 | Chr6:154,107,531 | SNV | C→T | ENST00000419506.6: c.1231C>T | ENSP00000403549.2: p.Gin411Ter | rs677830 | stop_gained |
| 9 | SEPT14P1 | Chr7:63,148,830 | DEL. | TG→T | ENST00000458703.1: n.76-2del | – | rs57585159 | splice_acceptor_variant&non_coding_transcript_variant |
| 10 | TRPA1 | Chr8:72,020,024 | SNV | C→T | ENST00000523582.5: c.2706-1G>A | – | rs66527710 | splice_acceptor_variant |
| 11 | CBWD5 | Chr9:65,733,001 | DEL. | GTCAA→G | ENST00000377392.9: c.308_311del | ENSP00000366609.6: p.lie103LysfsTer7 | ND | frameshift_variant |
| 12 | AC113331.2 | Chr11:5,070,346 | INS | A→AC | ENST00000445557.1: c.777dup | ENSP00000476576.1: p.lle260HisfsTer50 | rs5789358 | frameshift_variant |
| 13 | TP53I11 | Chr11:44,903,399 | DEL. | GC→G | ENST00000354556.8: c.705del | ENSP00000346564.4: p.Ter237AspfsTer58 | rs5791652 | frameshift_variant&NMD_transcript_variant |
| 14 | TP53I11 | Chr11:44,903,808 | INS | C→CT | ENST00000354556.8: c.416dup | ENSP00000346564.4: p.Ala140GlyfsTer13 | rs11456363 | Frameshift_variant&NMD_transcript_variant |

| | | | | | | | | |
|----|------------|-------------------|------|-----------------------------------|--------------------------------------|---|-------------|---|
| 15 | OR5M1 | Chr11:56,613,070 | DEL. | CCAGA→C | ENST00000526538.1: c.429_432del | ENSP00000435416.1: p.Cys143TrpfsTer19 | rs72003051 | frameshift_variant |
| 16 | SHANK2 | Chr11:70,952,751 | DEL. | CG→G | ENST00000458632.2: c.585del | ENSP00000390636.2: p.Arg196GlyfsTer? | rs35132270 | frameshift_variant |
| 17 | VSIG10L2 | Chr11:125,948,513 | SNV | G→A | ENST00000638636.1: c.642G>A | ENSP00000491467.1: p.Trp214Ter | rs635329 | stop_gained |
| 18 | WNK1 | Chr12:865,189 | INS | T→TC | ENST00000340908.8: c.2220dup | ENSP00000341292.5: p.Phe741LeufsTer30 | rs11441897 | frameshift_variant |
| 19 | AC092821.1 | Chr12:9,560,817 | SNV | A→G | ENST00000518709.6: c.2551-1A>G | - | rs200950634 | splice_donor_variant&NMD_transcript_variant |
| 20 | CELA1 | Chr12:51,346,623 | DEL | CATAA→C | ENST00000293636.1: c.12_15del | ENSP00000293636.1: p.Tyr5AspfsTer14 | rs753836828 | frameshift_variant&splice_region_variant |
| 21 | CELA1 | Chr12:51,346,631 | INS | A→AAG | ENST00000293636.1: c.7_8insCT | ENSP00000293636.1: p.Val3AlafsTer18 | rs370927847 | frameshift_variant |
| 22 | CELA1 | Chr12:51,346,632 | INS | C→CG | ENST00000293636.1: c.6_7insC | ENSP00000293636.1: p.Val3ArgfsTer22 | rs148235680 | frameshift_variant |
| 23 | OR6C1 | Chr12:55,320,622 | INS | C→CA | ENST00000379668.3: c.24dup | ENSP00000368990.2: p.Glu9ArgfeTer10 | rs5798345 | frameshift_variant |
| 24 | FARP1 | Chr13:98,396,439 | DEL | TTGAGTGAGGA ACAGGAGGAG GA→T | ENST00000596580.2: c.*4_*25del | - | rs6145192 | Frameshift_variant |
| 25 | MCF2L | Chr13:113,033,972 | DEL | TC→T | ENST00000486806.1: c.202del | ENSP00000471913.1: p.Arg68GlufsTer40 | rs57693403 | frameshift_variant |
| 26 | SIX1 | Chr14:60,658,222 | SNV | A→G | ENST00000554986.2: c.2T>C | ENSP000004522700.2: :p.Met1? | rs10143202 | start_lost |
| 27 | GABRG3 | Chr15:27,271,593 | DEL | GAGTC→G | ENST00000554696.5: c.66_69del | ENSP00000451862.1: p.Gln22HisfsTer33 | rs76888516 | frameshift_variant |
| 28 | PKD1L3 | Chr16:71,947,511 | DEL | CTTTG→C | ENST00000620267.1: c.3695_3698del | ENSP00000480090.1: p.Thr1232ArgfsTer26 | ND | frameshift_variant |

| | | | | | | | | |
|----|------------|------------------|-----|-------------|--|--|-------------|---|
| 29 | AC009120.3 | Chr16:74,315,145 | SNV | G→T | ENST00000564455.1: n.25-1G>T | - | rs12596957 | splice_acceptor_variant&non_coding_transcript_variant |
| 30 | KRT24 | Chr17:40,701,882 | DEL | CA→C | ENST00000264651.2: c.666del | ENSP00000264651.2: p.Asn222LysfsTer11 | rs11309872 | frameshift_variant |
| 31 | LRRC37A2 | Chr17:46,512,902 | DEL | TC→T | ENST00000333412.3: c.192del | ENSP00000333071.3: p.Ser65ProfsTer69 | ND | frameshift_variant |
| 32 | MRPL27 | Chr17:50,368,313 | DEL | CCTGGTCAG→C | ENST00000503633.5: c.*16-8_*16-1del | - | rs111380254 | splice_acceptor_variant&intron_variant |
| 33 | GAREM1 | Chr18:32,136,366 | SNV | A→G | ENST00000583696.1: c.*48+2T>C | - | ND | splice_donor_variant |
| 34 | PCSK4 | Chr19:1,490,421 | SNV | A→C | ENST00000591687.2: n.24+2T>G, | - | rs806528 | splice_donor_variant&non_coding_transcript_variant |
| 35 | GIPC1 | Chr19:14,483,329 | SNV | A→G | ENST00000591245.1: c.-31+2T>C | - | ND | splice_donor_variant |
| 36 | MTATP6P27 | Chr19:12,506,671 | SNV | A→G | ENST00000416987.1: n.95+2A>G | - | rs12981461 | splice_donor_variant&non_coding_transcript_variant |
| 37 | RPL13A | Chr19:49,490,126 | SNV | G→T | ENST00000477613.6: n.94-1G>T | - | rs55991145 | splice_acceptor_variant&non_coding_transcript_variant |
| 38 | ZNF83 | Chr19:52,616,816 | SNV | C→T | ENST00000595939.5: c.*119-1G>A | - | rs191291023 | splice_acceptor_variant&NMD_transcript_variant |
| 39 | LAMA5 | Chr20:62,336,902 | SNV | G→A | ENST00000474128.1: c.*3+2C>T | - | rs2890083 | splice_donor_variant&NMD_transcript_variant |
| 40 | AP001062.3 | Chr21:44,329,821 | SNV | T→C | ENST00000444409.1: n.143-2A>G | - | rs73907170 | splice_acceptor_variant&non_coding_transcript_variant |
| 41 | Z82186.1 | Chr22:47,363,710 | SNV | C→G | ENST00000434707.1: n.274+1C>G | - | rs2879000 | splice_donor_variant&non_coding_transcript_variant |
| 42 | AIFM1 | ChrX:130,149,545 | SNV | A→G | ENST00000527892.5: c.130T>C | ENSP00000435955.6: p.Ter44ArgextTer5 | rs1139851 | stop_lost&NMD_transcript_variant |

ND, Variante não descrita; VAR., variante; REF-AT, Referência-Alteração; ID, Identificação da variante; Cr., Cromossomo; SNV, Variante de nucleotídeo único; Del., Deleção; INS., Inserção; HGVSc, nome da sequência codificante sugerida pela *Human Genome Variation Society* (HGVS); HGVSp, HGVSc, nome da sequência da proteína sugerida pela *Human Genome Variation Society* (HGVS); *Nomenclaturas das consequências funcionais calculadas das variantes dadas pelo programa *Ensembl* (MCLAR et al., 2016).

Tabela S7 Caracterização das variantes de alto impacto em relação aos dados clínicos e diagnósticos histopatológicos.

| Gene/Variante | Total Am. | Sexo | | Idade | | Tabaco | | Álcool | | Localização | | | | | Dx. Clínico | | | Dx. Histop. | | Origem | |
|-----------------|-----------|------|---|-------|------|--------|---|--------|---|-------------|----|----|----|---|-------------|----|----|-------------|----|--------|----|
| | | F | M | < 50 | > 50 | S | N | S | N | LI | MJ | RA | AB | P | LH | LV | LE | BG | AG | CH | BR |
| DRAM2 | | | | | | | | | | | | | | | | | | | | | |
| C→T | 2 | 2 | 0 | 0 | 2 | 2 | 0 | 2 | 2 | 0 | 0 | 1 | 1 | 0 | 1 | 1 | 0 | 0 | 2 | 2 | 0 |
| OR10J1 | | | | | | | | | | | | | | | | | | | | | |
| T→A | 2 | 1 | 1 | 1 | 1 | 2 | 0 | 2 | 0 | 0 | 1 | 1 | 0 | 0 | 1 | 1 | 0 | 0 | 2 | 1 | 1 |
| KISS1 | | | | | | | | | | | | | | | | | | | | | |
| CT→C | 4 | 2 | 2 | 1 | 3 | 3 | 1 | 3 | 1 | 2 | 1 | 0 | 1 | 0 | 3 | 1 | 0 | 1 | 3 | 2 | 2 |
| ACTN2 | | | | | | | | | | | | | | | | | | | | | |
| TC→T | 7 | 5 | 2 | 2 | 4 | 5 | 2 | 5 | 2 | 2 | 2 | 1 | 1 | 1 | 4 | 2 | 1 | 3 | 4 | 4 | 3 |
| GLYCTK | | | | | | | | | | | | | | | | | | | | | |
| A→G | 6 | 4 | 2 | 2 | 4 | 4 | 2 | 4 | 2 | 2 | 1 | 1 | 1 | 1 | 3 | 2 | 1 | 2 | 4 | 3 | 3 |
| C4orf36 | | | | | | | | | | | | | | | | | | | | | |
| A→T | 2 | 1 | 1 | 1 | 1 | 2 | 0 | 2 | 0 | 0 | 1 | 0 | 1 | 0 | 2 | 0 | 0 | 0 | 2 | 1 | 1 |
| FAM198B | | | | | | | | | | | | | | | | | | | | | |
| C→A | 1 | 1 | 0 | 1 | 0 | 1 | 0 | 1 | 0 | 0 | 1 | 0 | 0 | 0 | 1 | 0 | 0 | 1 | 0 | 0 | 1 |
| OPRM1 | | | | | | | | | | | | | | | | | | | | | |
| C→T | 3 | 2 | 1 | 0 | 3 | 2 | 0 | 2 | 0 | 1 | 0 | 1 | 1 | 0 | 1 | 2 | 0 | 0 | 3 | 2 | 1 |
| SEPT14P1 | | | | | | | | | | | | | | | | | | | | | |
| TG→T | 2 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 0 | 1 | 0 | 1 | 0 | 2 | 0 | 0 | 0 | 2 | 1 | 1 |

| | | | | | | | | | | | | | | | | | | | | | |
|-------------------|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| TRPA1 | | | | | | | | | | | | | | | | | | | | | |
| C→T | 3 | 2 | 1 | 0 | 3 | 2 | 1 | 2 | 1 | 1 | 0 | 1 | 1 | 0 | 1 | 2 | 0 | 0 | 3 | 2 | 1 |
| CBWD5 | | | | | | | | | | | | | | | | | | | | | |
| GTCAA→G | 1 | 0 | 1 | 0 | 1 | 0 | 1 | 0 | 1 | 1 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 1 | 0 | 1 |
| OR52E1 | | | | | | | | | | | | | | | | | | | | | |
| A→AC | 3 | 2 | 1 | 2 | 1 | 1 | 2 | 2 | 1 | 1 | 1 | 1 | 0 | 0 | 1 | 1 | 1 | 1 | 2 | 2 | 1 |
| TP53I11 | | | | | | | | | | | | | | | | | | | | | |
| GC→G | 6 | 5 | 1 | 3 | 3 | 4 | 2 | 4 | 2 | 2 | 1 | 1 | 1 | 1 | 3 | 0 | 3 | 5 | 1 | 5 | 1 |
| C→CT | 6 | 4 | 2 | 2 | 4 | 5 | 1 | 5 | 1 | 2 | 2 | 1 | 1 | 0 | 4 | 2 | 0 | 2 | 4 | 3 | 3 |
| OR5M1 | | | | | | | | | | | | | | | | | | | | | |
| CCAGA→C | 4 | 3 | 1 | 2 | 2 | 4 | 0 | 3 | 1 | 2 | 1 | 1 | 0 | 0 | 3 | 0 | 1 | 4 | 0 | 3 | 1 |
| SHANK2 | | | | | | | | | | | | | | | | | | | | | |
| CG→G | 3 | 2 | 1 | 1 | 2 | 2 | 1 | 2 | 1 | 1 | 0 | 1 | 1 | 0 | 1 | 2 | 0 | 0 | 3 | 2 | 1 |
| VSIG10L2 | | | | | | | | | | | | | | | | | | | | | |
| G→A | 1 | 0 | 1 | 0 | 1 | 0 | 1 | 0 | 1 | 1 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 1 | 0 | 1 |
| WNK1 | | | | | | | | | | | | | | | | | | | | | |
| T→TC | 7 | 5 | 2 | 3 | 4 | 5 | 2 | 6 | 1 | 4 | 1 | 1 | 1 | 0 | 4 | 2 | 1 | 3 | 4 | 5 | 2 |
| AC092821.1 | | | | | | | | | | | | | | | | | | | | | |
| A→G | 2 | 2 | 0 | 1 | 1 | 0 | 2 | 1 | 1 | 1 | 0 | 0 | 0 | 1 | 0 | 0 | 2 | 2 | 0 | 2 | 0 |
| CELA1 | | | | | | | | | | | | | | | | | | | | | |
| CATAA→C | 6 | 3 | 3 | 2 | 4 | 4 | 2 | 4 | 2 | 2 | 1 | 2 | 1 | 0 | 2 | 2 | 2 | 2 | 4 | 4 | 2 |
| A→AAG | 6 | 3 | 3 | 2 | 4 | 4 | 2 | 4 | 2 | 2 | 1 | 2 | 1 | 0 | 2 | 2 | 2 | 2 | 4 | 4 | 2 |
| C→CG | 6 | 3 | 3 | 2 | 4 | 4 | 2 | 4 | 2 | 2 | 1 | 2 | 1 | 0 | 2 | 2 | 2 | 2 | 4 | 4 | 2 |
| OR6C1 | | | | | | | | | | | | | | | | | | | | | |
| C→CA | 8 | 5 | 3 | 4 | 4 | 3 | 2 | 3 | 2 | 4 | 2 | 1 | 0 | 1 | 4 | 1 | 3 | 6 | 2 | 5 | 3 |

| | | | | | | | | | | | | | | | | | | | | | |
|-------------------|-----------|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| FARP1 | | | | | | | | | | | | | | | | | | | | | |
| TTGAGTGAGGAAC | | | | | | | | | | | | | | | | | | | | | |
| AGGAGGAGGA→T | 3 | 2 | 1 | 1 | 2 | 3 | 0 | 2 | 1 | 1 | 1 | 1 | 0 | 0 | 2 | 0 | 1 | 3 | 0 | 2 | 1 |
| MCF2L | | | | | | | | | | | | | | | | | | | | | |
| TC→T | 2 | 2 | 0 | 1 | 1 | 1 | 1 | 2 | 0 | 1 | 1 | 0 | 0 | 0 | 1 | 0 | 1 | 2 | 0 | 1 | 1 |
| SIX1 | | | | | | | | | | | | | | | | | | | | | |
| A→G | 10 | 7 | 3 | 4 | 6 | 7 | 3 | 7 | 3 | 4 | 2 | 2 | 1 | 1 | 5 | 2 | 3 | 6 | 4 | 7 | 3 |
| GABRG3 | | | | | | | | | | | | | | | | | | | | | |
| GAGTC→G | 7 | 5 | 3 | 2 | 5 | 4 | 3 | 4 | 3 | 4 | 1 | 1 | 0 | 1 | 3 | 1 | 3 | 6 | 1 | 5 | 2 |
| PKD1L3 | | | | | | | | | | | | | | | | | | | | | |
| CTTTG→C | 3 | 2 | 1 | 0 | 3 | 2 | 1 | 2 | 1 | 1 | 0 | 1 | 1 | 0 | 1 | 2 | 0 | 0 | 3 | 2 | 1 |
| AC009120.3 | | | | | | | | | | | | | | | | | | | | | |
| G→T | 1 | 0 | 1 | 1 | 0 | 1 | 0 | 1 | 0 | 0 | 1 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 1 | 0 | 1 |
| KRT24 | | | | | | | | | | | | | | | | | | | | | |
| CA→C | 8 | 6 | 2 | 4 | 4 | 6 | 2 | 6 | 2 | 3 | 2 | 1 | 1 | 1 | 5 | 0 | 3 | 6 | 2 | 5 | 3 |
| LRRC37A2 | | | | | | | | | | | | | | | | | | | | | |
| TC→T | 1 | 0 | 1 | 1 | 0 | 1 | 0 | 1 | 0 | 0 | 1 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 1 | 0 | 1 |
| MRPL27 | | | | | | | | | | | | | | | | | | | | | |
| CCTGGTCAG→C | 8 | 5 | 3 | 4 | 4 | 4 | 3 | 4 | 3 | 4 | 2 | 1 | 0 | 1 | 4 | 1 | 3 | 6 | 2 | 5 | 3 |
| GAREM1 | | | | | | | | | | | | | | | | | | | | | |
| A→G | 1 | 0 | 1 | 1 | 0 | 1 | 0 | 1 | 0 | 0 | 1 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 1 | 0 | 1 |
| PCSK4 | | | | | | | | | | | | | | | | | | | | | |
| A→C | 4 | 3 | 1 | 2 | 2 | 3 | 1 | 3 | 1 | 2 | 1 | 1 | 0 | 0 | 2 | 0 | 2 | 4 | 0 | 3 | 1 |

| | | | | | | | | | | | | | | | | | | | | | |
|-------------------|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| GIPC1 | | | | | | | | | | | | | | | | | | | | | |
| A→G | 1 | 0 | 1 | 1 | 0 | 1 | 0 | 1 | 0 | 0 | 1 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 1 | 0 | 1 |
| MTATP6P27 | | | | | | | | | | | | | | | | | | | | | |
| A→G | 2 | 1 | 1 | 0 | 2 | 1 | 1 | 1 | 1 | 1 | 0 | 1 | 0 | 0 | 0 | 2 | 0 | 0 | 2 | 1 | 1 |
| RPL13A | | | | | | | | | | | | | | | | | | | | | |
| G→T | 3 | 2 | 1 | 0 | 3 | 1 | 2 | 1 | 2 | 1 | 0 | 1 | 0 | 1 | 0 | 2 | 1 | 1 | 2 | 2 | 1 |
| ZNF83 | | | | | | | | | | | | | | | | | | | | | |
| C→T | 1 | 1 | 0 | 1 | 0 | 1 | 0 | 1 | 0 | 0 | 1 | 0 | 0 | 0 | 1 | 0 | 0 | 1 | 0 | 0 | 1 |
| LAMA5 | | | | | | | | | | | | | | | | | | | | | |
| G→A | 8 | 6 | 2 | 4 | 4 | 6 | 2 | 6 | 2 | 3 | 2 | 1 | 1 | 1 | 5 | 0 | 3 | 6 | 2 | 5 | 3 |
| AP001062.3 | | | | | | | | | | | | | | | | | | | | | |
| T→C | 1 | 1 | 0 | 0 | 1 | 1 | 0 | 1 | 0 | 0 | 0 | 0 | 1 | 0 | 1 | 0 | 0 | 0 | 1 | 1 | 0 |
| Z82186.1 | | | | | | | | | | | | | | | | | | | | | |
| C→G | 2 | 1 | 1 | 1 | 1 | 2 | 0 | 2 | 0 | 0 | 1 | 0 | 1 | 0 | 2 | 0 | 0 | 0 | 2 | 1 | 1 |
| AIFM1 | | | | | | | | | | | | | | | | | | | | | |
| A→G | 4 | 3 | 1 | 1 | 3 | 3 | 1 | 2 | 2 | 2 | 0 | 1 | 0 | 1 | 2 | 0 | 2 | 4 | 0 | 4 | 0 |

Total Am., Total de Amostras; Idade (anos); S., Sim; N., Não; LI., Língua; MJ., Mucosa jugal; RA., Rebordo alveolar; AB., Assoalho bucal; P., Palato; Dx., Diagnóstico; LH., Leucoplasia homogênea; LV., Leucoplasia verrucosa; LE., Leucoeritroplasia; Histop., Histopatológico; DBG., Displasia epitelial oral de baixo grau; DAG., Displasia epitelial oral de algo grau; CH., Chile; BR., Brasil. Destaque em cinza para as variantes observadas em mais de 5 amostras.

