Universidade Estadual de Campinas
FACULDADE DE ODONTOLOGIA DE PIRACICABA

DOUGLAS VICTORINO ESPOSITO

ESTUDO DE ELEMENTOS CIS-REGULATÓRIOS DE GENES ENVOLVIDOS COM O DESENVOLVIMENTO DENTÁRIO

STUDY OF CIS-REGULATORY ELEMENTS OF GENES INVOLVED IN TOOTH DEVELOPMENT

Piracicaba
2022

DOUGLAS VICTORINO ESPOSITO

ESTUDO DE ELEMENTOS CIS-REGULATÓRIOS DE GENES ENVOLVIDOS COM O DESENVOLVIMENTO DENTÁRIO

STUDY OF CIS-REGULATORY ELEMENTS OF GENES INVOLVED IN TOOTH DEVELOPMENT

Dissertação apresentada à Faculdade de Odontologia de Piracicaba/Instituto da Universidade Estadual de Campinas como parte dos requisitos exigidos para a obtenção do título de Mestre em Biologia Buco-Dental, na Área de Histologia e Embriologia.

Dissertation presented to the Faculty of Dentistry of Piracicaba/Institute of the University of Campinas in partial fulfillment of the requirements for the degree of Master in Oral-dental Biology, in the area of Histology and Embryology.

Orientador: Prof. Dr. Marcelo Rocha Marques

ESTE TRABALHO CORRESPONDE À VERSÃO FINAL DA DISSERTAÇÃO DEFENDIDA PELO ALUNO DOUGLAS VICTORINO ESPOSITO E ORIENTADA PELO PROF. DR. MARCELO ROCHA MARQUES.

PIRACICABA

2022

Ficha catalográfica
Universidade Estadual de Campinas
Biblioteca da Faculdade de Odontologia de Piracicaba
Marilene Girello - CRB 8/6159

Es65e

Esposito, Douglas Victorino, 1989-
Estudo de elementos cis-regulatórios de genes envolvidos com o desenvolvimento dentário / Douglas Victorino Esposito. – Piracicaba, SP : [s.n.], 2022.

Orientador: Marcelo Rocha Marques.
Dissertação (mestrado) – Universidade Estadual de Campinas, Faculdade de Odontologia de Piracicaba.

1. Crista neural. 2. Sequenciamento de cromatina por imunoprecipitação. 3. Elementos facilitadores genéticos. I. Marques, Marcelo Rocha, 1976-. II. Universidade Estadual de Campinas. Faculdade de Odontologia de Piracicaba. III. Título.

Informações para Biblioteca Digital

Título em outro idioma: Study of cis-regulatory elements of genes involved in tooth development
Palavras-chave em inglês:
Neural crest
Chromatin immunoprecipitation sequencing
Enhancer elements, genetic
Área de concentração: Histologia e Embriologia
Titulação: Mestre em Biologia Buco-Dental
Banca examinadora:
Marcelo Rocha Marques [Orientador]
Juliana Gusson Roscito
Nilva de Karla Cervigne Furlan
Data de defesa: 28-04-2022
Programa de Pós-Graduação: Biologia Buco-Dental

Identificação e informações acadêmicas do(a) aluno(a)
- ORCID do autor: https://orcid.org/0000-0003-1279-3987
- Currículo Lattes do autor: http://lattes.cnpq.br/7605573357746115

UNIVERSIDADE ESTADUAL DE CAMPINAS

Faculdade de Odontologia de Piracicaba

A Comissão Julgadora dos trabalhos de Defesa de Dissertação de Mestrado, em sessão pública realizada em 28 de abril de 2022, considerou o candidato DOUGLAS VICTORINO ESPOSITO aprovado.

PROF. DR. MARCELO ROCHA MARQUES

PROFª. DRª. JULIANA GUSSON ROSCITO

PROFª. DRª. NILVA DE KARLA CERVIGNE FURLAN

A Ata da defesa, assinada pelos membros da Comissão Examinadora, consta no SIGA/Sistema de Fluxo de Dissertação/Tese e na Secretaria do Programa da Unidade.

AGRADECIMENTOS

O presente trabalho foi realizado com apoio da Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq), processo nº 131559/2019-2”

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

Ao Prof. Dr. Marcelo Rocha Marques, pela orientação neste projeto, além de sua paciência, excelente suporte, e disponibilidade durante todas as etapas do mesmo.

À Faculdade de Odontologia de Piracicaba, na pessoa do seu Diretor, Prof. Dr. Francisco Haiter Neto, pela disponibilidade de sua infraestrutura, além de todos os docentes do programa de Biologia Buco-Dental, técnicos de laboratório e funcionários.

Aos membros da banca examinadora, Dra. Juliana Gusson Roscito, Dra. Nilva de Karla Cervigne, Dra. Ana Paula de Souza e Dra. Juliana do Santos Neves pela disposição em presenciar e participar desta defesa de mestrado.

A todas as pessoas que participaram indiretamente na execução deste trabalho, como minha família pelo suporte emocional e acolhimento, e aos meus amigos da FOP pelas conversas e pelos cafezinhos da tarde.

RESUMO

O desenvolvimento dentário é um processo complexo que foi conservado durante a evolução de diversas espécies de vertebrados. Embora já sejam conhecidos os genes que participam da formação dos dentes, o conhecimento de como estes genes são regulados, ainda é limitado. O entendimento da regulação dos genes que formam os dentes é importante para se determinar as causas das variabilidades entre dentes normais, bem como para entender a origem de algumas alterações que levam a formação de dentes defeituosos ou mesmo a ausência de elementos dentários como acontece com muitos indivíduos com agenesia dentária. O objetivo deste estudo foi realizar uma anotação genômica de candidatos a regiões regulatórias distais (*enhancers*) de genes chaves para o desenvolvimento dentário. Células ectomesenquimais da papila apical (SCAP), as quais são derivadas de células da crista neural craniana (CNCC) e que dão origem à dentina e polpa de raízes dentárias, foram obtidas de dentes terceiros molares humanos em formação que foram removidos e seriam descartados. Após a realização de ensaios de imunoprecipitação de cromatina (ChIP-Seq) para marcas de H3K27ac, H3K4me1, H3K4me3 e do cofator p300 de células SCAP, os dados obtidos foram analisados conjuntamente com seguintes dados públicos abertos: 1) de ChIP-Seq para as mesmas marcas obtidas para SCAP e de acessibilidade da cromatina (ATAC-Seq) de CNCC; 2) de conservação de regiões genômicas entre diferentes espécies de vertebrados e; 3) de dados de regiões genômicas já previamente identificadas com atividade de *enhancer*. A análise dos dados possibilitou identificar regiões do genoma que são candidatas a serem elementos regulatórios distais de genes como PAX9, MSX1, BMP4, LEF1, DLX1, DLX2, genes estes envolvidos com o desenvolvimento dentário em humanos e em outros vertebrados.

Palavras-chave: Células ectomesenquimais da papila apical; Células da crista neural craniana; ChIP-seq; *Enhancers*; Elemento cis-regulatório; Promotores; Desenvolvimento dentário

ABSTRACT

Tooth development is a complex process that has been conserved during the evolution of different vertebrate species. Although the genes involved in tooth formation are well known, the knowledge on how they are finely regulated is still fairly limited. Understanding the regulation of these tooth-forming genes is crucial in determining the causes of variability between normal teeth, as well as the origin of some genetic alterations that lead to defective teeth, or even the absence of teeth which occurs in individuals with tooth agenesis. This study aimed to create a genomic annotation of candidate distant regulatory regions (*enhancers*) of genes involved in the tooth formation process. Apical papilla ectomesenchymal stem cells (SCAP), which are derived from cranial neural crest cells (CNCC), and give rise to the dentin and pulp formed in dental roots, were obtained from human third molars still in development, which would have been discarded otherwise. Following assays for Chromatin Immunoprecipitation Sequencing (ChIP-Seq) targeting markers such as H3K27ac, H3K4me1, H3K4me3 and the p300 cofactor derived from SCAP, the data was compared in conjunction with the following open source data: 1) ChIP-Seq for the same markers evaluated for SCAP and Assay for Transposase-Accessible Chromatin using sequencing (ATAC-Seq) derived CNCC; 2) conserved sequence data between different vertebrate species; 3) and of genomic regions previously identified with *enhancer* activity. The conjoined analysis of the data allowed us to identify genomic regions that may be distal regulatory elements of genes such as PAX9, MSX1, BMP4, LEF1, DLX1, DLX2, which are involved in tooth development in humans and other vertebrates.

Keywords: Apical papilla ectomesenchymal stem cells; Cranial neural crest cells; ChIP-seq; *Enhancers*; Cis regulatory element; Promoters; Tooth development

SUMÁRIO

1 INTRODUÇÃO 9

2 REVISÃO DA LITERATURA 11

3 PROPOSIÇÃO 19

4 MATERIAL E MÉTODOS 21

5 RESULTADOS 25

6 DISCUSSÃO 76

7 CONCLUSÕES 81

REFERÊNCIAS 83

APÊNDICE 92

ANEXOS 97

Anexo 1 - Verificação de originalidade e prevenção de plágio 97

Anexo 2 - Parecer Consubstanciado do Comitê de Ética em Pesquisa – FOP 98

1 INTRODUÇÃO

Os dentes são órgãos essenciais responsáveis pela sobrevivência e diversidade em vertebrados devido às suas funções de corte, esmagamento e mastigação da comida, e também como sua utilidade em ataque e defesa (Bergqvist 2003, Mu 2021). A morfogênese dentária foi descrita primeiramente há mais de 100 anos e é muito similar entre todos os vertebrados (Owen, 1845; Leche, 1895; Jernvall, 2012). A formação do dente é regulada por interações epitélio-mesenquimais, na qual o mesênquima se deriva da crista neural e o epitélio do ectoderma (Soukup et al., 2008; Fraser et al., 2009; Jernvall 2012). No desenvolvimento, os dentes se tornam visíveis durante as fases de botão e de capuz, nas quais a área da coroa dentária aparece junto da invaginação do epitélio ao redor. O estágio de capuz é seguido pelo estágio de campânula, onde os padrões de cúspides tendem a surgir, dependendo da espécie de vertebrado. Logo após, os odontoblastos de origem mesenquimal, os ameloblastos de origem epitelial, se diferenciam na interface epitelial-mesenquimal para formar a dentina e o esmalte, respectivamente. Os tecidos dentais duros possuem a mesma composição entre vertebrados, sendo o esmalte até 98% de hidroxiapatita. (Jernvall, 2012).

A determinação do embasamento genético na qual se origina a diversidade animal é um grande desafio na biologia atual. Estudos têm demonstrado que o desenvolvimento dentário em diferentes espécies possui vias moleculares em comum, as quais são chave na regulação do formato e do número de dentes, e que durante a evolução foram apresentando variações espaciais e temporais nas vias de sinalização compartilhadas entre essas espécies. As similaridades presentes entre espécies se dão pela quantidade atual limitada de genes disponíveis nessas vias de sinalização (Jernvall, 2012). A procura por tais genes e as mutações as quais impulsionam a evolução animal é um objetivo importante na biologia (Carroll, 2008; Cleves et al., 2018). Além de mutações no código genético, a hipótese de cis-regulação propõe que mudanças nesse tipo de regulação são os principais catalisadores da evolução morfológica, já que essas mutações não causam pleiotropia antagonista, onde mutações em regiões codificantes de genes regulatórios desenvolvimentais podem gerar alelos tanto benéficos quanto prejudiciais para o organismo (Wray, 2007; Cleves et al., 2018). Embora alguns estudos conseguiram

demonstrar diferenças fenotípicas em variações de espécies através de diferentes alelos cis-regulatórios, pouco se sabe de como e porquê esses alelos são utilizados na natureza (Stern e Orgogozo, 2008; Martin e Orgogozo, 2013; Cleves et al., 2018).

A regulação da expressão genética durante o desenvolvimento é essencial no embasamento da vida biológica de um organismo complexo. Grande parte dessa regulação se dá através de elementos cis-regulatórios chamados *enhancers* (ou acentuadores), que são definidos como pequenas sequências de DNA as quais não darão origem à uma proteína, e que podem elevar a transcrição, independente da sua distância e orientação relativa ao seu promotor cognato (Schaffner, 2015; Long et al., 2016). *Enhancers* tendem a ter muitas similaridades com outros tipos de elementos cis-regulatórios, assim como os promotores, porém com a vantagem de ativar a transcrição sobre longas distâncias no genoma, permitindo uma única região promotora ser regulada por diversos *enhancers* em períodos diferentes do desenvolvimento do organismo (Kim e Shiekhattar, 2015; Long et al., 2016). Além disso, *enhancers* são bons alvos para mudanças evolutivas pois tendem a ser específicos para tipos celulares, permitindo o controle de certos genes em diferentes tecidos sem afetar a função desses genes em outros tecidos. *Enhancers* também costumam se encontrar em grupos de elementos redundantes no genoma, permitindo com que variações genéticas nos mesmos não aumentem o risco de letalidade no organismo (Levine, 2010; Wittkopp e Kalay, 2012; Long et al., 2016).

Embora existam muitas informações acerca de genes que são determinantes para o processo odontogênese, ainda é limitado o conhecimento sobre as regiões que regulam a expressão destes genes. O avanço no reconhecimento das regiões regulatórias de genes envolvidos com o desenvolvimento dentário, potencialmente pode melhorar o entendimento de como o refinamento no controle da expressão de genes odontogênicos tem impacto por exemplo no tamanho, forma e número de dentes em diferentes vertebrados. Portanto, este trabalho utilizou diferentes ferramentas para identificar sequências de DNA não codificantes que seriam candidatas a *enhancers* de genes envolvidos com o processo de odontogênese.

2 REVISÃO DA LITERATURA

2.1 Células da Crista Neural

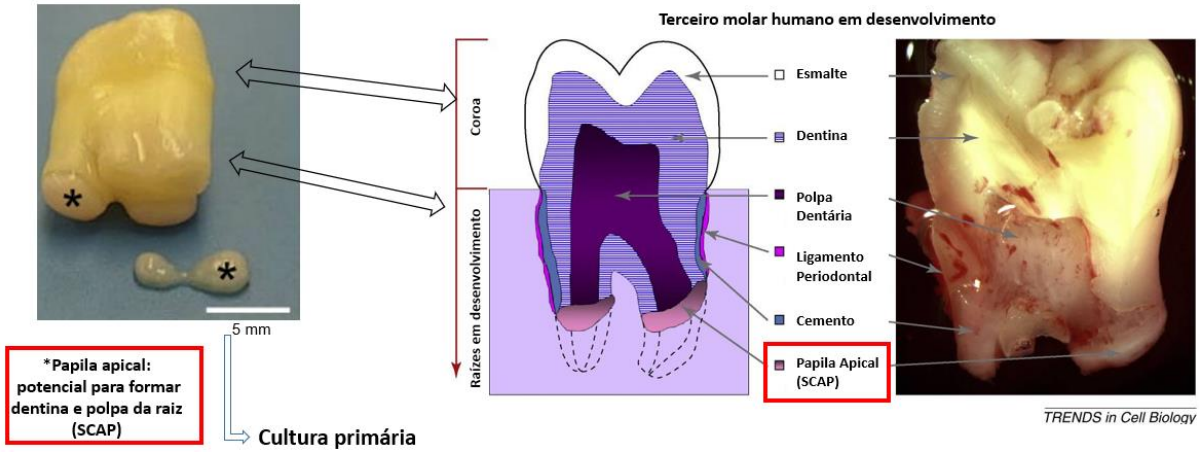
Células da crista neural são uma população celular específica em vertebrados, presente no início da embriogênese e localizadas na fronteira da placa neural, separando o neuroectoderma da epiderme (Gammill e Bronner-Fraser, 2003; Sauka-Spengler e Bronner-Fraser, 2008; Rada-Iglesias et al., 2013). Os progenitores de células da crista sofrem uma transição epitélio-mesenquimal, delaminando do tubo neural dorsal, migrando pelo corpo e adquirindo um vasto potencial de diferenciação, dando origem aos elementos do esqueleto craniofacial, a maior parte do elemento dentário, o ouvido médio, o sistema nervoso periférico, células pigmentares e certas estruturas cardíacas (Gammill e Bronner-Fraser, 2003; Sauka-Spengler e Bronner-Fraser, 2008; Rada-Iglesias et al., 2013). Em relação aos dentes e tecidos de suporte dentários, as CNCCs se diferenciam inicialmente em células ectomesenquimais e estas originam: dentina, polpa dentária, cemento, ligamento periodontal e osso alveolar. Interações entre grupos de células e diferentes domínios do embrião em desenvolvimento constituem mecanismos chave da morfogênese (Balic e Thesleff, 2015). As interações entre tecidos são sequenciais e recíprocas, e são mediadas por diversas famílias de moléculas sinalizadoras altamente conservadas, como as Wnts (proteínas de sinalização importantes no desenvolvimento embrionário), Bmps (*Bone Morphogenetic Proteins*), Shh (*Sonic Hedgehog Protein*) e Fgfs (*Fibroblast Growth Factor*). Essas moléculas regulam a expressão genética, e em particular, a expressão de fatores de transcrição que são compartilhados entre vários tecidos e órgãos.

2.2 Desenvolvimento Dentário

A geração da variedade de tecidos e órgãos é o resultado de combinações únicas e exclusivas de fatores de transcrição que regulam o destino da célula e determinam a identidade da célula alvo durante estágios sucessivos de um processo morfogenético (Davidson, 1993; Balic e Thesleff, 2015). O desenvolvimento dos órgãos formados a partir das regiões ectodérmicas e endodérmicas do embrião é regulado pela comunicação entre o epitélio e o mesênquima subjacente. Assim como acontece em diferentes órgãos, uma combinação de interações epiteliais-

mesenquimais recíprocas, regulam todos os aspectos de desenvolvimento dentário (Balic e Thesleff, 2015).

O processo de odontogênese é iniciado pelo epitélio oral, e este tem a capacidade de alterar o mesênquima subjacente a ele que é derivado de CNCCs (Mina e Kollar, 1987; Balic e Thesleff, 2015). As CNCCs se diferenciam em diversos tipos celulares que darão origem à diversas estruturas craniofaciais, como as células ectomesenquimais da papila apical (SCAP), que possuem um grande potencial proliferativo e respondem pela formação da maior parte do elemento dentário. O elongamento e a maturação da raiz dentária é guiada pelas SCAP, as quais se diferenciam em odontoblastos que darão origem à dentina, eventualmente dando origem à polpa do dente (Huang et al., 2008; Driesen et al., 2021) (Figura 1). Após a maturação completa do dente, as SCAP já terão sido integralmente diferenciadas, não estando mais presentes em um dente completamente formado.



Fontes: Adaptado de Egusa et al., 2012; Volponi et al., 2010 em *TRENDS in Cell Biology*.

Figura 1. Terceiro molar com raiz ainda em desenvolvimento, sendo as células ectomesenquimais da papila apical encontradas na extremidade da raiz, conforme essas células se diferenciam em odontoblastos para formação da polpa e da dentina radicular.

As SCAP possuem potenciais de diferenciação osteogênicos, adipogênicos e condrogênicos, além de ter capacidade imunomoduladora, fatores que demonstram uma capacidade regenerativa tecidual nesse tipo celular, além de ser uma fonte de células ectomesenquimais de fácil coleta (Huang et al., 2009; Gaudin et al., 2018; Liu et al., 2019; Fehrmann et al., 2020; Driesen et al., 2021). Sinais derivados da lâmina dentária estão envolvidos na transmissão de potencial odontogênico para o

mesênquima e são preservados ao longo de subsequentes estágios de desenvolvimento dentário em centros de sinalização epiteliais, possivelmente contribuindo para a manutenção do potencial odontogênico no mesênquima (Balic e Thesleff, 2015).

A identidade celular é determinada pela combinação única de fatores de transcrição expressas pela célula, e diferentes tipos celulares podem ser reprogramados por combinações específicas de fatores de transcrição (Iwafuchi-Doi e Zaret, 2014; Balic e Thesleff, 2015). A mudança no potencial odontogênico coincide com a condensação das células do mesênquima dentário, e é portanto provável que os mesmos sinais epiteliais estão envolvidos em ambos processos (Balic e Thesleff, 2015). Além disso, os fatores epiteliais Fgf8 e Sema3f (*Semaphorin-3F*) demonstraram atrair e repelir, respectivamente, as células dentárias mesenquimais, e o estímulo mecânico associado com a compactação das mesmas seriam suficiente para induzir expressão de alguns genes específicos do mesênquima dentário, como o MSX1, PAX9 (*Paired Box* 9) e BMP4 (Mammoto et al., 2011; Balic e Thesleff, 2015). Fatores de transcrição induzidos no mesênquima de camundongos, entre a lâmina dentária e os estágios de placódio includem Msx1 (*Msx Homeobox* 1, Msx2, Pax9, Lhx6 (*LIM Homeobox* 6), Lhx7, Dlx1 (*Distal-Less Homeobox* 1), Dlx2 e Runx2 (*Runt-related transcription factor* 2), e é provável que pelo menos alguns desses sejam responsáveis pela aquisição do potencial odontogênico no mesênquima. Todos esses genes listados são necessários para a morfogênese inicial do dente, como demonstrado em camundongos *knockouts* singulares e duplos (Bei, 2009; Balic e Thesleff, 2015). O desenvolvimento dentário é interrompido no estágio de lâmina dentária em alguns mutantes conjugados, incluindo Msx1;Msx2, Dlx1;Dlx2 e Lhx6;Lhx7 (interrupção somente dos molares), enquanto a interrupção durante o estágio de botão ocorre quando somente Pax9, Msx1 ou Runx2 é deletado. Adicionalmente, o desenvolvimento interrompido no estágio de lâmina em mutantes conjugados de Gli2;Gli3 (*GLI Family Zinc Finger*) e no estágio de botão em nocautes de Lef1 (*Lymphoid Enhancer Binding Factor*) indicam as funções necessárias dos mecanismos de Shh e Wtn/β-catenina. Ainda sim, é necessário demonstrar quais fatores de transcrição específicos podem programar a aquisição da identidade dentária mesenquimal (Balic e Thesleff, 2015).

A morfologia dos dentes é altamente conservada e análises dos padrões de expressão genética em epitélio dentário e mesênquima revelaram a associação de

diversos genes com o desenvolvimento dentário. O que irá definir uma dentição normal ou um fenótipo defeituoso, é em grande parte dependente da expressão de uma rede de genes relacionados aos dentes (Townsend et al., 2008; Line, 2003; Kapadia et al., 2007; Thesleff, 1995; Kurtén, 1963; Koussoulakou, 2009). Mesmo com o conhecimento em expressão espaço-temporal de genes específicos durante a formação do dente, nenhum gene foi diretamente relacionado experimentalmente ou naturalmente com a ontogênese ou a falta de algum dente específico (Line, 2003; Kapadia et al., 2007; Kurtén, 1963; Koussoulakou, 2009).

Defeitos do desenvolvimento geralmente ocorrem em tecidos dentários como resultado de mutações em genes que expressam fatores de transcrição dentários. Em humanos, esses defeitos podem aparecer sozinhos (isolados, como na hipodontia em mutantes de PAX9) ou em combinação com defeitos em outros tecidos (sindrômicos, como na hipodontia através de mutações em diversos genes, como MSX1, AXIN2 (Axin 2), EDA (*Ectodysplasin A*), PITX2 (*Paired Like Homeodomain 2*) e SHH) (Fleischmannova et al., 2008; Mostowska, 2003; Bailleul-Forestier, 2008; Koussoulakou, 2009). Na hereditariedade dos aspectos dentários, os aspectos são herdados de maneira mendeliana. A maioria das polimorfias dentárias são controladas por mais de 300 genes (Stock, 2001; Koussoulakou, 2009). Como exemplo, na displasia ectodérmica (EDA) e na displasia ectodérmica hipoidrótica (HED), os genes MSX1 e PAX9 cumprem papéis importantes na determinação do fenótipo da dentição (Courtney, 2005; Koussoulakou, 2009).

2.3 Doenças associadas à má-formação dentária

Uma das anomalias craniofaciais mais comum em humanos é a agenesia dentária, devido a falhas prematuras no desenvolvimento dentário. Estudos populacionais revelaram que até 20% da população pode perder pelo menos um terceiro molar, e que a dentição primária não é afetada como a permanente. Além dos terceiros molares, os segundos pré-molares e os incisivos superiores laterais também são afetados nas variantes de agenesia dentária (Ye e Attaie, 2016). Quando estão ausentes mais de seis dentes, a condição é chamada de hipodontia severa, ou oligodontia. Em casos extremos, a anodontia representa a perda total de todos os dentes.

A agenesia dentária pode ocorrer através de uma condição isolada, chamada não sindrômica, ou através de uma síndrome genética, onde há defeitos em diversos

genes, incluindo os relacionados ao desenvolvimento dentário. A agenesia dentária não sindrômica pode ocorrer esporadicamente, ou herdada de maneira mendeliana, e pode estar relacionada a outros defeitos como tamanho e formato dos dentes anormais e fendas labiais/palatinas (Ye e Attaie, 2016).

Além de defeitos em genes envolvidos no desenvolvimento dentário, alguns dos casos de agenesia dentária não sindrômica também podem ser atribuídos para fatores exógenos como quimioterapia, radioterapia, e tratamentos durante o desenvolvimento da dentição permanente (Letra et al, 2021). A agenesia dentária não sindrômica era comumente vista como uma condição monogênica, porém estudos mais recentes demonstraram que sua herança pode ser envolver dois ou mais genes. Defeitos em alguns genes como AXIN2, EDAR, FGFR1, IRF6, MSX1, PAX9 e WNT10A tendem a causar hipodontia/oligodontia, herdadas de maneira autossômica majoritariamente dominante. Defeitos em PAX9 também demonstraram fenótipos de dentes menores e irregulares (Letra et al, 2021).

Algumas síndromes também estão correlacionadas com os genes acima listados, como por exemplo os genes EDA/EDAR e WNT10A relacionados à Displasia Ectodérmica Anidrótica (Letra et al, 2021). Adicionalmente, o gene WNT10A também foi relacionado à Displasia Odonto-onico-dérmica e à Síndrome de Schopf-Schulz-Passarge (Letra et al, 2021). Outro exemplo é o gene MSX1 que foi relacionado à Síndrome de Wiktop, ou Displasia Ectodérmica Tipo 3 (Letra et al, 2021).

2.4 Expressão gênica e Elementos cis-regulatórios

Além da informação que pode ser herdada através da sequência de DNA, são conhecidas outras maneiras na qual a informação pode ser passada de geração para geração, as quais não envolvam a sequência de DNA propriamente dita, que é chamada de Epigenética (Zhang et al, 2021). As histonas, complexos proteicos nos quais o DNA se envolve para compactação e descompactação, podem sofrer modificações químicas em amino ácidos de suas porções N-terminais que afetam ultimamente a acessibilidade ao DNA e sua regulação. A metilação do DNA também pode regular a expressão gênica, e também é vista como uma modificação epigenética.

A determinação destas regiões com potencial cis-regulatório é diversa, sendo a conservação de uma dada região entre espécies diferentes um dos indicativos mais utilizados inicialmente (Chen et al., 2018). Além disso, ensaios que investigam a

acessibilidade da cromatina (Ensaio de cromatina transposase-acessível seguida de sequenciamento - ATAC-Seq) em regiões não codificantes também podem indicar quais estão mais acessíveis, e que pode conferir um potencial de interação com outras regiões, talvez de forma regulatória (Davie et al., 2015). Também podem ser investigadas modificações químicas nas histonas (através de imunoprecipitação da cromatina seguida de sequenciamento - ChIP-Seq), complexo de proteínas no qual o DNA se envolve, permitindo a compactação ou descompactação do mesmo.

As modificações epigenéticas variam entre acetilação, metilação, fosforilação e ubiquitinação em aminoácidos específicos da sequência peptídica de certas subunidades do complexo de histonas. Por exemplo, a acetilação da lisina 27 na histona 3 (H3K27ac) está geralmente associada à regiões genômicas que estão ativas, e quando essa marca se encontra em conjunto com a trimetilação da lisina 4 na histona 3 (H3K4me3), as marcas em conjunto denotam uma região transcricionalmente ativa, geralmente em genes essenciais, de acordo com o tipo de célula (Beacon et al., 2021). Outra marca prominente e que geralmente está associada à regiões com potencial cis-regulatório é a monometilação da lisina 4 na histona 3 (H3K4me1), que quando encontrada, pode indicar uma possível região de *enhancer* (Creyghton et al., 2010). Estudos prévios também identificaram que a presença da marca H3K4me1, com a ausência de H3K27ac, indica que a região não age como *enhancer*, mas sim que a mesma está preparada para se tornar uma região enhancer. Neste caso, assim que a acetilação da lisina 27 se der por uma histona acetiltransferase, possivelmente por diferenciação ou especialização da célula, essa região poderá ser utilizada como *enhancer* para um determinado gene (Creyghton et al., 2010). O cofator de transcrição p300 também pode ser investigado através de ChIP-Seq. Este cofator é uma proteína normalmente conjugada com a proteína CBP (*CREB Binding Protein*), ambas sendo da categoria de histona acetiltransferases, capaz de acetilar histonas. O complexo p300-CBP é conhecido por ser um regulador importante na atividade da transcrição mediada por RNA Polimerase II (Kalkhoven, 2004).

A identidade celular é majoritariamente determinada por conjuntos de genes específicos expressos por um tipo celular, o que depende de tipos distintos de elementos cis-regulatórios próximos ao gene (ex: promotores) e distantes ao gene (ex: *enhancers*). Atualmente é amplamente aceito que, enquanto os promotores determinam o nível transcricional de um gene, *enhancers* são responsáveis por

controlar quando e onde um gene deve ser expresso (Yánez-Cuna et al., 2013; Rada-Iglesias, 2014). Portanto, o estabelecimento de um perfil de expressão genética específico de um tipo celular depende principalmente de elementos cis-regulatórios distais. Consequentemente, levando em consideração essas funções importantes, *enhancers* cumprem papéis fundamentais durante a embriogênese, e deleções ou mutações em sequências de *enhancer* podem levar a doenças (Lettice et al., 2003; Rada-Iglesias, 2014). Porém, apesar da importância desenvolvimental e patológica dos *enhancers*, até recentemente, poucos elementos regulatórios foram caracterizados em humanos ou em outro organismo vertebrado. Novos avanços em tecnologia genômica têm expandido grandemente o catálogo de enhancers identificados em diversos tipos celulares e organismos (Rada-Iglesias, 2014).

Vários estudos já foram efetuados em genes que darão origem aos dentes, e também como ausência de tais genes surte alteração na formação dentária, mas o conhecimento sobre a regulação transcricional dos mesmos ainda é limitado. O potencial regulatório que certas regiões possuem em determinados genes, em um determinado tipo celular, constitui uma complexa rede de interações que pode vir a possuir certa variabilidade entre dois indivíduos de uma mesma espécie, ou até entre espécies, podendo por exemplo, causar diferenças no tamanho ou formato de um ou vários dentes. Vale lembrar também que mesmo regiões altamente conservadas entre espécies podem demonstrar uma grande diferença fenotípica.

Polimorfismos (SNPs) também são chave quando se analisa variabilidade intra e entre espécies que possam aumentar ou diminuir a afinidade de uma região regulatória por uma região codificante. Além disso, como a maioria desses genes está presente em grande parte dos vertebrados, é também interessante analisar como a variabilidade regulatória desses mesmos genes afeta o fenótipo entre espécies.

Alguns estudos prévios ilustram que a variação genética não-codificante tem o potencial de influenciar a regulação transcricional e a expressão genética, o que em alguns casos podem ter consequências patológicas e/ou influenciar a variação fenotípica (Rada-Iglesias, 2014). Estudos de casos específicos têm validado essas predições e demonstrado que mutações ou deleções afetando elementos regulatórios distais podem afetar características ecologicamente relevantes (Gompel et al., 2005; Shapiro et al., 2004; Attanasio et al., 2013; Prescott et al., 2015). Alterações nos elementos regulatórios causados pela introdução de variantes genéticas podem conferir mudanças regulatórias que resultam em variedades morfológicas. Avanços

recentes em tecnologia de sequenciamento têm revelado um número de potenciais variantes regulatórias que podem alterar a expressão genética. Porém, um número limitado de estudos demonstra dependência causal entre genética e mudanças morfológicas (Amano, 2020).

3 PROPOSIÇÃO

A comunicação entre os tecidos é mediada por moléculas sinalizadoras conservadas que agem em redes complexas de regulação genética com fatores de transcrição (Balic e Thesleff, 2015). Embora os genes e a sinalização molecular durante o desenvolvimento dentário sejam altamente conservados evolutivamente, é notável a variação morfológica que existe entre os dentes ao se comparar diferentes espécies (ex: mamíferos). Além disso, distúrbios de desenvolvimento dentário são comumente associados a mutações dentro de genes que participam da sinalização em alguma fase de formação do dente. No entanto, pouco se sabe a respeito das causas genéticas associadas com as variações morfológicas. Como já vem sendo descrito para outros órgãos, variações genéticas em regiões regulatórias de genes chaves do desenvolvimento dos dentes são, provavelmente, parte das causas da diversidade fenotípica dos elementos dentários entre as espécies. Além disso, variantes genéticas em regiões regulatórias de genes envolvidos com a odontogênese, podem estar relacionadas a distúrbios de desenvolvimento dentário.

3.1 Anotação de potenciais elementos cis-regulatórios de genes odontogênicos

O objetivo principal desta pesquisa foi de expandir o conhecimento sobre a regulação de transcrição de genes importantes no desenvolvimento dentário, realizando uma anotação de potenciais elementos cis-regulatórios distais (*enhancers*) de genes odontogênicos a partir de dados de marcação de histona de células ectomesenquimais da papila apical (SCAP).

3.2 Comparação de enhancers preditos com dados públicos de enhancers de células da crista neural

O objetivo secundário desta pesquisa foi efetuar a comparação dos *enhancers* (preditos) encontrados com dados públicos de *enhancers* de células da crista neural humanas (CNCC). Assim, podendo elucidar maneiras as quais essas células da crista neural têm seu destino comprometido para diferenciação em células ectomesenquimais voltadas para o desenvolvimento dentário, investigando regiões que regulam genes desse mesmo processo. Adicionalmente, realizamos a análise de

conservação entre vertebrados dos candidatos a *enhancers* encontrados e correlacionar com fenótipos dentários, quando houver dados disponíveis.

4 MATERIAIS E MÉTODOS

4.1 Amostras

O estudo foi aprovado pelo Comitê de Ética em Pesquisa da Faculdade de Odontologia de Piracicaba, Universidade Estadual de Campinas (CAAE 47869121.2.0000.5418)

A amostra (terceiro molar) foi obtida de um paciente saudável com idade de 18 anos (n=1), sem distinção de gênero, com indicação de extração de terceiro molar por razões ortodônticas ou profiláticas. O dente apresentava pelo menos 1/3 das raízes ainda não formadas e saiu íntegro após a exodontia.

O protocolo de isolamento das células ectomesenquimais da papila apical (SCAP) foi baseado em Liu et al. (2006) e Akiyama et al. (2012) e os ensaios de citometria de fluxo para confirmação do tipo celular foram realizados em estudo prévio. Esse estudo utilizou as SCAP por ser uma fonte de células ectomesenquimais de fácil coleta, com um grande potencial proliferativo e de diferenciação, as quais dão origem aos odontoblastos, que produzem a maior parte do elemento dentário como a dentina e a polpa dentária, e por serem (as SCAPs) um tipo celular proveniente das CNCCs através de diferenciação (Huang et al., 2008; Huang et al., 2009; Gaudin et al., 2018; Liu et al., 2019; Fehrmann et al., 2020; Driesen et al., 2021).

4.2 Avaliação das marcas de cromatina por meio de ChIP-Seq

O ensaio de ChIP-Seq foi realizado com o objetivo de avaliar marcas de cromatina nas SCAP que indiquem a atividade de *enhancer* e sítios de início de transcrição (TSS) ativos (H3K27ac, H3K4me3 e H3K4me1), além da presença do cofator transcricional p300. Esses ensaios foram realizados pela empresa Epigentek, localizada em Farmingdale, no estado de Nova Iorque, nos Estados Unidos da América. Brevemente, as células foram fixadas com paraformaldeído a 1% por 10 min à temperatura ambiente e, em seguida, foi adicionado glicina a uma concentração final de 0,125 M. A cromatina foi então sonicada a um tamanho de aproximadamente 0,5-2 kb. Um total de 3-5 µg de anticorpo (H3K4me1 - ab8895 - Abcam; H3K27ac - 39133 - Active Motif; H3K4me3 - 39159 - Active Motif; p300 - ab275378 - Abcam) foi adicionado à cromatina sonicada e incubados durante a noite sobre rotação leve a 4ºC. Foram mantidos 10% da cromatina usada, de cada reação ChIP, como DNA

input. Posteriormente, 75 µL de proteína A ou beads magnéticas de proteína G DynaI (dependendo da espécie de anticorpo e isotipo Ig) foram adicionados às reações ChIP e incubadas por mais quatro horas a 4ºC. As beads magnéticas foram lavadas e a cromatina, então eluída, passou pela reversão dos crosslinks e purificação. Em seguida, foram preparadas as bibliotecas de DNA para sequenciamento em um equipamento Illumina. Após o sequenciamento, a qualidade dos dados gerados foi verificada e os adaptadores removidos (*Trim Galore* 0.6.2, *Cutadapt* 2.3, e *Python* 3.7.3). Em seguida, os dados foram alinhados ao genoma humano de referência versão hg19 (*bowtie2* 2.3.5.1), as duplicatas foram removidas (*Samtools* 1.9), e então a denominação dos picos (*Peak calling*) foi realizada usando MACS 2.1.2. As regiões de interesse foram visualizadas por meio da ferramenta *The Integrative Genomics Viewer* (IGV), em arquivos no formato TDF, e também no *UCSC Genome Browser* em arquivos do formato WIG.

Adicionalmente, foi efetuada a remoção de "ruído" dos dados de marcação de histona das SCAP, utilizando a ferramenta SICER (Versão 1.1 - *Galaxy*), efetivamente subtraindo os dados de DNA input de cada uma das amostras, a fim de manter somente os picos mais estatisticamente importantes. As configurações na ferramenta SICER incluíram um *window size* de 200 bp, *fragment size* de 150 bp, *effective genome fraction* 0.8, *Gap size* de 200 bp e *Statistic threshold value* de 0.01.

4.3 Predição de *enhancers* de genes envolvidos com odontogênese

Após análises de sequenciamento, a predição foi dada verificando regiões com enriquecimentos para as marcas de histona previamente mencionadas (H3K27ac, H3K4me3, H3K4me1) próximos a genes já identificados como genes importantes para a formação dentária usando o website Bite It (http://bite-it.helsinki.fi/ - acessado em Junho de 2021), que são: PAX9, BMP4, MSX1, MSX2, DLX1, DLX2, AXIN2, INHBA, BARX1, BMP2, BMP6, WNT5A, LEF1. Também foram verificados genes também já identificados em relação à distúrbios do desenvolvimento dentário em humanos, utilizando bases de artigos e dados como o Pubmed e o GWAS (*Genome-Wide Association Studies*) (acessadas em Junho/Julho de 2021). Adicionalmente, foram analisadas regiões as quais apresentam atividade de *enhancer*, principalmente na região orofacial, já previamente estudadas e depositadas da base de dados VISTA *Enhancer Browser* (acessada em Julho de 2021).

Além das marcas de histona, foram incluídos também parâmetros de conservação destas regiões entre diferentes espécies de vertebrados (*Vertebrate MultiZ Alignment & Conservation (100 species)*), os dados de hipersensibilidade à DNAseI (*DNaseI Hypersensitivity clusters in 125 cell types from ENCODE (v3)*) para identificar regiões do genoma que não estão ocupadas por complexos proteicos/regiões que estão mais acessíveis, e o enriquecimento de motifs (sequências reconhecidas) para fatores de transcrição. Os dados de conservação foram gerados originalmente usando a ferramenta *MultiZ* da *UCSC/Penn State Bioinformatics*. Os dados de hipersensibilidade à DNAseI à partir da base de dados ENCODE (v3). Por último, o enriquecimento de motifs foi gerado com o *Transfac Matrix Database* (v7.0) criado pela *Biobase*. Estes três últimos tracks já estão todos integrados e disponíveis para visualização no *UCSC Genome Browser*.

4.4 Comparação com *enhancers* da células da crista neural (CNCC)

As regiões selecionadas foram comparadas às regiões de *enhancers* enriquecidas e descritas em células da crista neural, assim demonstradas por Prescott et al. (2015), com o objetivo de determinar se as CNCC possuem regiões regulatórias pré-definidas, que possam estar relacionadas à regulação transcricional de genes envolvidos no processo de desenvolvimento dentário. Foram utilizados os dados de marcação de H3K27ac, H3K4me3, H3K4me1, p300, e ATAC-Seq, provenientes de células-tronco embrionárias H9 diferenciadas em CNCCs, e iPSCs diferenciadas em CNCCs, de Prescott et al. (2015), depositados no *GEO* (*Gene Expression Omnibus*) *Accession number* GSE70751 (NCBI) (acessado em Junho de 2021), como comparação para com os nossos dados das células SCAP. Para possibilitar o uso e a comparação dos dados de marcação de histona de CNCC de Prescott et al. (2015), foi necessário converter o formato dos mesmos de WIG para Bedgraph, utilizando a ferramenta Wig-to-Bedgraph 1.0.1 (*Galaxy*). Foi efetuada a intersecção das regiões com enriquecimento para as marcas de histona já descritas em CNCC (dados de Prescott et al. (2015)), com as regiões enriquecidas em SCAP utilizando a ferramenta Intersect 1.0.0 (*Galaxy*), com no mínimo 1 bp de intersecção, para demonstrar as regiões similares nas quais essas marcas estão nos dois tipos celulares. Após a sobreposição dos dados experimentais das SCAPs com os dados das CNCCs, as regiões foram selecionadas manualmente, e tiveram suas marcações de histona categorizadas em uma tabela, de acordo com a cobertura total da região

selecionada para cada marca de ChIP, sendo separadas em regiões com cobertura total, cobertura parcial, ou sem a presença de sinal.

Também foi investigada a topologia das marcas estudadas (H3K27Ac, H3K4me3, HEK4me1 e p300), para determinar a porcentagem de cobertura das mesmas no genoma humano, utilizando a ferramenta *ChIP Seeker* 1.18.0 (*Galaxy*), nas células SCAP e CNCC (dados de Prescott et al. (2015)). Foram utilizadas as configurações padrões da ferramenta *ChIP Seeker*, além de incluir em consideração a anotação de regiões para o genoma humano inteiro do ENCODE, Release 19 (para o genoma hg19), em formato de arquivo GTF, que serviu de base para nomear as regiões enriquecidas nos dados de ChIP-Seq das SCAP e CNCC. Também foi efetuada uma análise de *ChIP Seeker* com as mesmas configurações acima nas intersecções das marcas de histona entre SCAP e CNCC.

Por último, foi feito um levantamento através de bases de dados como o GWAS, ou PubMed, de polimorfismos (SNPs) já investigados previamente em outros estudos os quais possuem efeitos nos genes de interesse para desenvolvimento dentário e craniofacial. Para investigar se tais SNPs alteram a capacidade de fatores de transcrição de interagirem com as possíveis regiões regulatórias afetadas, foi feita uma análise de motifs dessas mesmas regiões, com e sem SNPs, utilizando a ferramenta STAMP (http://www.benoslab.pitt.edu/stamp/), usando o input de sequência consenso, filtrando para exibir os dados de fatores de transcrição humanos somente. As sequências tinham tamanho de 10 bp, onde a base afetada se encontra na posição 4 ou 5 dessa sequência.

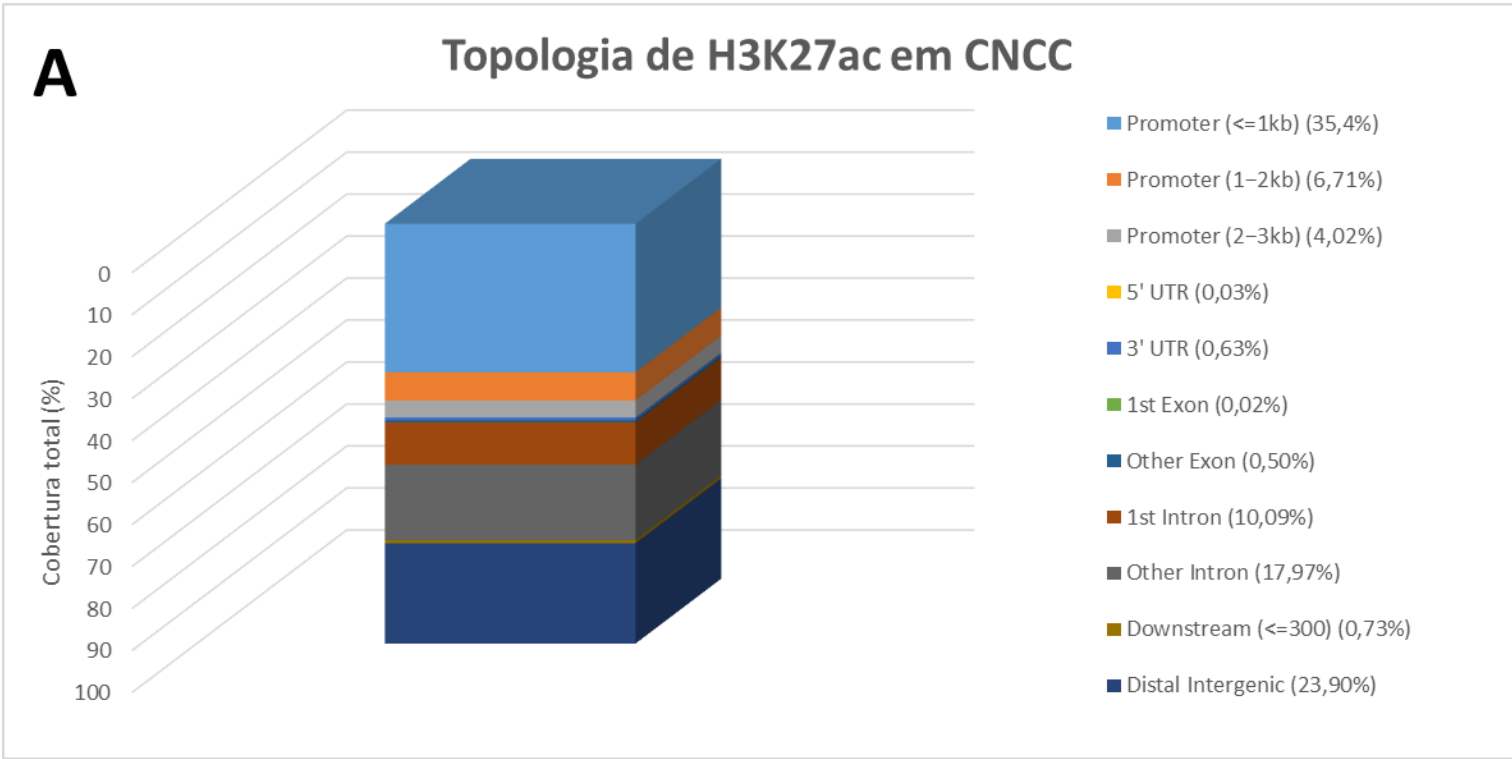
5 RESULTADOS

5.1 Topologia das marcas de histona e de cofator de transcrição p300

As marcas de histona geralmente estão distribuídas ao longo de regiões genômicas onde servirão como sinalizadores para fatores de transcrição que alteram a conformação da cromatina. Certas marcas, quando presentes em conjunto, podem representar um sítio de início de transcrição próximo à um promotor, ou uma região cis-regulatória, por exemplo (Andersson e Sandelin, 2020).

5.1.1 Marca de histona H3K27ac

Ao analisar a marca de histona H3K27ac em CNCC, se observa um enriquecimento de 46,13% em regiões promotoras, 28,06% para regiões intrônicas, 23,90% para regiões distais intergênicas e menos de 1% para exons e regiões 5’ e 3’ não traduzidas (Figura 2A). Nas SCAP, do total de marcas de histona H3K27ac 31,19% se encontram em regiões promotoras, 39,45% para regiões distais intergênicas e cerca de 3% para regiões exônicas e 5’ e 3’ não traduzidas Figura 2B).



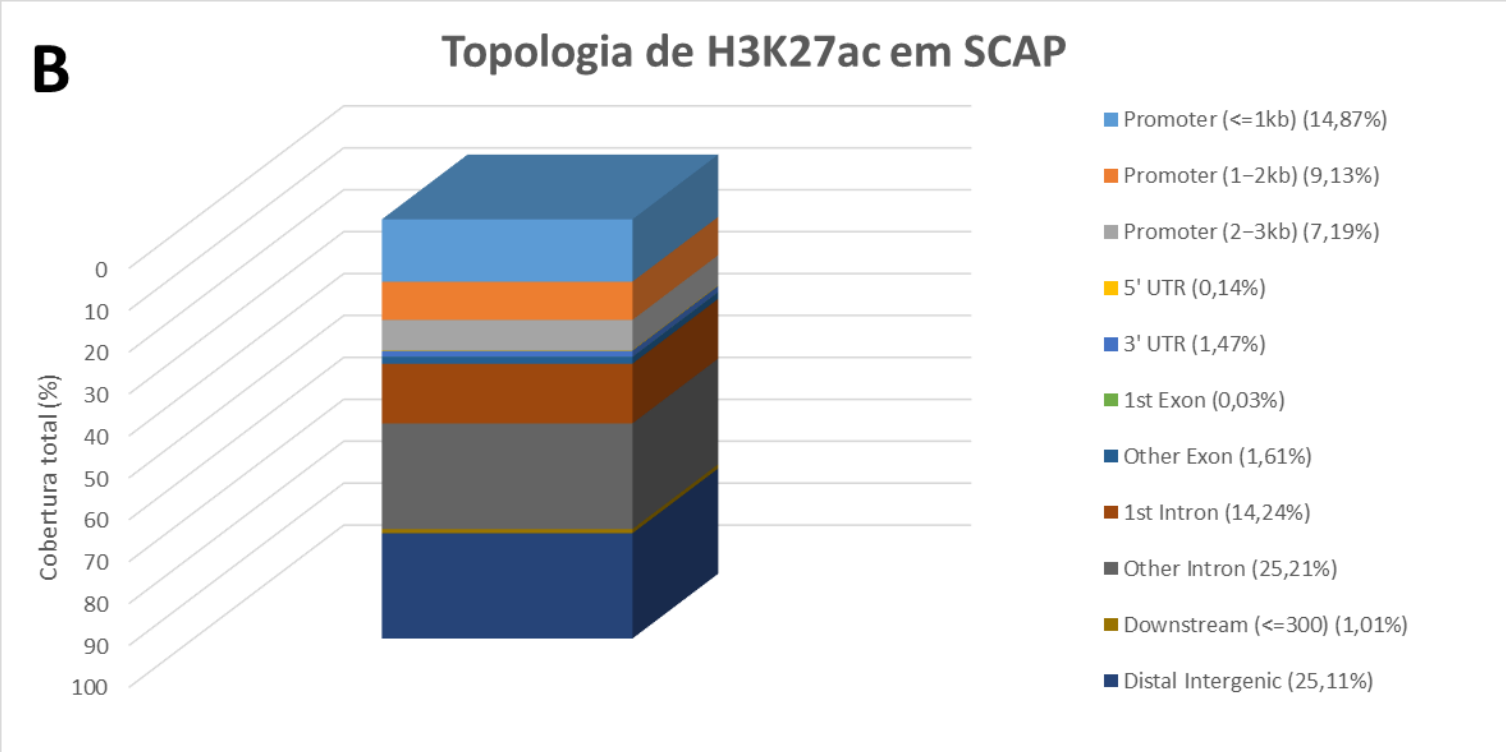


Figura 2. Topologia da marca H3K27ac ao longo dos genomas de **A)** CNCC e **B)** SCAP, com porcentagens obtidas através da ferramenta *ChIP Seeker*.

Quando as marcas de H3K27ac nas CNCCs e SCAPs são sobrepostas, existe um enriquecimento de aproximadamente 62,16% em regiões promotoras, 21,86% para regiões intrônicas e 13,88% para regiões distais intergênicas, e o restante para exons e regiões 5’ e 3’ não traduzidas (Figura 3).

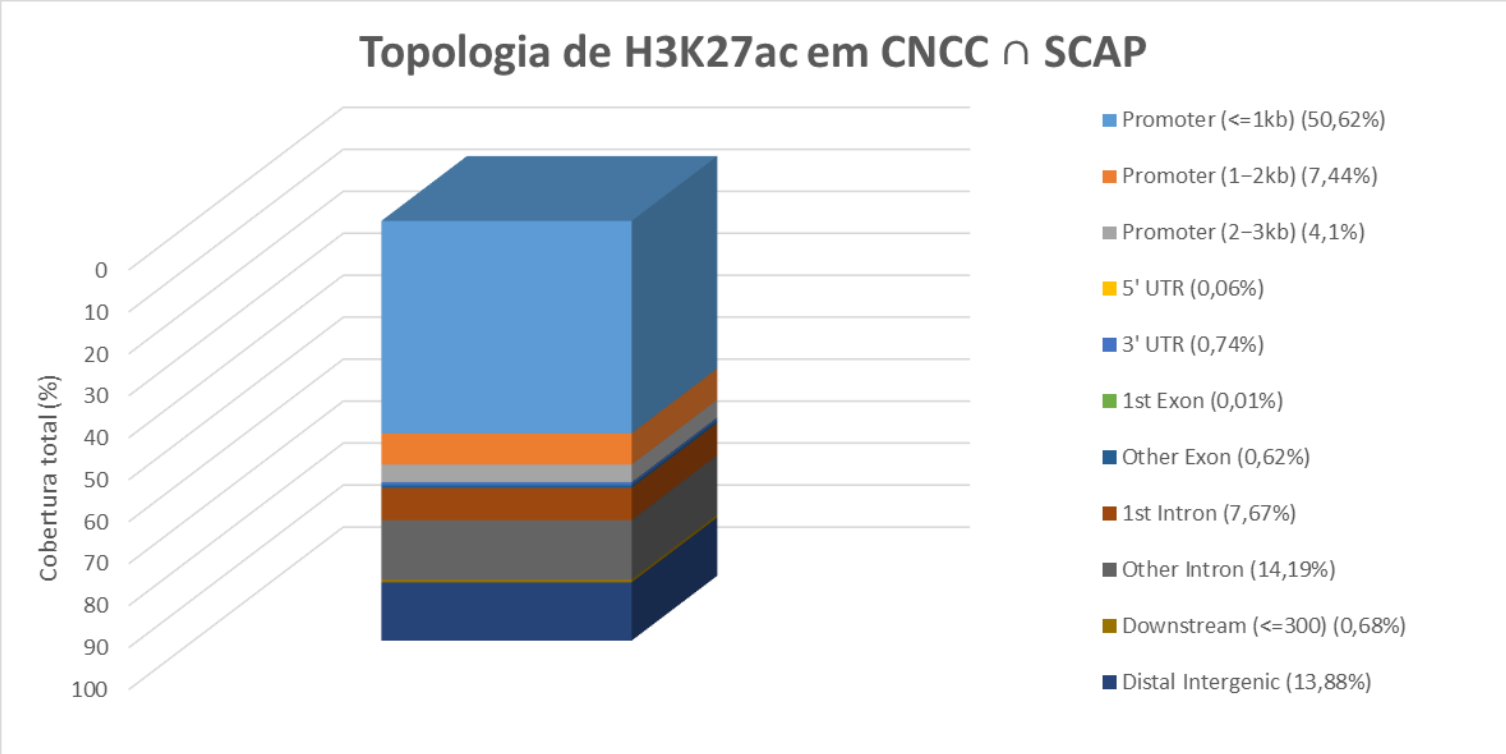
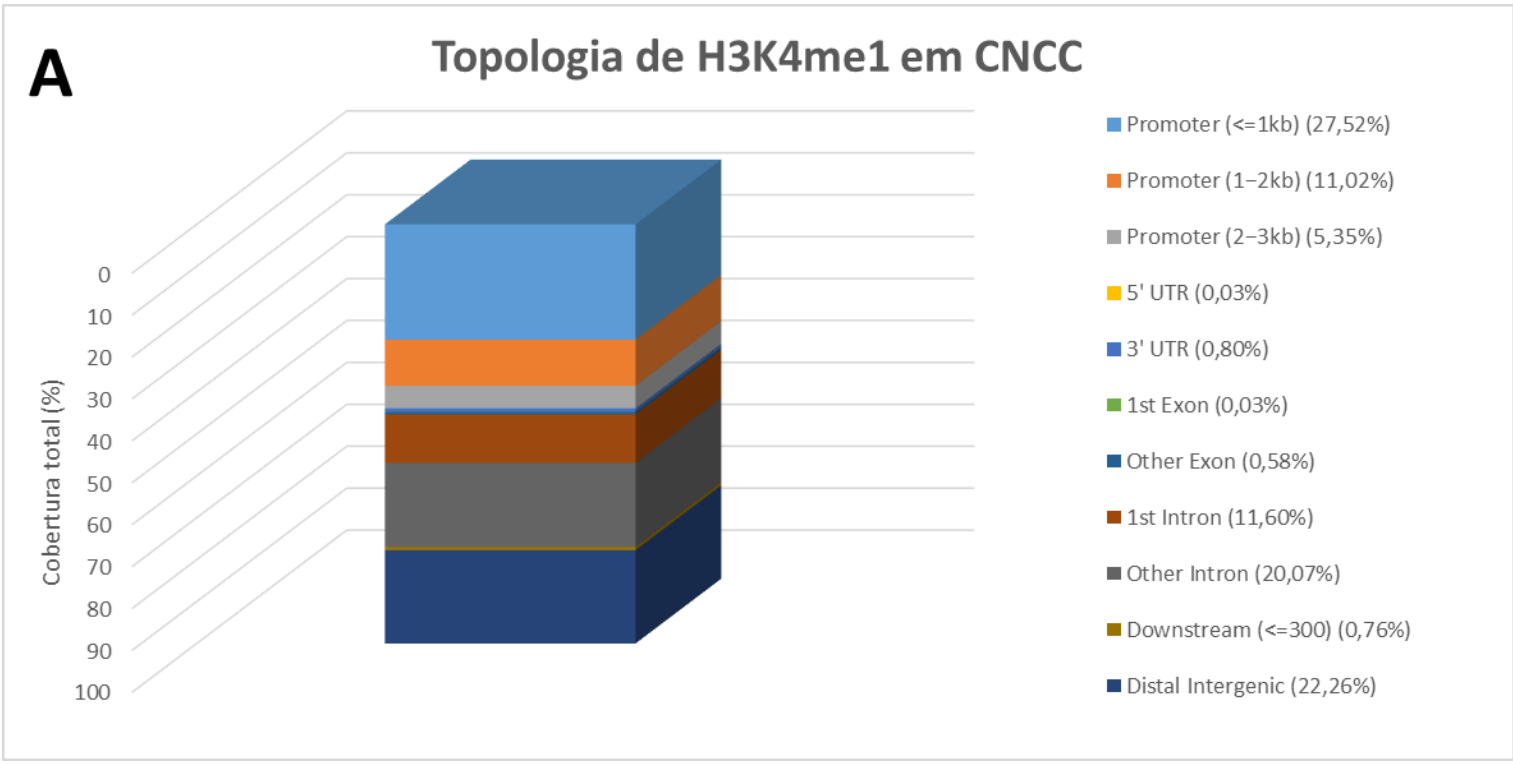


Figura 3. Topologia da marca H3K27ac ao longo dos genomas de CNCC e SCAP, filtrando apenas para regiões que se sobrepõem nos dois tipos celulares, com porcentagens obtidas através da ferramenta *ChIP Seeker*.

5.1.2 Marca de histona H3K4me1

Dos dados de marcação de histona H3K4me1 em CNCC, 43,89% foram encontradas em regiões promotoras, 31,67% em regiões intrônicas, 22,26% em regiões distais intergênicas e cerca de 2% em regiões exônicas e regiões 5’ e 3’ não traduzidas Figura 4A). Em relação às SCAP, cerca de 34,28% das marcas de H3K4me1 estão presentes em regiões promotoras, 39,9% em regiões intrônicas, 22,03% em regiões distais intergênicas, e aproximadamente 4% das marcas foram encontradas em regiões de exon e regiões 3’ e 5’ não traduzidas (Figura 4B).



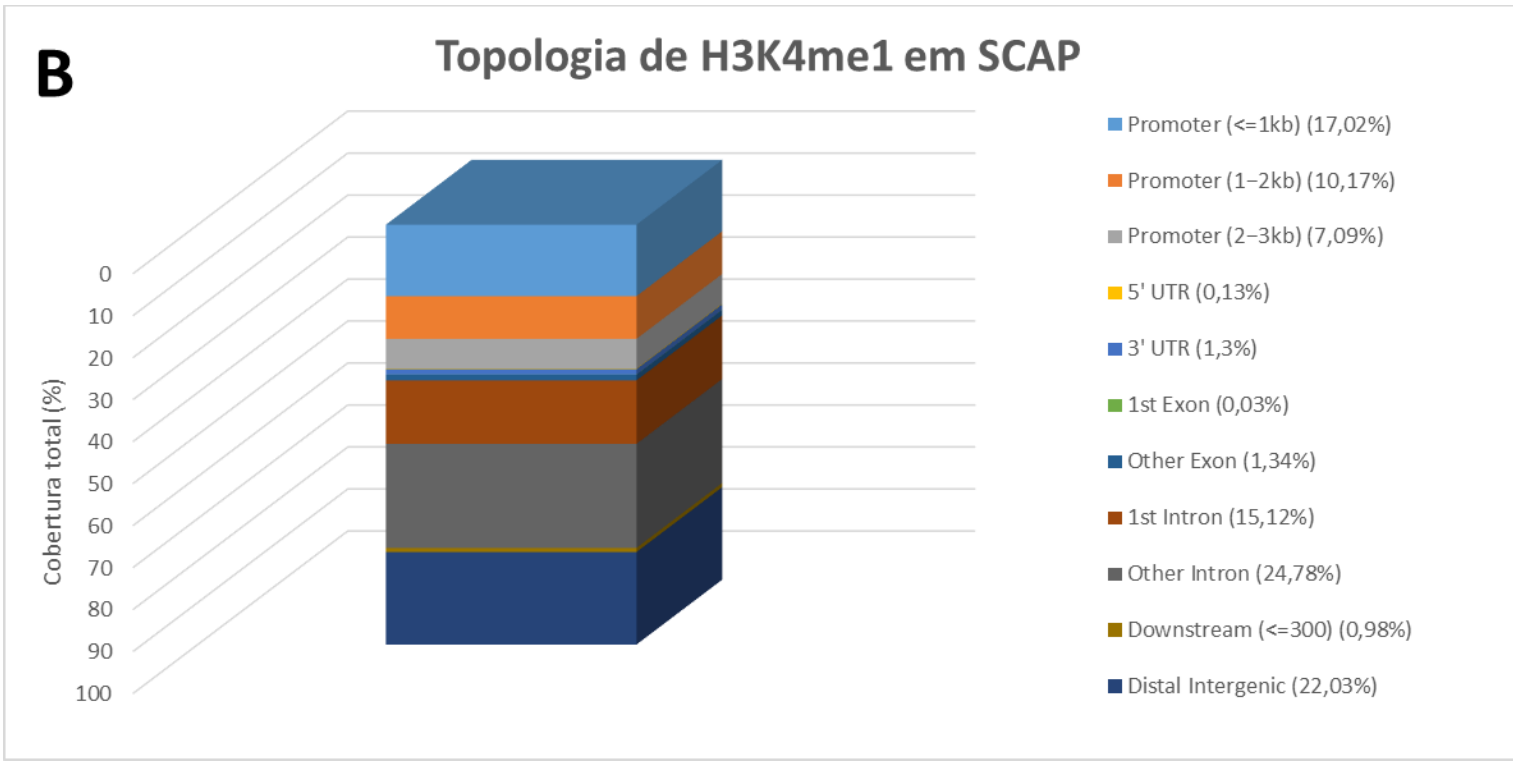


Figura 4. Topologia da marca H3K4me1 ao longo dos genomas de **A)** CNCC e **B)** SCAP, com porcentagens obtidas através da ferramenta *ChIP Seeke*r.

Quando as marcas de H3K4me1 nas CNCCs e SCAPs são sobrepostas, há um enriquecimento de aproximadamente 54,45% em regiões promotoras, 28,08% para regiões intrônicas, 15,13% para regiões distais intergênicas, e o restante para exons e regiões 5’ e 3’ não traduzidas (Figura 5).

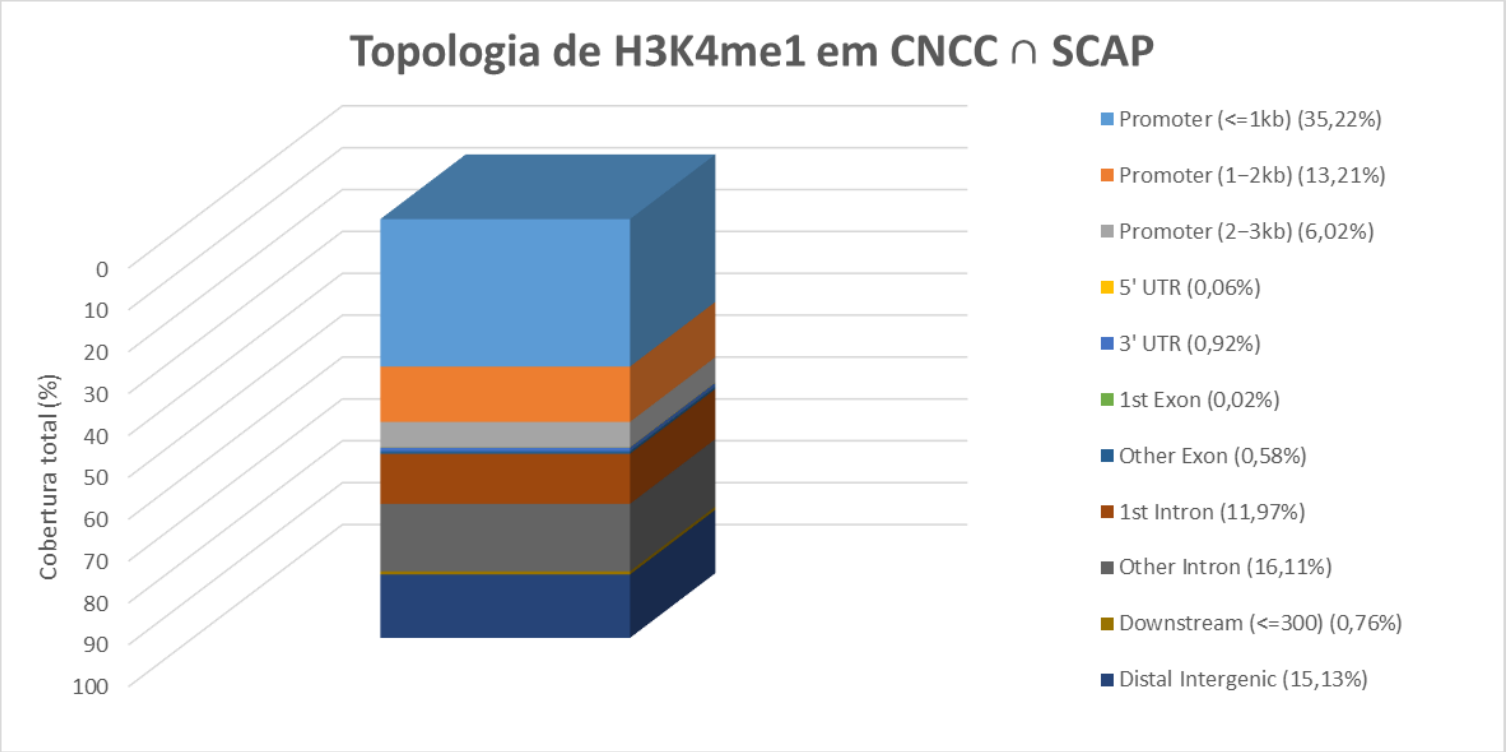


Figura 5. Topologia da marca H3K4me1 ao longo dos genomas de CNCC e SCAP, filtrando apenas para regiões que se sobrepõem nos dois tipos celulares, com porcentagens obtidas através da ferramenta *ChIP Seeker*.

5.1.3 Marca de histona H3K4me3

A topologia encontrada para a marcação de H3K4me3 em CNCC demonstrou que 91,48% dessas marcas se encontraram em regiões promotoras, 2,83% para regiões intrônicas, 5,06% para regiões distais intergênicas e menos de 1% para regiões exônicas e 5’ e 3’ não traduzidas (Figura 6A). Tendo em vista as SCAP, a marca de H3K4me3 teve uma presença de 24,24% em regiões promotoras, 24,43% em regiões intrônicas, 49,67% em regiões distais intergênicas e aproximadamente 1% para regiões exônicas e regiões 5’ e 3’ não traduzidas (Figura 6B).

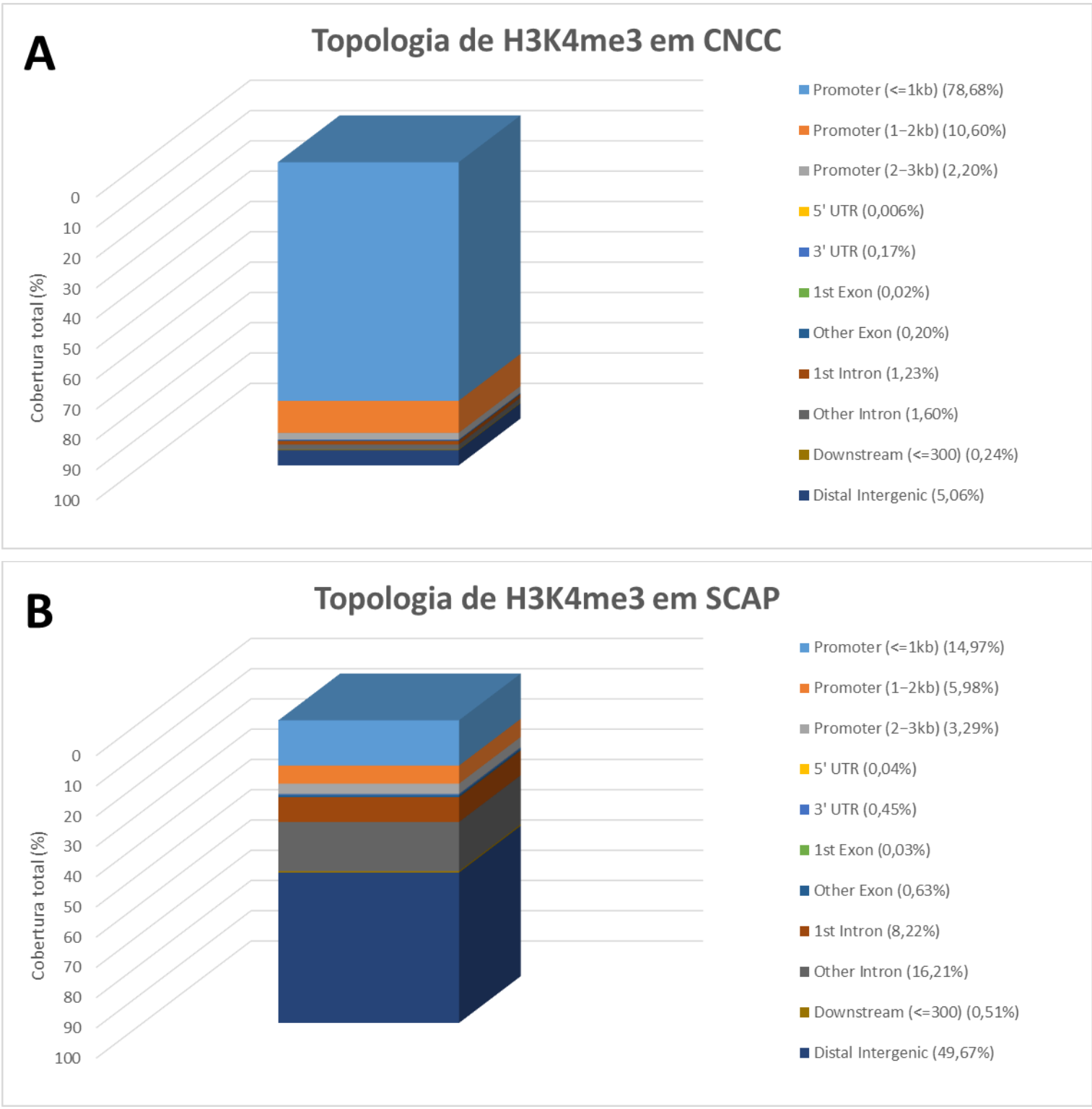


Figura 6. Topologia da marca H3K4me3 ao longo dos genomas de **A)** CNCC e **B)** SCAP, com porcentagens obtidas através da ferramenta *ChIP Seeker*.

Ao se sobrepôr as marcas de H3K4me3 das CNCCs nas SCAPs, pode-se observar que a grande maioria dessas marcas se encontram em regiões promotoras, com 93,94% das marcas totais. Quanto às regiões intrônicas, são sobrepostas

somente 1,69%, e para regiões intergênicas 3,84% das marcas totais. O restante para exons e regiões 5’ e 3’ não traduzidas (Figura 7).

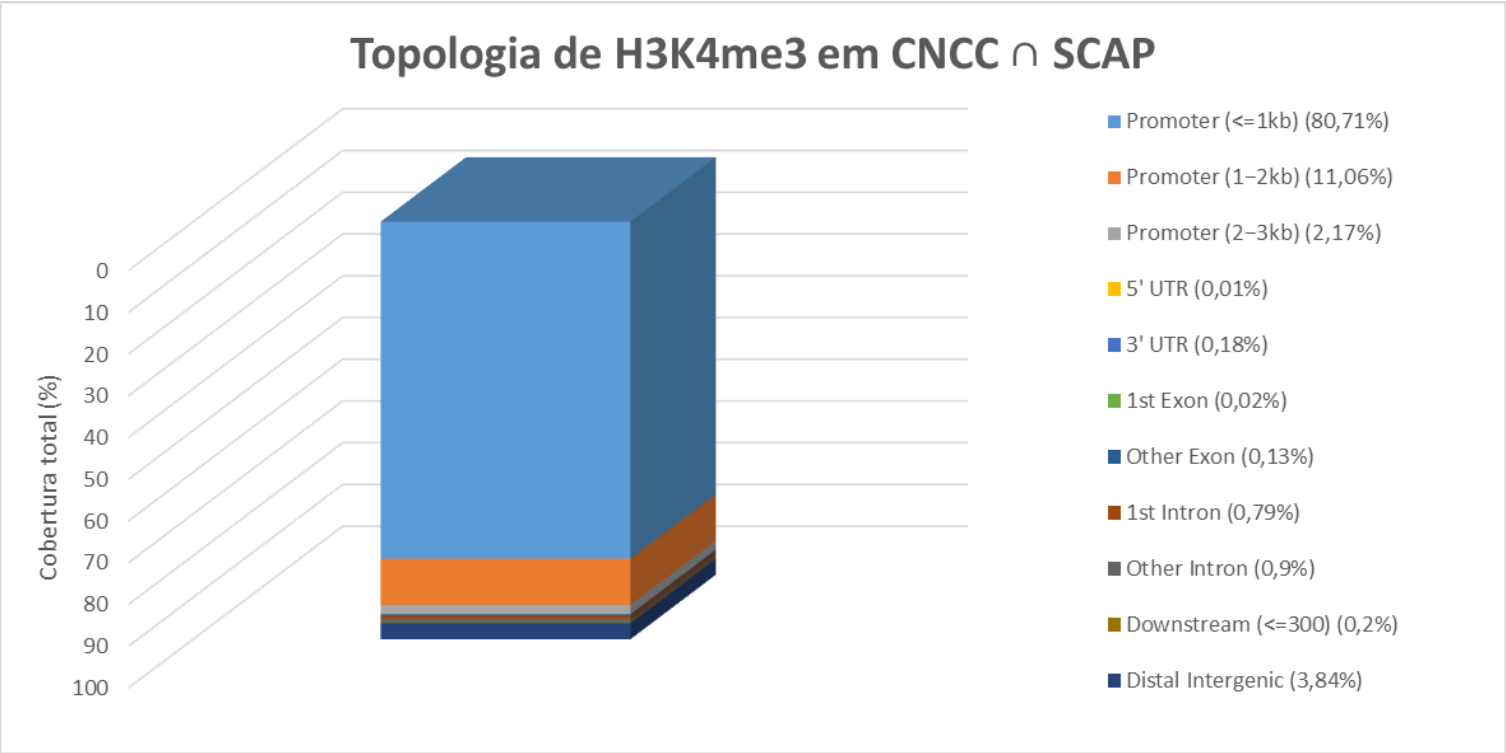


Figura 7. Topologia da marca H3K4me3 ao longo dos genomas de CNCC e SCAP, filtrando apenas para regiões que se sobrepõem nos dois tipos celulares, com porcentagens obtidas através da ferramenta *ChIP Seeker*.

5.1.4 Cofator de Transcrição p300

Em relação ao cofator de transcrição p300, a ferramenta ChIPSeeker identificou que cerca de 52,52% do sinal deste cofator foi encontrado em regiões promotoras em CNCC, 22,38% em regiões intrônicas, 23,28% em regiões distais intergênicas e 1,09% em regiões exônicas e 5’ e 3’ não traduzidas (Figura 8A). Nas SCAP, esse cofator se encontrava 18,6% em regiões promotoras, 36,73% em regiões intrônicas, 41% em regiões distais intergênicas e 2,77% em regiões exônicas, e 5’ e 3’ não traduzidas (Figura 8B)

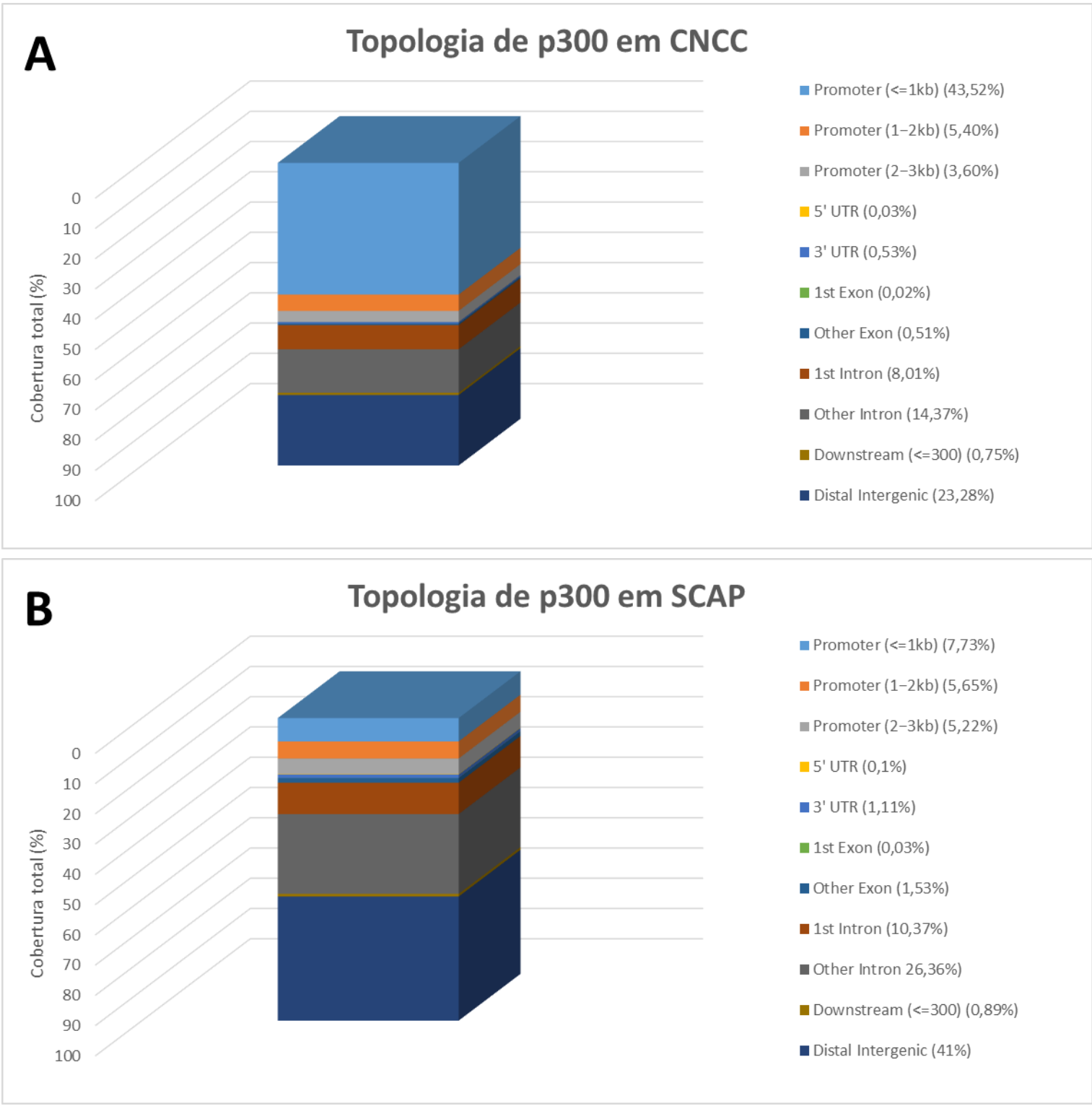


Figura 8. Topologia do cofator de transcrição p300 ao longo dos genomas de **A)** CNCC e **B)** SCAP, com porcentagens obtidas através da ferramenta *ChIP Seeker*.

Quando as marcas do cofator de transcrição p300 nas CNCCs e SCAPs são sobrepostas, há um enriquecimento de aproximadamente 23% em regiões

promotoras, 18,72% para regiões intrônicas, 56,6% para regiões distais intergênicas, e o restante para exons e regiões 5’ e 3’ não traduzidas (Figura 9).

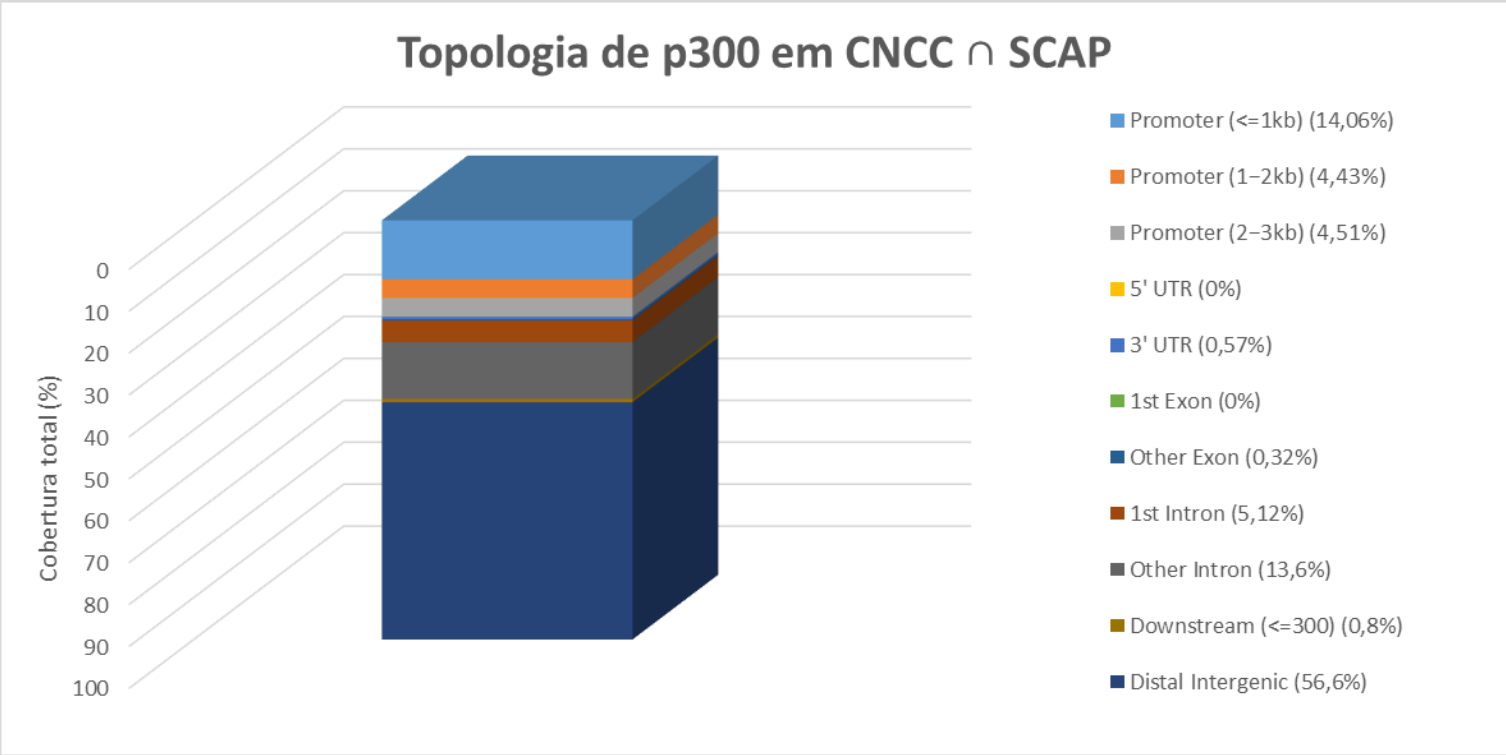


Figura 9. Topologia do cofator de transcrição p300 ao longo dos genomas de CNCC e SCAP, filtrando apenas para regiões que se sobrepõem nos dois tipos celulares, com porcentagens obtidas através da ferramenta *ChIP Seeker*.

5.2 Comparação direta entre regiões selecionadas das CNCCs e SCAPs

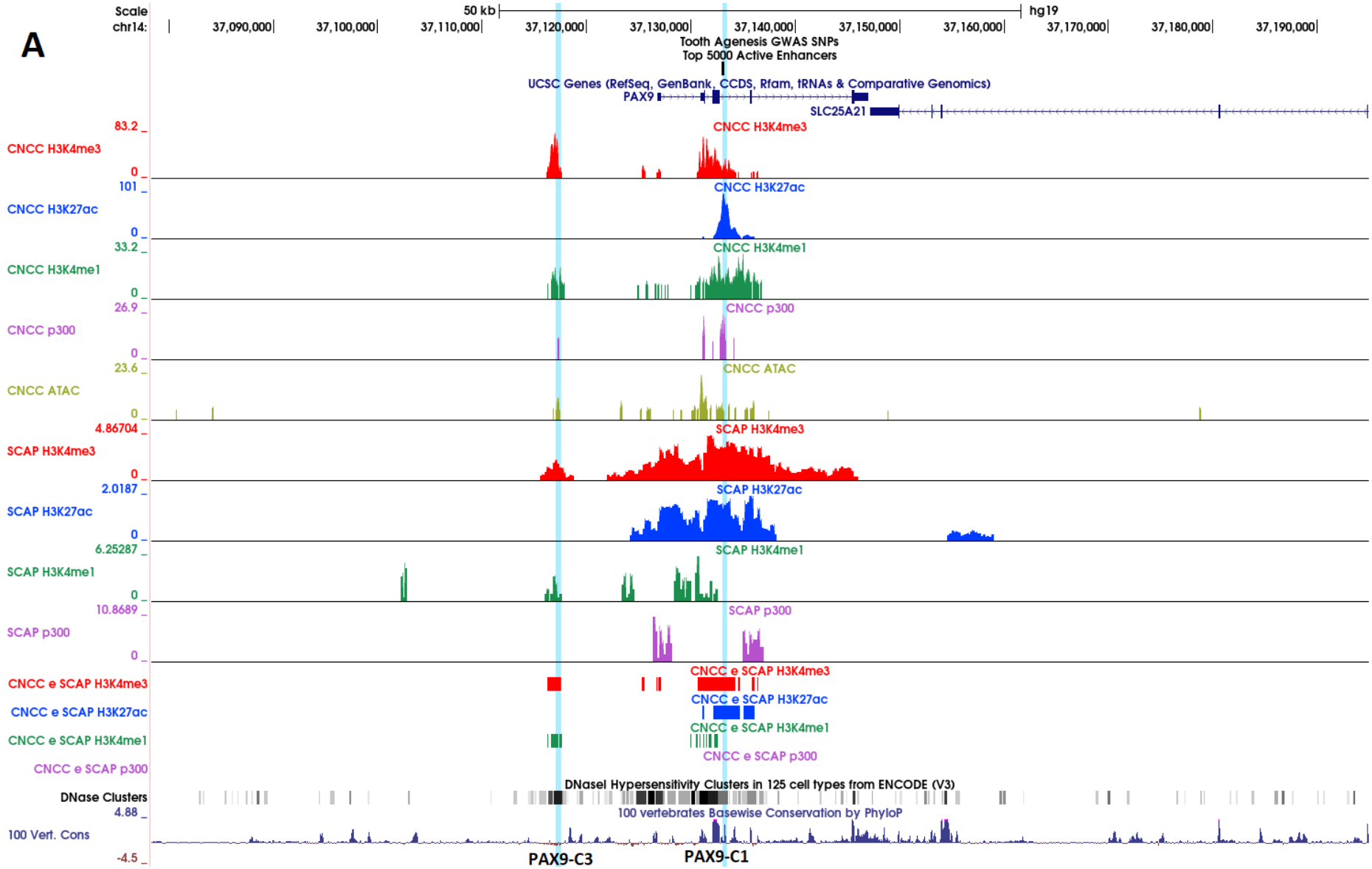
A partir da lista de genes relacionados ao desenvolvimento dentário obtidas previamente, foram selecionadas algumas regiões exemplo com potencial cis-regulatório em células da crista neural ou em células da papila apical, além de regiões que contém potencial cis-regulatório em ambas as células. A tabela suplementar S1 apresenta os dados caracterizados de acordo com a cobertura e enriquecimento de cada marca para as regiões investigadas a seguir, além de todas as outras regiões candidatas não exemplificadas.

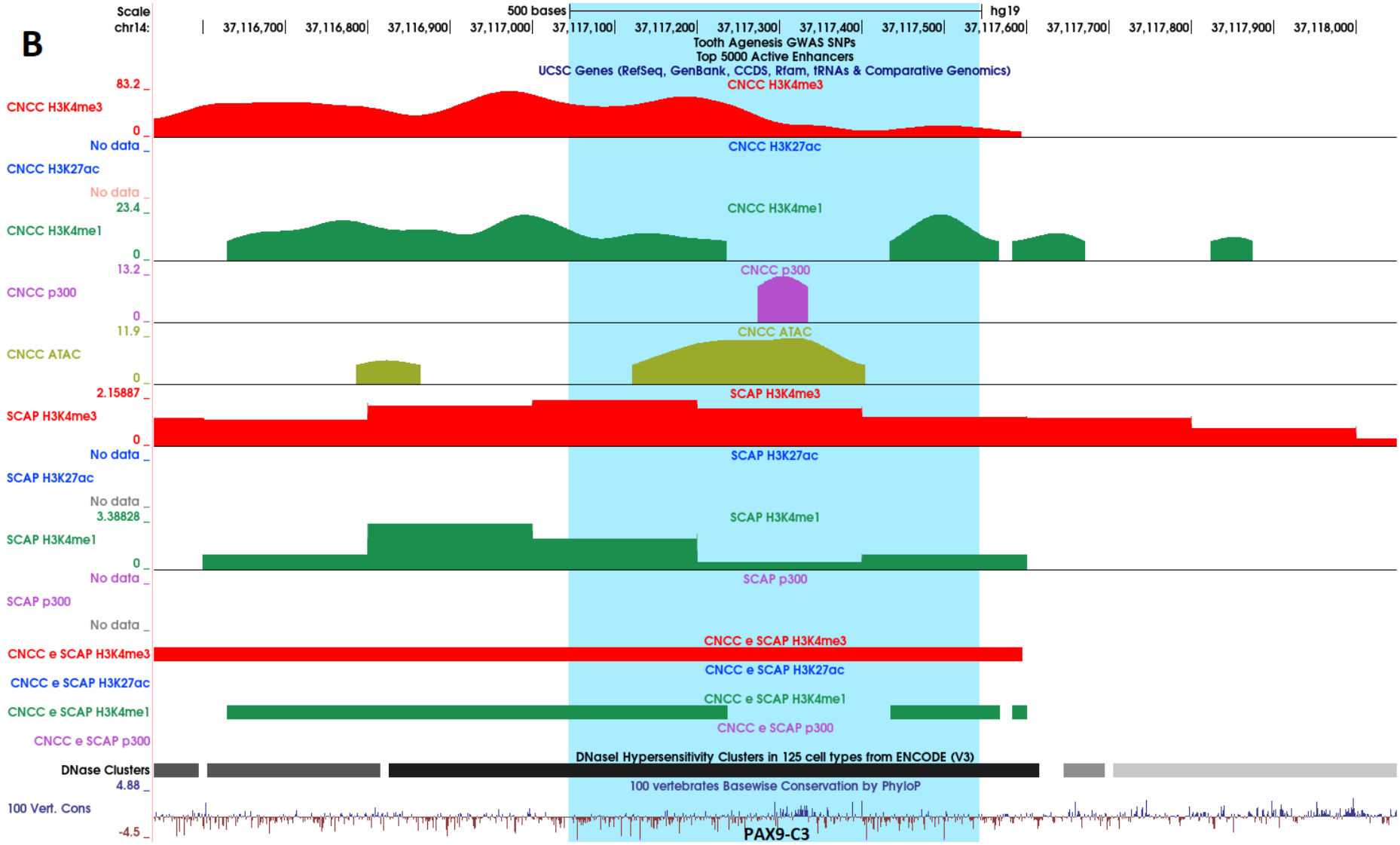
5.2.1 Candidatos a *enhancer* de PAX9

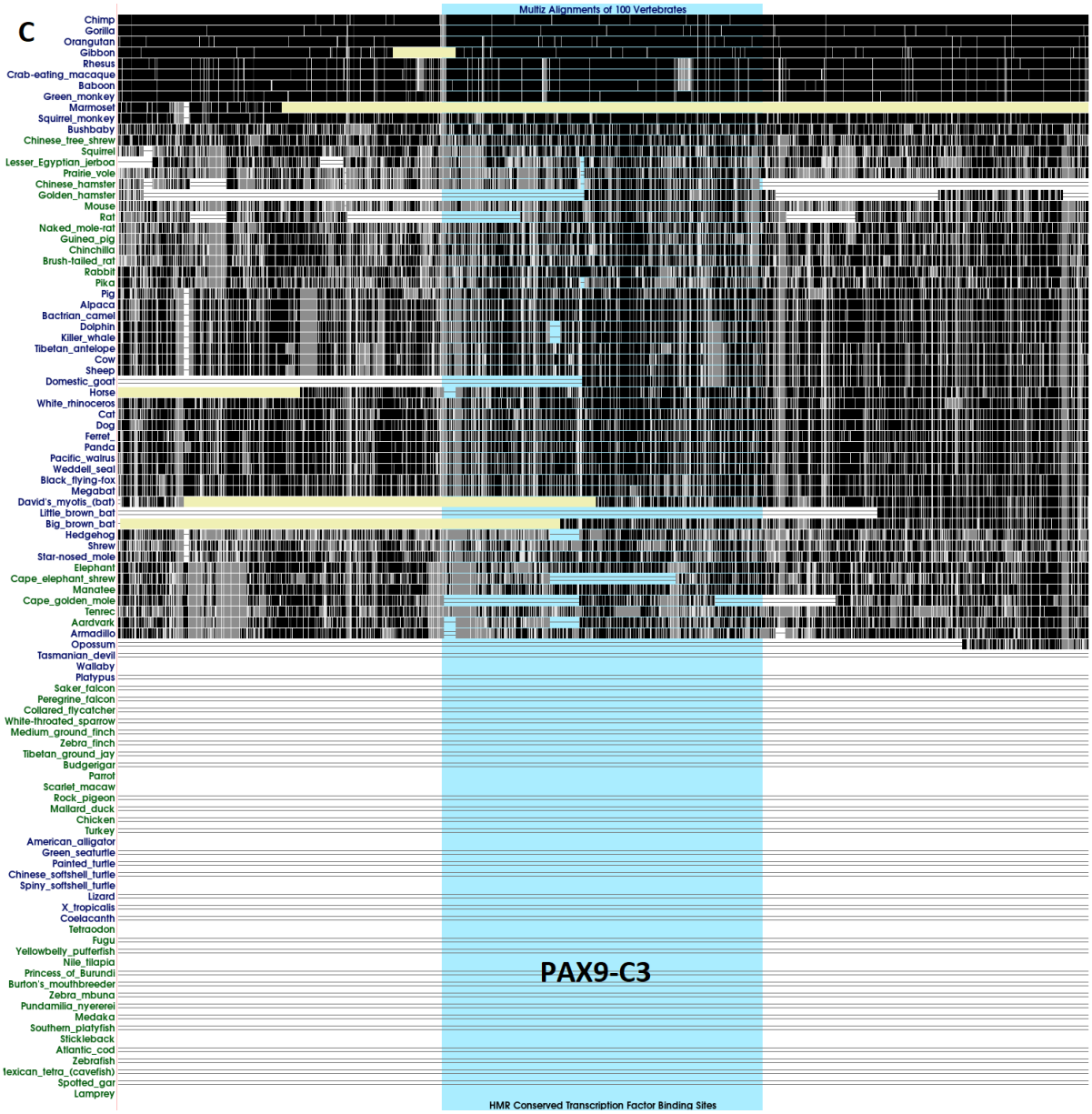
Uma região de aproximadamente 500 bp com uma distância de aproximadamente 10 kbp (10000 pares de base) upstream do promotor do gene PAX9 (PAX9-C3 - Candidato 3 de *enhancer* do gene PAX9) demonstra enriquecimento nas marcações de H3K4me3, H3K4me1 e ATAC-seq em CNCCs e enriquecimento similar em H3K4me3 e H3K4me1 em células SCAP (Figuras 10A e

10B), com ausência de H3K27ac nos dois tipos celulares. A região demonstra grande sensibilidade à DNAseI, moderada conservação entre espécies de vertebrados, porém nenhuma sequência reconhecida como fator de transcrição em humanos, camundongos e ratos (*HMR Conserved Transcription Factor Binding Sites)* (Figura 10C).

Também para o PAX9, outra região com cerca de 500 bp encontrada no intron 3 do próprio PAX9 (PAX9-C1) contém alto enriquecimento das marcas de H3K27Ac, H3K4me3, H3K4me1, p300 e ATAC-seq nas CNCCs (Figura 10A e 10D), além de ter sido denominada como uma região de *enhancer* altamente ativa em CNCCs por Prescott et al. (2015) (de acordo com o enriquecimento visto da marca H3K27ac e comparado entre humanos e chimpanzés). Essa mesma região para as células SCAPs possui grande enriquecimento para as marcas de H3K27ac e H3K4me3, que demarcam possível presença de um *enhancer* também nesse tipo celular. Essa região também possui alta conservação entre diferentes espécies de vertebrados, além de conter motifs para alguns fatores de transcrição comuns em humanos, ratos e camundongos, como PAX6, HOXA3 e ELK1 (Figura 10E).


A
Scale
chr14:
50 kb
hg19
37,090,000
37,100,000
37,110,000
37,120,000
37,130,000
37,140,000
37,150,000
37,160,000
37,170,000
37,180,000
37,190,000
Tooth Agenesis GWAS SNPs
Top 5000 Active Enhancers
UCSC Genes (RefSeq, GenBank, CCDS, Rfam, tRNAs & Comparative Genomics)
PAX9
SLC25A21
CNCC H3K4me3
83.2
0
CNCC H3K27ac
101
0
CNCC H3K4me1
33.2
0
CNCC p300
26.9
0
CNCC ATAC
23.6
0
SCAP H3K4me3
4.86704
0
SCAP H3K27ac
2.0187
0
SCAP H3K4me1
6.25287
0
SCAP p300
10.8689
0
CNCC e SCAP H3K4me3
CNCC e SCAP H3K27ac
CNCC e SCAP H3K4me1
CNCC e SCAP p300
DNase Clusters
DNaseI Hypersensitivity Clusters in 125 cell types from ENCODE (V3)
4.88
100 Vert. Cons
100 vertebrates Basewise Conservation by PhyloP
-4.5
PAX9-C3
PAX9-C1


B
Scale
chr14:
500 bases
hg19
37,116,700
37,116,800
37,116,900
37,117,000
37,117,100
37,117,200
37,117,300
37,117,400
37,117,500
37,117,600
37,117,700
37,117,800
37,117,900
37,118,000
Tooth Agenesis GWAS SNPs
Top 5000 Active Enhancers
UCSC Genes (RefSeq, GenBank, CCDS, Rfam, tRNAs & Comparative Genomics)
CNCC H3K4me3
83.2
CNCC H3K4me3
0
No data
CNCC H3K27ac
CNCC H3K27ac
No data
CNCC H3K4me1
23.4
CNCC H3K4me1
0
CNCC p300
13.2
CNCC p300
0
CNCC ATAC
11.9
CNCC ATAC
0
SCAP H3K4me3
2.15887
SCAP H3K4me3
0
No data
SCAP H3K27ac
SCAP H3K27ac
No data
SCAP H3K4me1
3.38828
SCAP H3K4me1
0
No data
SCAP p300
SCAP p300
No data
CNCC e SCAP H3K4me3
CNCC e SCAP H3K4me3
CNCC e SCAP H3K27ac
CNCC e SCAP H3K27ac
CNCC e SCAP H3K4me1
CNCC e SCAP H3K4me1
CNCC e SCAP p300
CNCC e SCAP p300
DNase Clusters
DNaseI Hypersensitivity Clusters in 125 cell types from ENCODE (V3)
4.88
100 Vert. Cons
100 vertebrates Basewise Conservation by PhyloP
-4.5
PAX9-C3


C
Multiz Alignments of 100 Vertebrates
Chimp
Gorilla
Orangutan
Gibbon
Rhesus
Crab-eating_macaque
Baboon
Green_monkey
Marmoset
Squirrel_monkey
Bushbaby
Chinese_tree_shrew
Squirrel
Lesser_Egyptian_jerboa
Prairie_vole
Chinese_hamster
Golden_hamster
Mouse
Rat
Naked_mole-rat
Guinea_pig
Chinchilla
Brush-tailed_rat
Rabbit
Pika
Pig
Alpaca
Bactrian_camel
Dolphin
Killer_whale
Tibetan_antelope
Cow
Sheep
Domestic_goat
Horse
White_rhinoceros
Cat
Dog
Ferret
Panda
Pacific_walrus
Weddell_seal
Black_flying-fox
Megabat
David's_myotis_(bat)
Little_brown_bat
Big_brown_bat
Hedgehog
Shrew
Star-nosed_mole
Elephant
Cape_elephant_shrew
Manatee
Cape_golden_mole
Tenrec
Aardvark
Armadillo
Opossum
Tasmanian_devil
Wallaby
Platypus
Saker_falcon
Peregrine_falcon
Collared_flycatcher
White-throated_sparrow
Medium_ground_finch
Zebra_finch
Tibetan_ground_jay
Budgerigar
Parrot
Scarlet_macaw
Rock_pigeon
Mallard_duck
Chicken
Turkey
American_alligator
Green_seaturtle
Painted_turtle
Chinese_softshell_turtle
Spiny_softshell_turtle
Lizard
X_tropicalis
Coelacanth
Tetraodon
Fugu
Yellowbelly_pufferfish
Nile_tilapia
Princess_of_Burundi
Burton's_mouthbreeder
Zebra_mbuna
Pundamilia_nyererei
Medaka
Southern_platyfish
Stickleback
Atlantic_cod
Zebrafish
Mexican_tetra_(cavefish)
Spotted_gar
Lamprey
PAX9-C3
HMR Conserved Transcription Factor Binding Sites

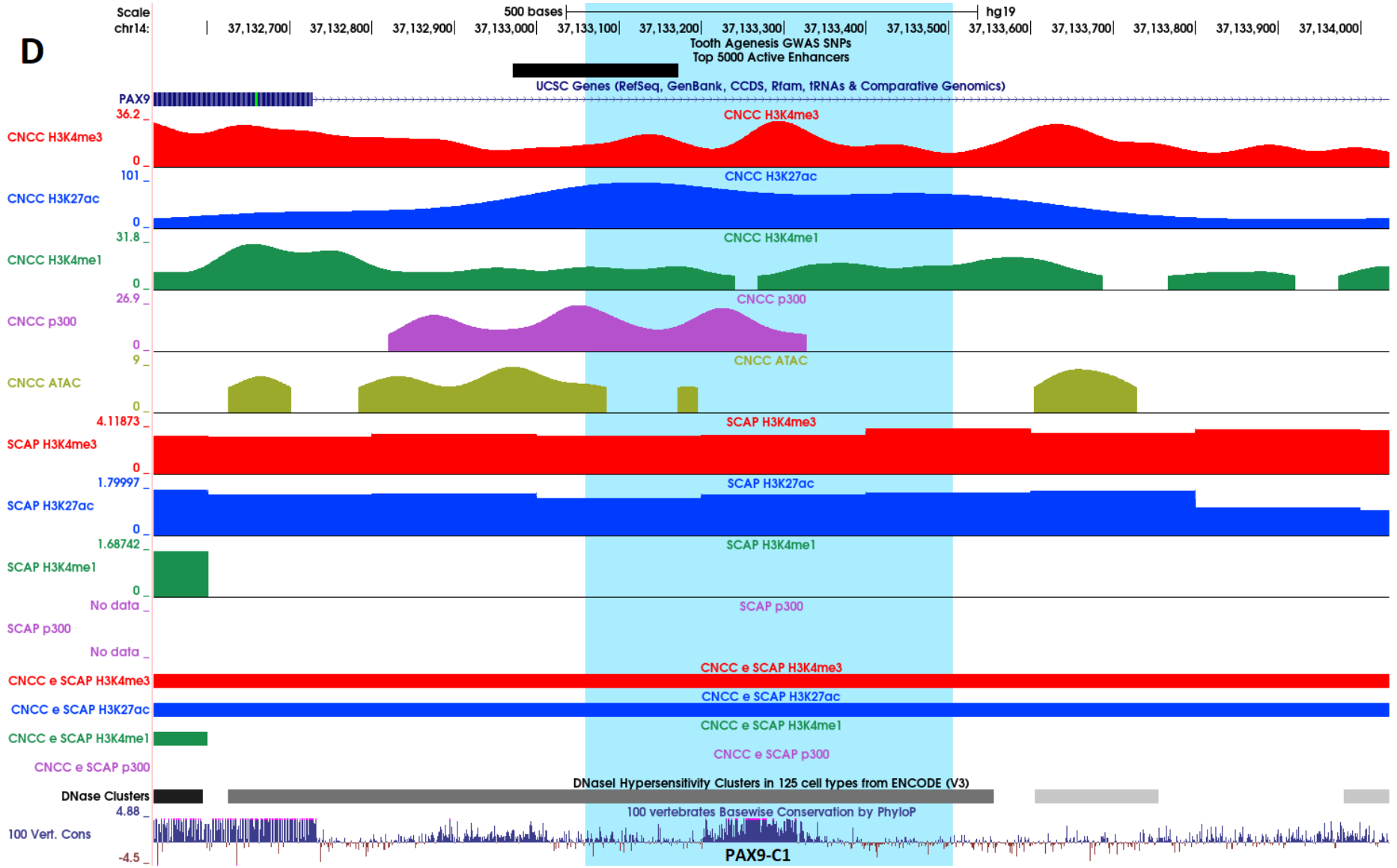
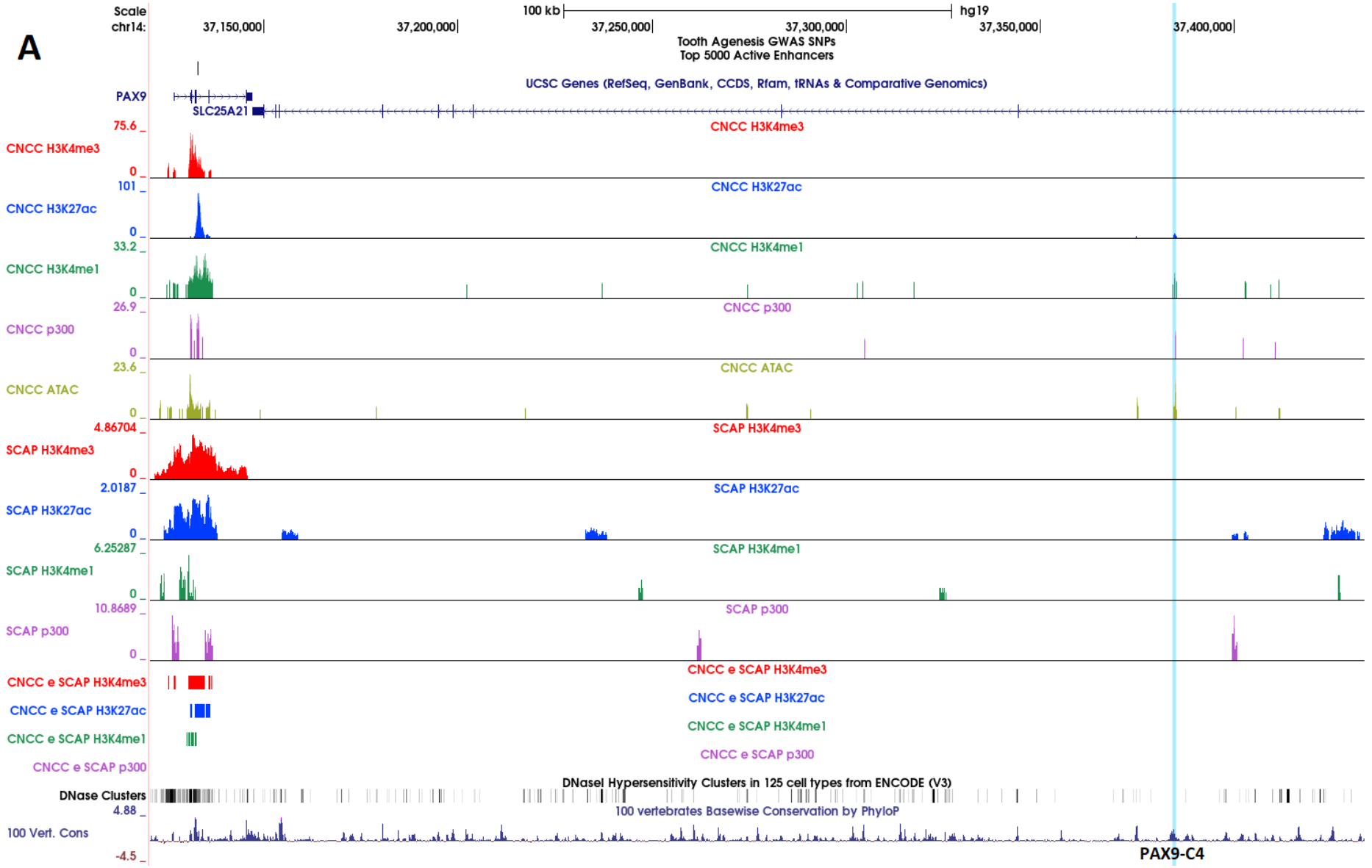

D
Scale
chr14:
500 bases
hg19
37,132,700
37,132,800
37,132,900
37,133,000
37,133,100
37,133,200
37,133,300
37,133,400
37,133,500
37,133,600
37,133,700
37,133,800
37,133,900
37,134,000
Tooth Agenesis GWAS SNPs
Top 5000 Active Enhancers
UCSC Genes (RefSeq, GenBank, CCDS, Rfam, tRNAs & Comparative Genomics)
PAX9
36.2
CNCC H3K4me3
0
101
CNCC H3K27ac
0
31.8
CNCC H3K4me1
0
26.9
CNCC p300
0
9
CNCC ATAC
0
4.11873
SCAP H3K4me3
0
1.79997
SCAP H3K27ac
0
1.68742
SCAP H3K4me1
0
No data
SCAP p300
No data
CNCC e SCAP H3K4me3
CNCC e SCAP H3K27ac
CNCC e SCAP H3K4me1
CNCC e SCAP p300
DNase Clusters
DNaseI Hypersensitivity Clusters in 125 cell types from ENCODE (V3)
4.88
100 Vert. Cons
-4.5
100 vertebrates Basewise Conservation by PhyloP
PAX9-C1

Figura 10. **A)** Mapa genômico da região ao redor do gene PAX9. As regiões grifadas com as linhas azul claro são as regiões candidatas mais próximas para *enhancer* do gene PAX9, PAX9-C3 (chr14:37,117,044-37,117,546) e PAX9-C1 (chr14:37,133,061-37,133,510) respectivamente. **B)** Mapa genômico com visualização ampliada para a região candidata a *enhancer* PAX9-C3. **C)** Conservação entre 100 espécies de vertebrados (*MultiZ Alignment*) e sítios de ligação de fatores de transcrição (*HMR Conserved Transcription Factor Binding Sites*) da região ao redor do candidato a *enhancer* PAX9-C3. **D)** Mapa genômico com visualização ampliada para a região candidata a *enhancer* PAX9-C1. **E)** Conservação entre 100 espécies de vertebrados e sítios de ligação de fatores de transcrição da região ao redor do candidato a *enhancer* PAX9-C1. As imagens foram obtidas através do *UCSC Genome Browser.*

Uma terceira região, candidata para *enhancer* do gene PAX9 (PAX9-C4), com cerca de 900 bp de tamanho, se situa no intron 1 do gene SLC25A21, aproximadamente 250 kbp de distância do promotor de PAX9, a qual contém moderado enriquecimento em CNCC das marcas de H3K27ac, H3K4me1 e p300 (Figura 11A e 11B), além de grande acessibilidade à cromatina em uma região de alta conservação entre vertebrados (Figura 11C). A mesma região não possui enriquecimento para nenhuma das marcas estudadas em SCAP (Figura 11B), demonstrando uma mudança na topologia epigenética em algum momento da diferenciação das CNCCs até SCAPs, onde a região perde seu potencial cis-regulatório. A região também apresenta sequências reconhecidas por fatores de transcrição como IRF7, FREAC7 (FOXL1) e HMX1, este último importante no desenvolvimento de estruturas craniofaciais (Munroe et al., 2009)


A
Scale
chr14:
100 kb
hg19
37,150,000
37,200,000
37,250,000
37,300,000
37,350,000
37,400,000
Tooth Agenesis GWAS SNPs
Top 5000 Active Enhancers
UCSC Genes (RefSeq, GenBank, CCDS, Rfam, tRNAs & Comparative Genomics)
PAX9
SLC25A21
CNCC H3K4me3
CNCC H3K27ac
CNCC H3K4me1
CNCC p300
CNCC ATAC
SCAP H3K4me3
SCAP H3K27ac
SCAP H3K4me1
SCAP p300
CNCC e SCAP H3K4me3
CNCC e SCAP H3K27ac
CNCC e SCAP H3K4me1
CNCC e SCAP p300
DNaseI Hypersensitivity Clusters in 125 cell types from ENCODE (V3)
DNase Clusters
100 vertebrates Basewise Conservation by PhyloP
100 Vert. Cons
PAX9-C4

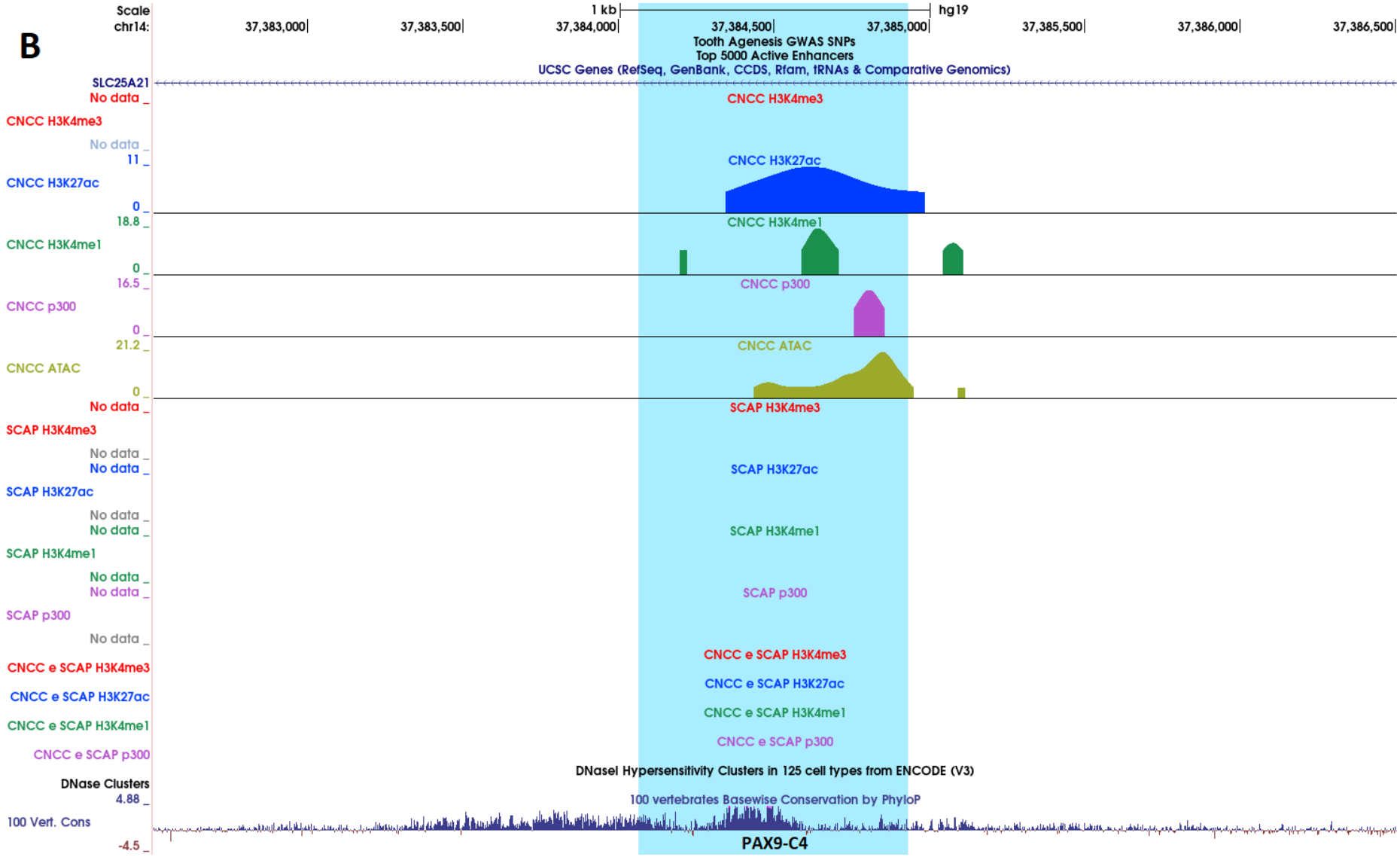
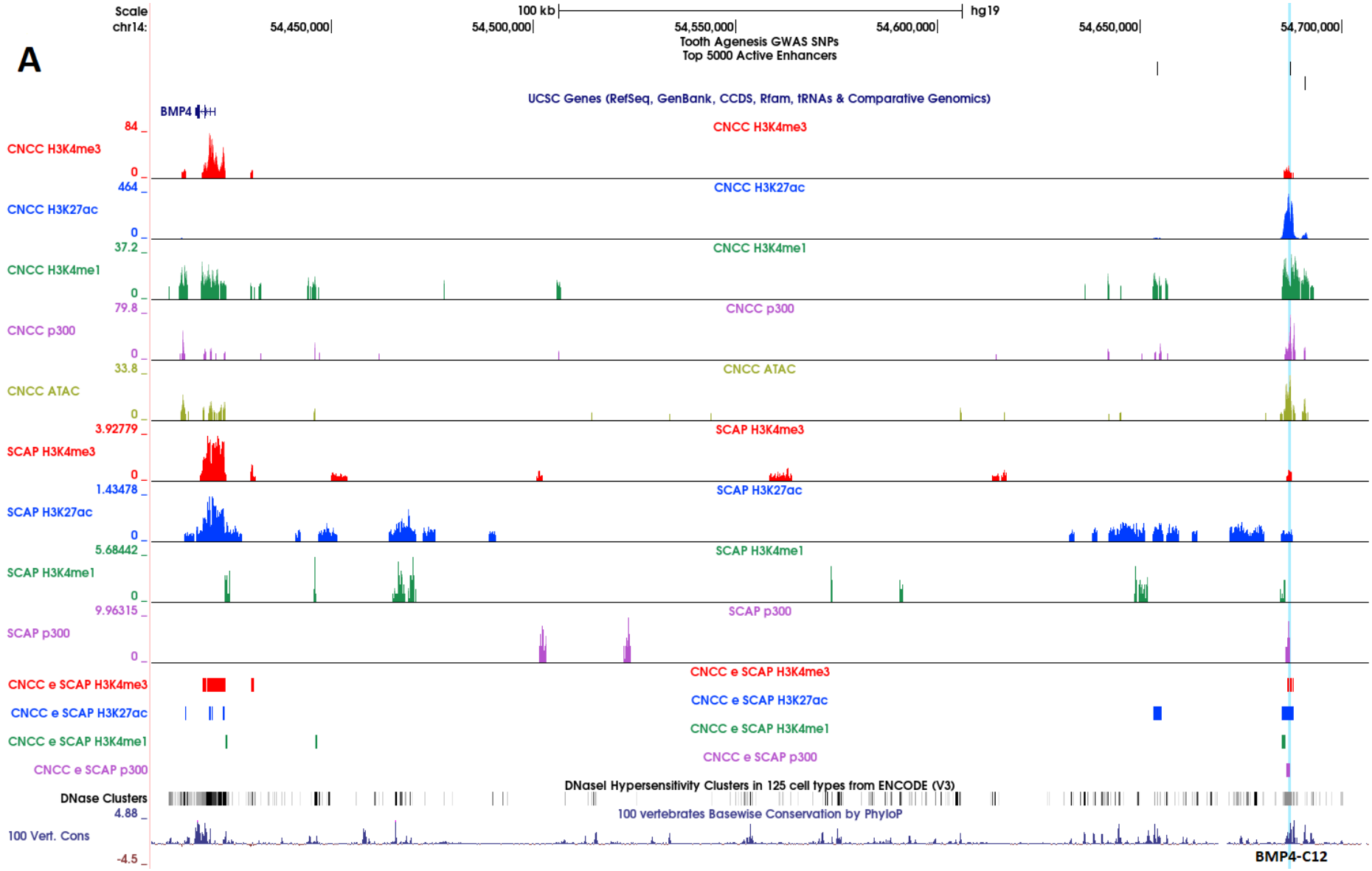

B
Scale
chr14:
1 kb
hg19
37,383,000
37,383,500
37,384,000
37,384,500
37,385,000
37,385,500
37,386,000
37,386,500
Tooth Agenesis GWAS SNPs
Top 5000 Active Enhancers
UCSC Genes (RefSeq, GenBank, CCDS, Rfam, tRNAs & Comparative Genomics)
SLC25A21
No data
CNCC H3K4me3
CNCC H3K4me3
No data
11
CNCC H3K27ac
CNCC H3K27ac
0
18.8
CNCC H3K4me1
CNCC H3K4me1
0
16.5
CNCC p300
CNCC p300
0
21.2
CNCC ATAC
CNCC ATAC
0
No data
SCAP H3K4me3
SCAP H3K4me3
No data
No data
SCAP H3K27ac
SCAP H3K27ac
No data
No data
SCAP H3K4me1
SCAP H3K4me1
No data
No data
SCAP p300
SCAP p300
No data
CNCC e SCAP H3K4me3
CNCC e SCAP H3K4me3
CNCC e SCAP H3K27ac
CNCC e SCAP H3K27ac
CNCC e SCAP H3K4me1
CNCC e SCAP H3K4me1
CNCC e SCAP p300
CNCC e SCAP p300
DNase Clusters
DNaseI Hypersensitivity Clusters in 125 cell types from ENCODE (V3)
4.88
100 Vert. Cons
100 vertebrates Basewise Conservation by PhyloP
-4.5
PAX9-C4

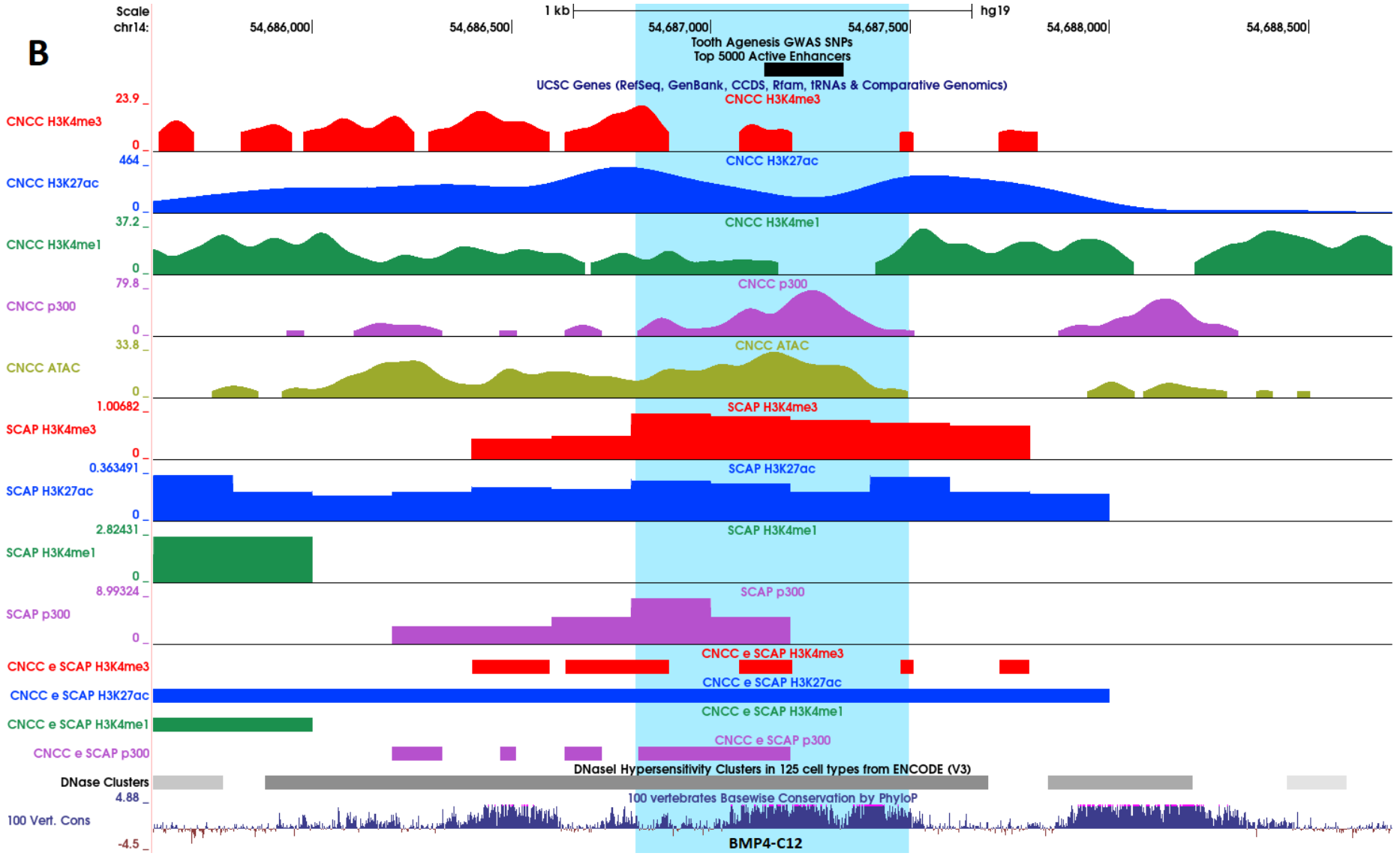
Figura 11. **A)** Mapa genômico contendo parte dos genes PAX9 e SLC25A21. A região candidata a enhancer de PAX9 grifada em azul claro, PAX9-C4 (chr14:37,384,062-37,384,950), se encontra dentro do intron 1 do gene SLC25A21. **B)** Mapa genômico com visualização ampliada para a região candidata a *enhancer* PAX9-C4. **C)** Conservação entre 100 espécies de vertebrados e sítios de ligação de fatores de transcrição da região ao redor do candidato a *enhancer* PAX9-C4. As imagens foram obtidas através do *UCSC Genome Browser.*

5.2.2 Candidato a *enhancer* de BMP4

No que diz respeito ao gene BMP4, uma região situada cerca de 265 kbp downstream do mesmo, com um tamanho de aproximadamente 2 kbp (BMP4-C12) possui um grande enriquecimento para todas as marcas investigadas nas células da crista neural, em especial nas marcas H3K27ac, H3K4me1 e p300, indicando grande atividade de *enhancer*, e nesta mesma região, no que diz respeito às células da

papila, também estão enriquecidas todas as marcas investigadas, exceto H3K4me1 (Figuras 12A e 12B). Essa região também possui alta abertura da cromatina segundo os dados de ATAC-seq nas CNCCs, alta conservação e hipersensibilidade à DNAseI e é um sítio de ligação para fatores de transcrição como GATA1, EVI1 e NKX2-2 (e NKX2-5) conservados entre humanos, ratos e camundongos (Figura 12C). Além disso, esta região também está denominada como um *enhancer* altamente ativo em CNCCs por Prescott et al. (2015).


A
Scale
chr14:
100 kb
hg19
54,450,000
54,500,000
54,550,000
54,600,000
54,650,000
54,700,000
Tooth Agenesis GWAS SNPs
Top 5000 Active Enhancers
UCSC Genes (RefSeq, GenBank, CCDS, Rfam, tRNAs & Comparative Genomics)
BMP4
CNCC H3K4me3
CNCC H3K27ac
CNCC H3K4me1
CNCC p300
CNCC ATAC
SCAP H3K4me3
SCAP H3K27ac
SCAP H3K4me1
SCAP p300
CNCC e SCAP H3K4me3
CNCC e SCAP H3K27ac
CNCC e SCAP H3K4me1
CNCC e SCAP p300
DNase Clusters
DNaseI Hypersensitivity Clusters in 125 cell types from ENCODE (V3)
100 Vert. Cons
100 vertebrates Basewise Conservation by PhyloP
BMP4-C12


B
Scale
chr14:
1 kb
hg19
54,686,000
54,686,500
54,687,000
54,687,500
54,688,000
54,688,500
Tooth Agenesis GWAS SNPs
Top 5000 Active Enhancers
UCSC Genes (RefSeq, GenBank, CCDS, Rfam, tRNAs & Comparative Genomics)
CNCC H3K4me3
23.9
0
CNCC H3K27ac
464
0
CNCC H3K4me1
37.2
0
CNCC p300
79.8
0
CNCC ATAC
33.8
0
SCAP H3K4me3
1.00682
0
SCAP H3K27ac
0.363491
0
SCAP H3K4me1
2.82431
0
SCAP p300
8.99324
0
CNCC e SCAP H3K4me3
CNCC e SCAP H3K27ac
CNCC e SCAP H3K4me1
CNCC e SCAP p300
DNase Clusters
DNaseI Hypersensitivity Clusters in 125 cell types from ENCODE (V3)
100 Vert. Cons
4.88
-4.5
100 vertebrates Basewise Conservation by PhyloP
BMP4-C12

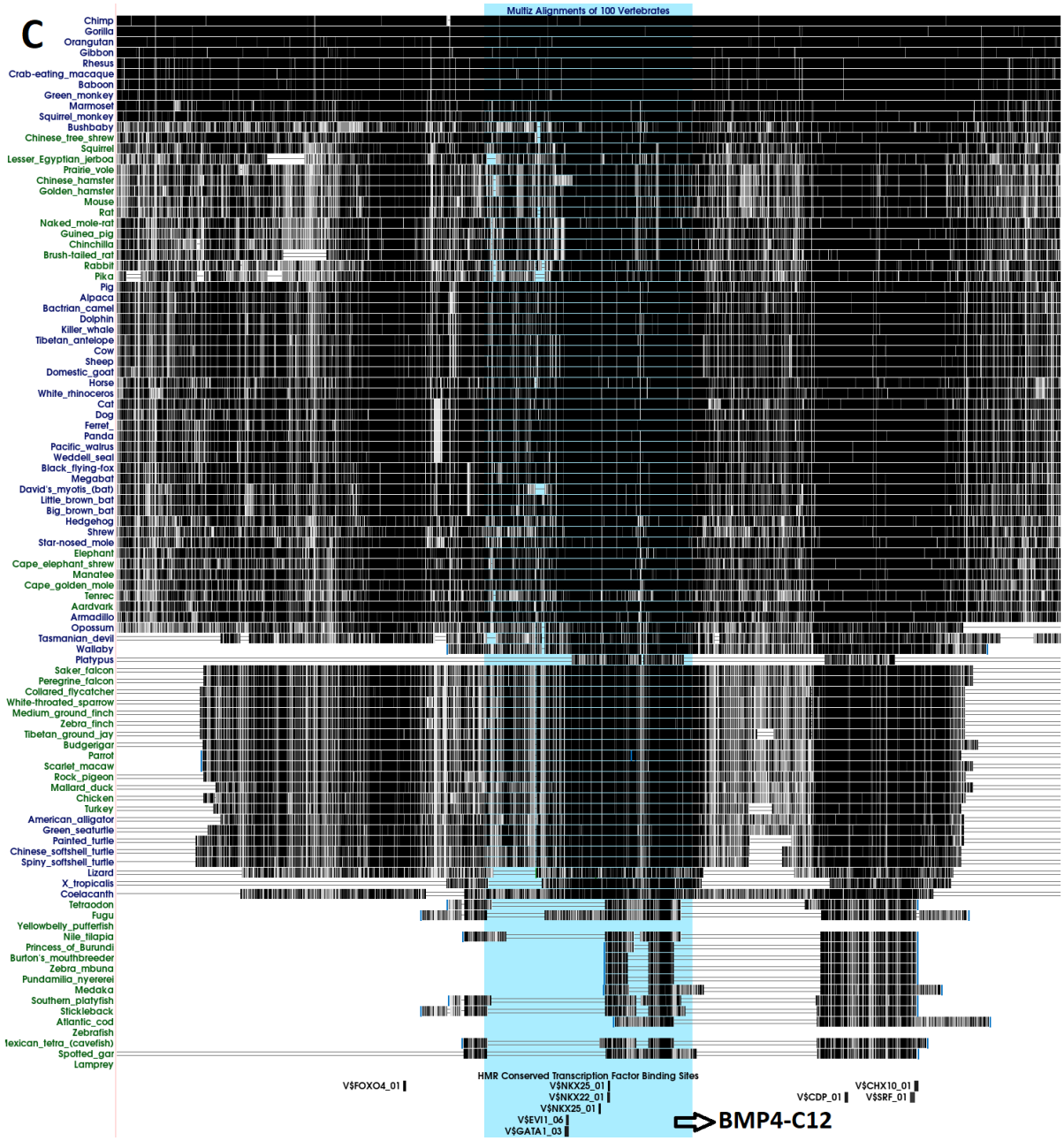
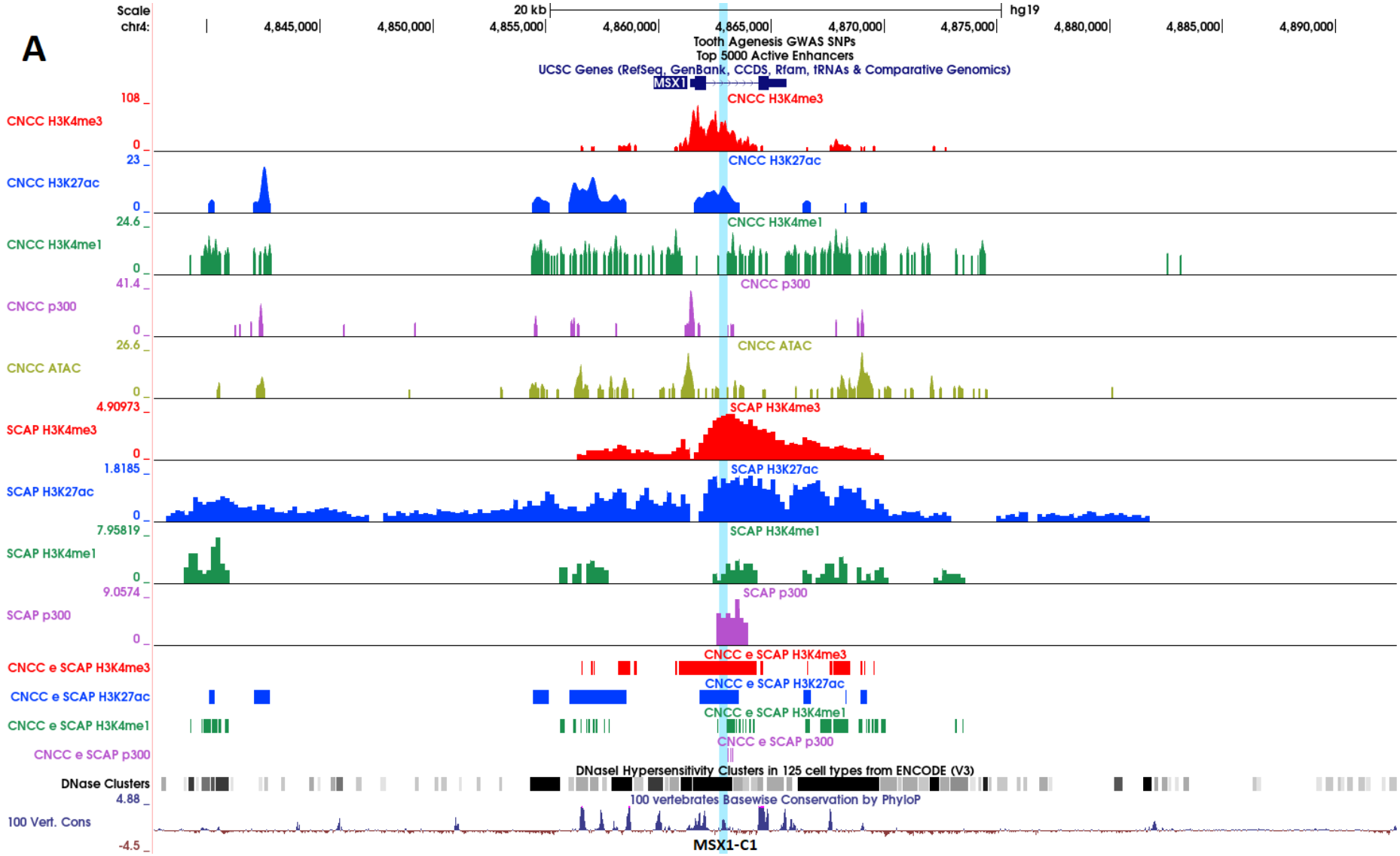
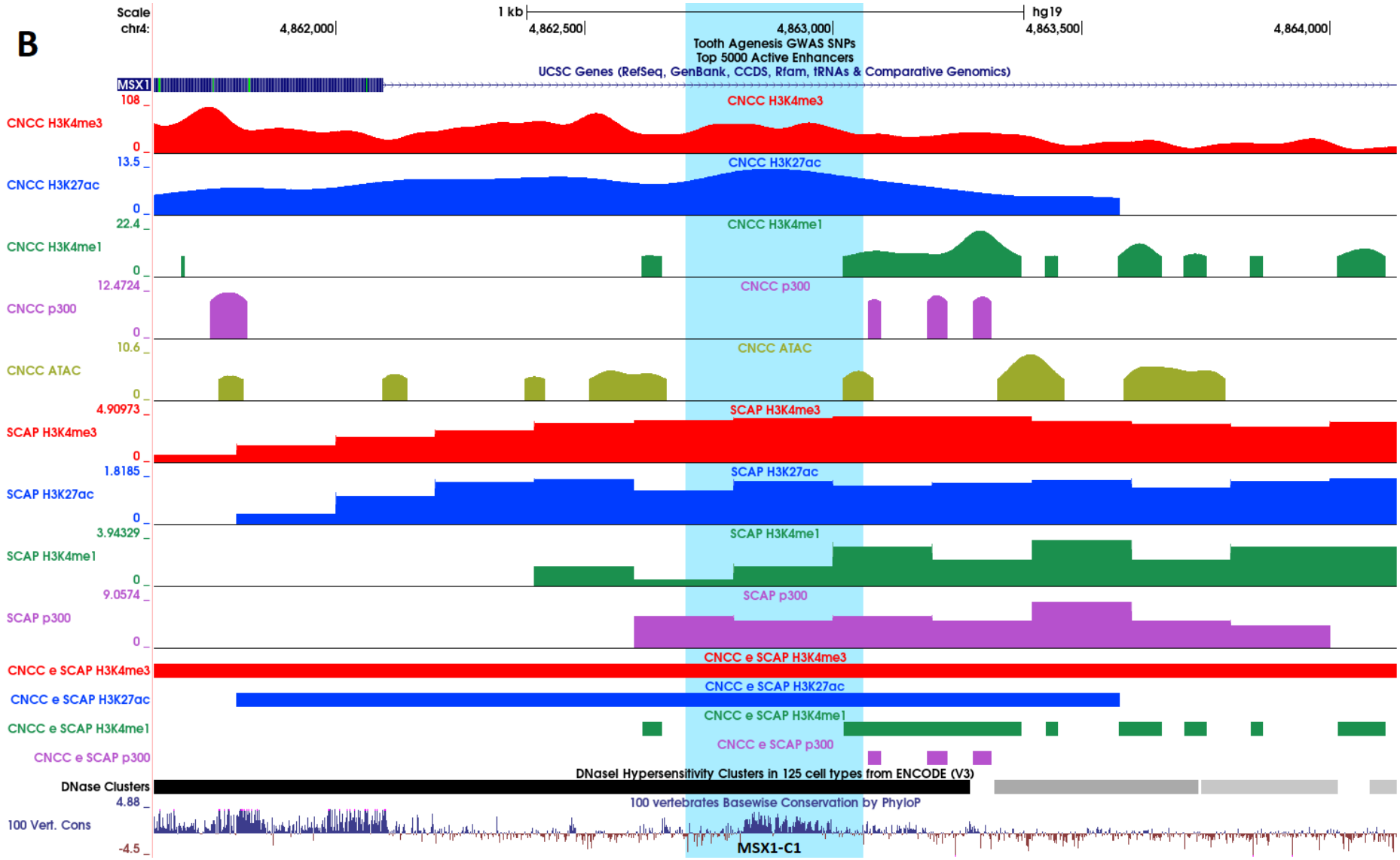


Figura 12. **A)** Mapa genômico contendo o gene BMP4 e uma região não codificante grifada em azul claro, a qual representa uma das regiões candidatas a *enhancer* do gene BMP4, a região BMP4-C12 (chr14:54,686,814-54,687,504). **B)** Mapa genômico com visualização ampliada para a região candidata a *enhancer* BMP4-C12. **C)** Conservação entre 100 espécies de vertebrados e sítios de ligação de fatores de transcrição da região ao redor do candidato a *enhancer* BMP4-C12. As imagens foram obtidas através do *UCSC Genome Browser.*

5.2.3 Candidato a *enhancer* de MSX1

Uma região dentro do intron 1 do gene MSX1 (MSX1-C1) é uma das candidatas a *enhancer* desse mesmo gene. Essa região tem aproximadamente 300 pares de base, com enriquecimento nos sinais de H3K27ac e H3K4me3 para CNCC (Figuras 13A e 13B). Nos sinais obtidos das células SCAPs, essa região demonstra enriquecimento para todas as marcas investigadas (Figuras 13A e 13B) e uma grande evidência de conservação entre espécies de vertebrados, uma boa hipersenbilidade à DNAseI e motifs para fatores de transcrição importantes como GATA1 e NFKB (Figura 13C).


A
Scale
chr4:
20 kb
hg19
4,845,000
4,850,000
4,855,000
4,860,000
4,865,000
4,870,000
4,875,000
4,880,000
4,885,000
4,890,000
Tooth Agenesis GWAS SNPs
Top 5000 Active Enhancers
UCSC Genes (RefSeq, GenBank, CCDS, Rfam, tRNAs & Comparative Genomics)
MSX1
CNCC H3K4me3
108
0
CNCC H3K27ac
23
0
CNCC H3K4me1
24.6
0
CNCC p300
41.4
0
CNCC ATAC
26.6
0
SCAP H3K4me3
4.90973
0
SCAP H3K27ac
1.8185
0
SCAP H3K4me1
7.95819
0
SCAP p300
9.0574
0
CNCC e SCAP H3K4me3
CNCC e SCAP H3K27ac
CNCC e SCAP H3K4me1
CNCC e SCAP p300
DNase Clusters
DNaseI Hypersensitivity Clusters in 125 cell types from ENCODE (V3)
4.88
100 Vert. Cons
100 vertebrates Basewise Conservation by PhyloP
-4.5
MSX1-C1


B
Scale
chr4:
1 kb
4,862,000
4,862,500
4,863,000
4,863,500
4,864,000
hg19
Tooth Agenesis GWAS SNPs
Top 5000 Active Enhancers
UCSC Genes (RefSeq, GenBank, CCDS, Rfam, tRNAs & Comparative Genomics)
MSX1
CNCC H3K4me3
CNCC H3K27ac
CNCC H3K4me1
CNCC p300
CNCC ATAC
SCAP H3K4me3
SCAP H3K27ac
SCAP H3K4me1
SCAP p300
CNCC e SCAP H3K4me3
CNCC e SCAP H3K27ac
CNCC e SCAP H3K4me1
CNCC e SCAP p300
DNaseI Hypersensitivity Clusters in 125 cell types from ENCODE (V3)
DNase Clusters
100 vertebrates Basewise Conservation by PhyloP
100 Vert. Cons
MSX1-C1

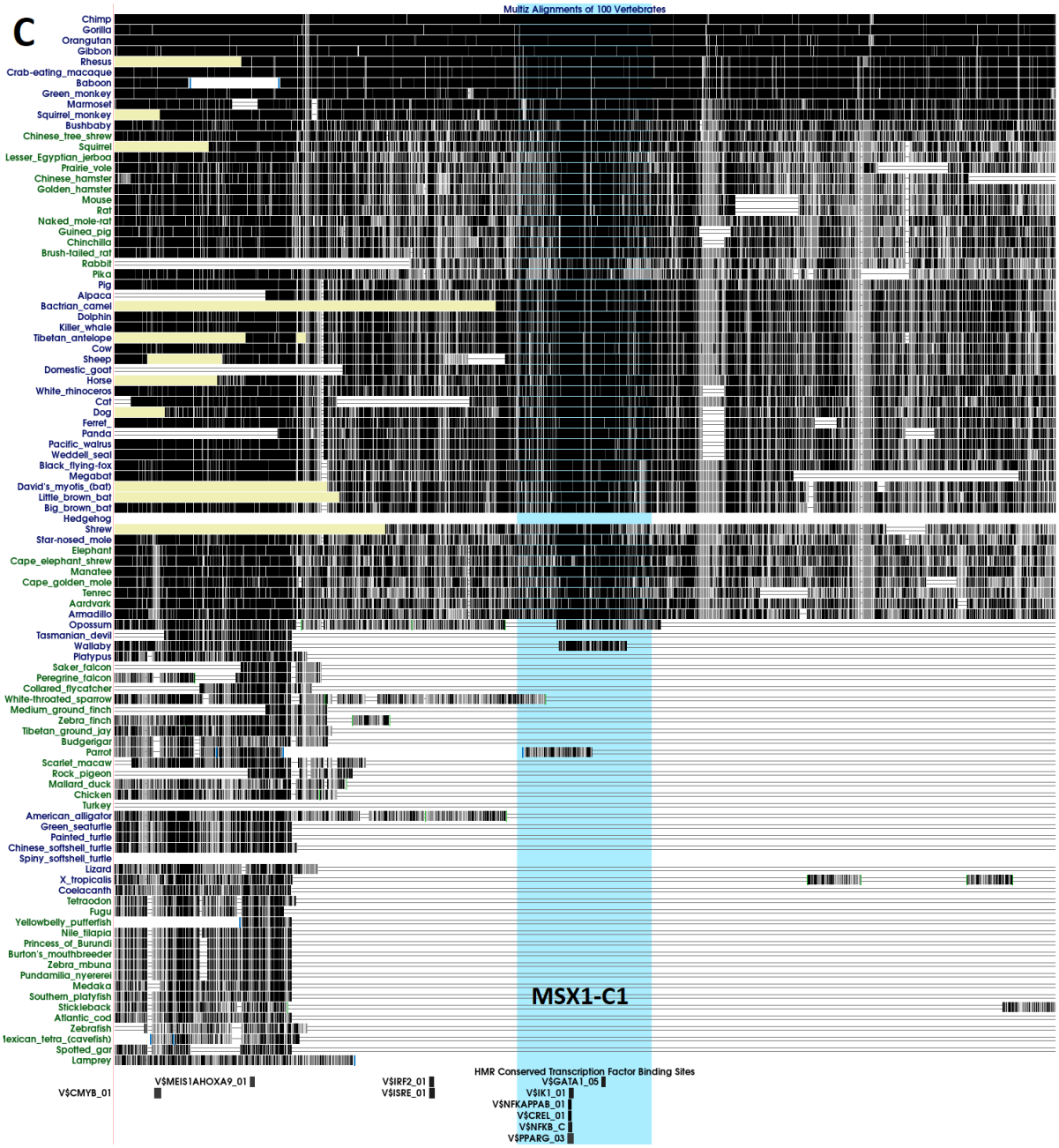
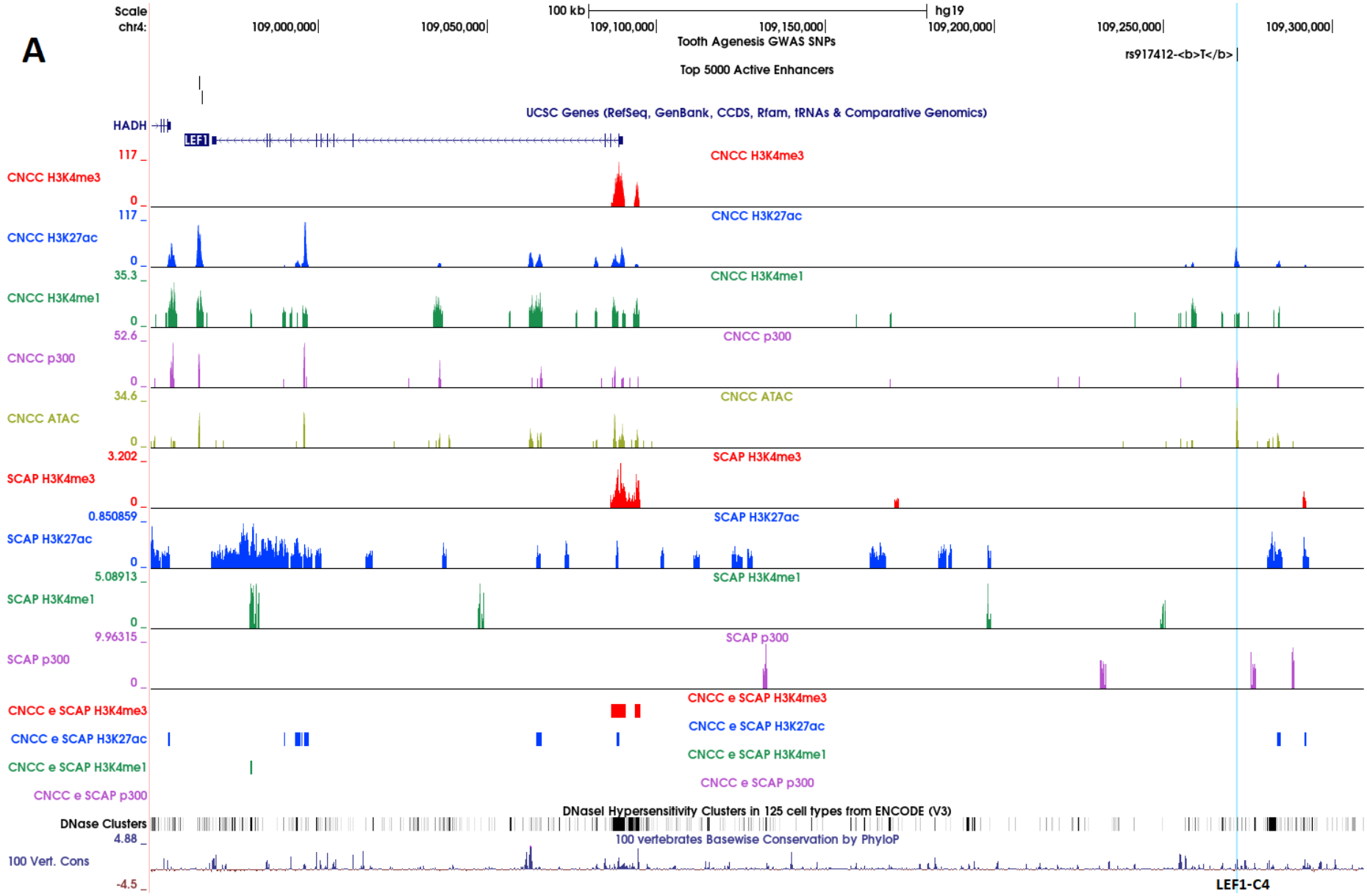
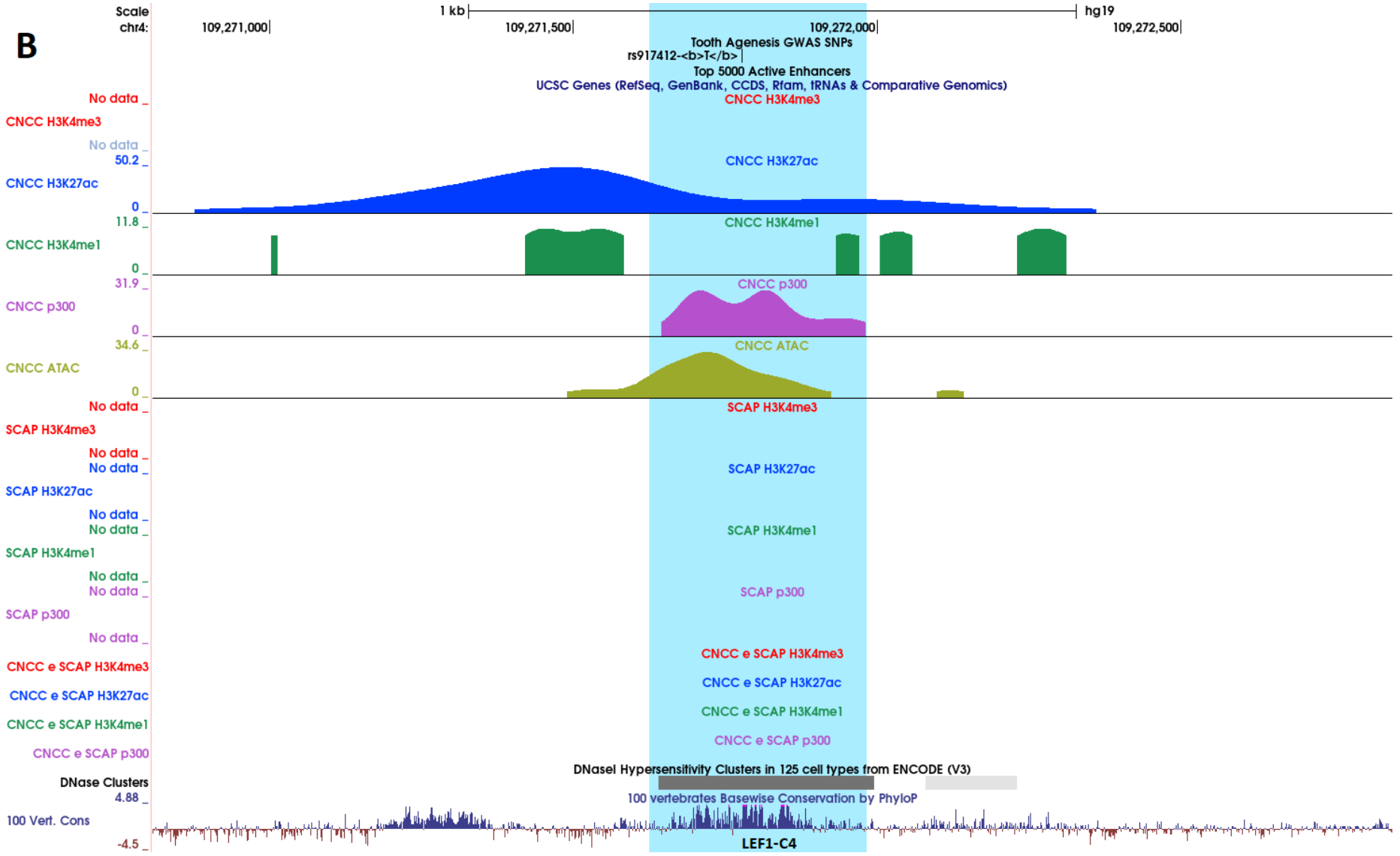


Figura 13. **A)** Mapa genômico contendo o gene MSX1. A região grifada com a linha azul claro representa uma das regiões candidatas a *enhancer* do gene MSX1, presente dentro do intron 1 do próprio gene MSX1, a região MSX1-C1 (chr4:4,862,698-4,863,072). **B)** Mapa genômico com visualização ampliada para a região candidata a *enhancer* MSX1-C1. **C)** Conservação entre 100 espécies de vertebrados e sítios de ligação de fatores de transcrição da região ao redor do candidato a *enhancer* MSX1-C1. As imagens foram obtidas através do *UCSC Genome Browser.*

5.2.4 Candidato a *enhancer* de LEF1

Em relação ao gene LEF1, uma região de aproximadamente 350 pares de base (LEF1-C4) contém um SNP (rs917412) que foi previamente identificado por Jonsson et al. (2018). Essa região se encontra a 182 kbp downstream do gene LEF1, e possui enriquecimentos para as marcas de H3K27ac, p300, H3K4me1 e ATAC-seq para as CNCCs, além de ter alta conservação entre espécies de vertebrados, uma alta hipersensibilidade à DNAseI em diferentes tipos celulares e apresentar motifs reconhecidos por diversos fatores de transcrição (Figura 14A, 14B e 14C). Curiosamente, essa mesma região não apresenta sinal algum para as SCAPs (Figura 14B), mesmo contendo esse SNP, que está relacionado à agenesia dentária, investigada em populações da Islândia, diminuindo a expressão do gene LEF1 como demonstrado por Jonsson et al. (2018). Essa região aparenta ser altamente essencial para a expressão do gene LEF1, justificando sua altíssima conservação entre espécies de vertebrados, o que é uma característica importante de uma região cis-regulatória.


A
Scale
chr4:
100 kb
hg19
109,000,000
109,050,000
109,100,000
109,150,000
109,200,000
109,250,000
109,300,000
Tooth Agenesis GWAS SNPs
rs917412-T
Top 5000 Active Enhancers
UCSC Genes (RefSeq, GenBank, CCDS, Rfam, tRNAs & Comparative Genomics)
HADH
LEF1
117
CNCC H3K4me3
0
CNCC H3K27ac
35.3
CNCC H3K4me1
52.6
CNCC p300
34.6
CNCC ATAC
3.202
SCAP H3K4me3
0.850859
SCAP H3K27ac
5.08913
SCAP H3K4me1
9.96315
SCAP p300
CNCC e SCAP H3K4me3
CNCC e SCAP H3K27ac
CNCC e SCAP H3K4me1
CNCC e SCAP p300
DNaseI Hypersensitivity Clusters in 125 cell types from ENCODE (V3)
DNase Clusters
4.88
100 vertebrates Basewise Conservation by PhyloP
100 Vert. Cons
-4.5
LEF1-C4


B
Scale
chr4:
109,271,000
1 kb
109,271,500
109,272,000
hg19
109,272,500
Tooth Agenesis GWAS SNPs
rs917412-T
Top 5000 Active Enhancers
UCSC Genes (RefSeq, GenBank, CCDS, Rfam, tRNAs & Comparative Genomics)
No data
CNCC H3K4me3
CNCC H3K4me3
No data
50.2
CNCC H3K27ac
CNCC H3K27ac
0
11.8
CNCC H3K4me1
CNCC H3K4me1
0
31.9
CNCC p300
CNCC p300
0
34.6
CNCC ATAC
CNCC ATAC
0
No data
SCAP H3K4me3
SCAP H3K4me3
No data
No data
SCAP H3K27ac
SCAP H3K27ac
No data
No data
SCAP H3K4me1
SCAP H3K4me1
No data
No data
SCAP p300
SCAP p300
No data
CNCC e SCAP H3K4me3
CNCC e SCAP H3K4me3
CNCC e SCAP H3K27ac
CNCC e SCAP H3K27ac
CNCC e SCAP H3K4me1
CNCC e SCAP H3K4me1
CNCC e SCAP p300
CNCC e SCAP p300
DNase Clusters
DNaseI Hypersensitivity Clusters in 125 cell types from ENCODE (V3)
4.88
100 Vert. Cons
100 vertebrates Basewise Conservation by PhyloP
-4.5
LEF1-C4

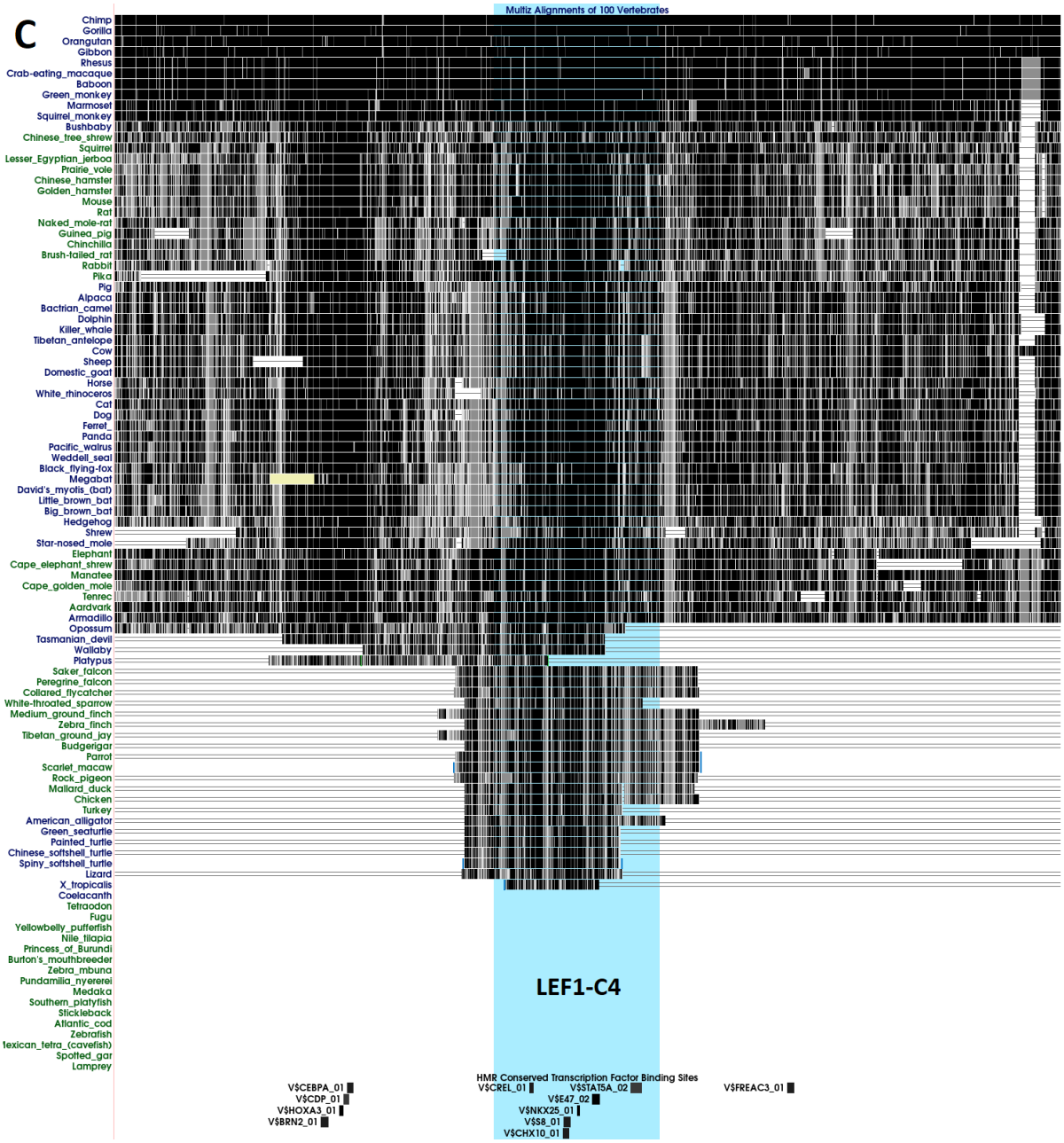
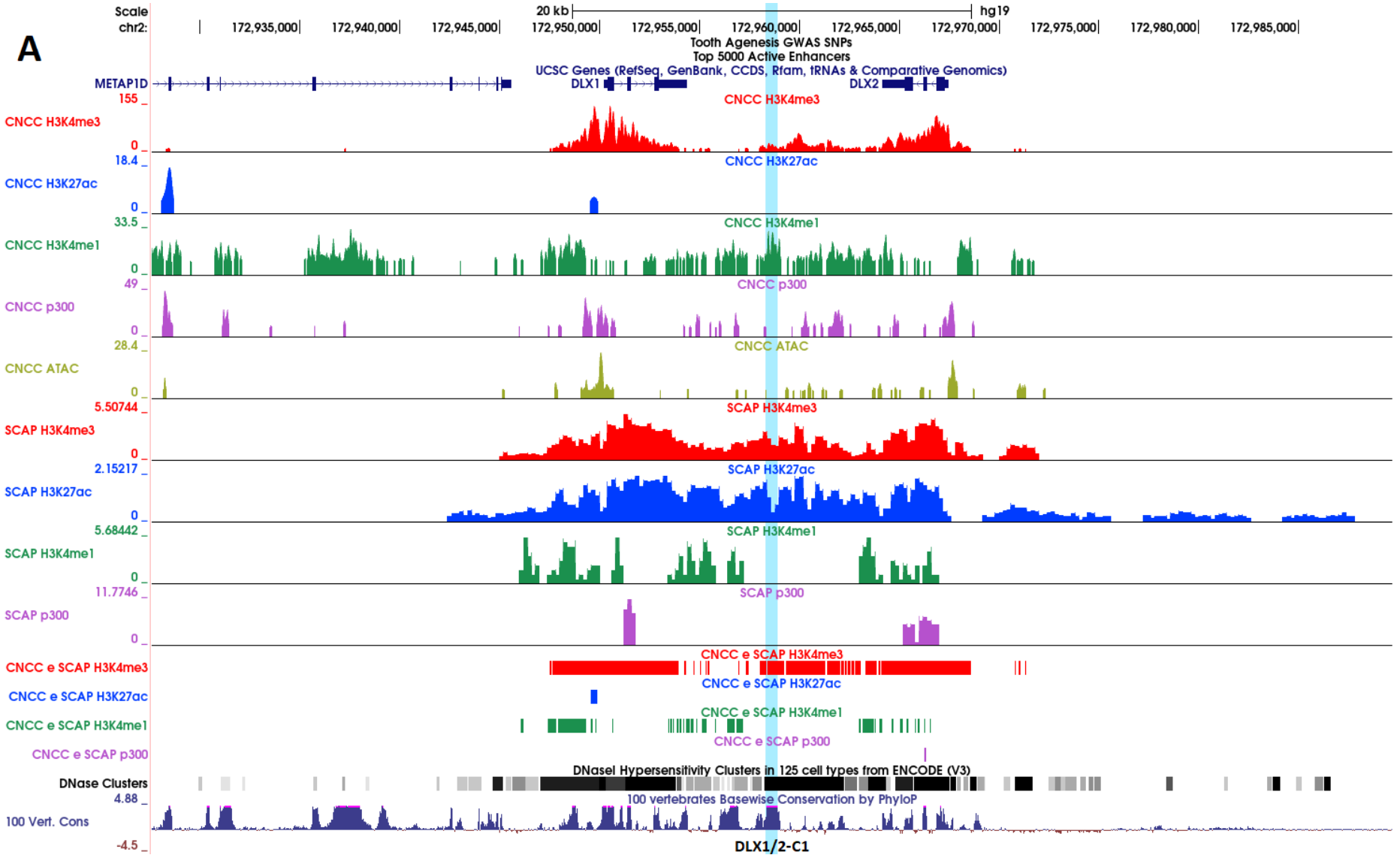
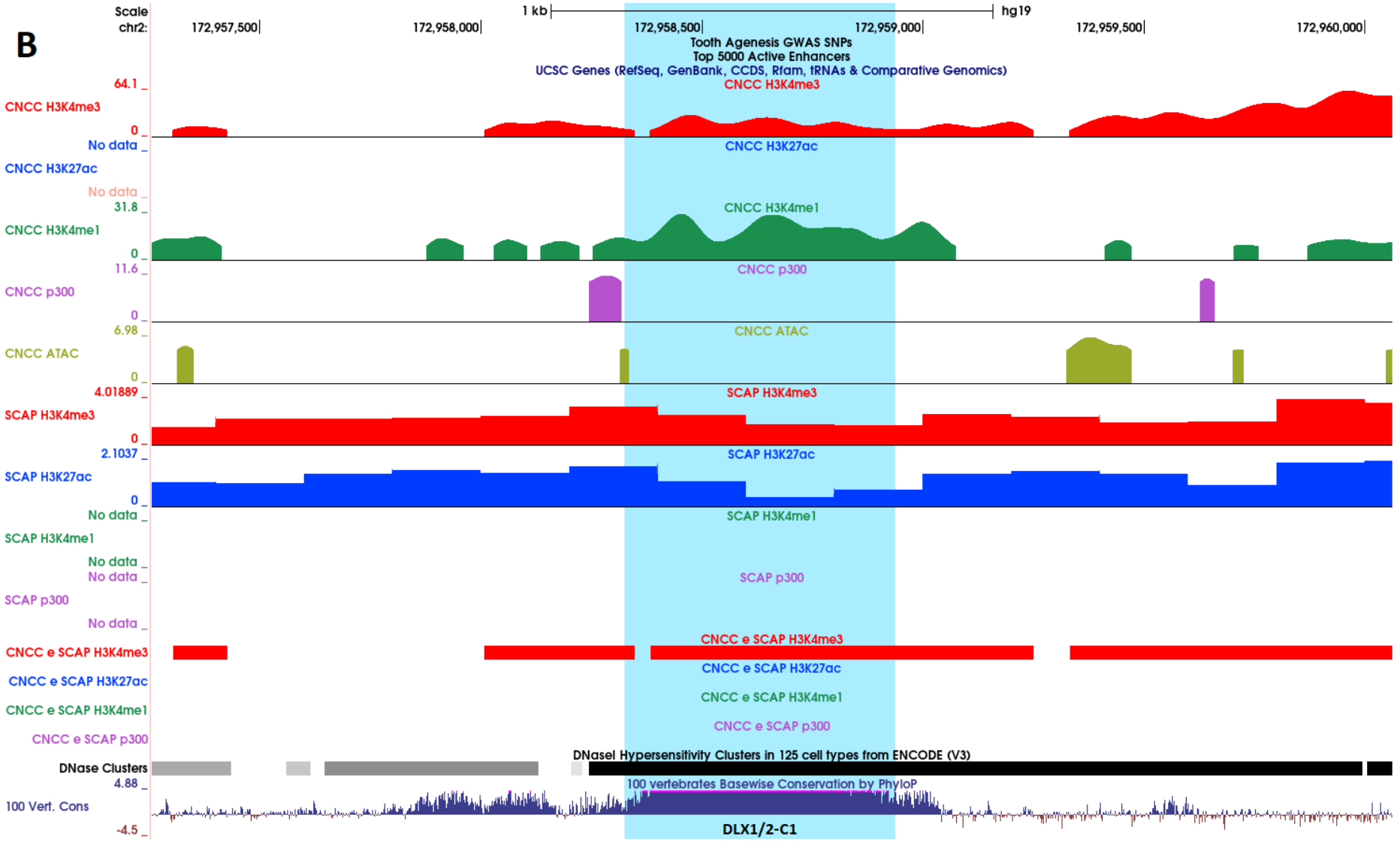


Figura 14. **A)** Mapa genômico contendo o gene LEF1 à esquerda. A fina região grifada com a linha azul claro representa uma das regiões candidatas a *enhancer* do gene LEF1 , o candidato LEF1-C4 (chr4:109271626-109271983), a qual contém um SNP (rs917412) previamente relacionado à casos de agenesia dentária. **B)** Mapa genômico com visualização ampliada para a região candidata a *enhancer* LEF1-C4. **C)** Conservação entre 100 espécies de vertebrados e sítios de ligação de fatores de transcrição da região ao redor do candidato a *enhancer* LEF1-C4. As imagens foram obtidas através do *UCSC Genome Browser.*

5.2.5 *Enhancer* da dupla DLX1-DLX2

Uma região já previamente estudada com atividade de enhancer Park et al., 2004), de aproximadamente 600 bp (DLX1/2-C1) se encontra praticamente equidistante entre a dupla DLX1 e DLX2, estando cerca de 4 kbp *downstream* de DLX1 e 5 kbp *upstream* de DLX2 (Figura 15A). Essa região não codificante exibe grande conservação entre espécies e hipersensibilidade à DNAseI (Figura 15C), além de apresentar grande sinal de H3K4me3 e H3K4me1 para as CNCCs (Figura 15B). Nas SCAPs, o enriquecimento acontece nas marcas de H3K4me3 e H3K27ac (Figura 15B), sendo um possível *poised* (preparado) *enhancer* durante o estágio de CNCC, e um *enhancer* ativo durante o estágio de SCAP.


A
Scale
chr2:
20 kb
hg19
172,935,000
172,940,000
172,945,000
172,950,000
172,955,000
172,960,000
172,965,000
172,970,000
172,975,000
172,980,000
172,985,000
Tooth Agenesis GWAS SNPs
Top 5000 Active Enhancers
UCSC Genes (RefSeq, GenBank, CCDS, Rfam, tRNAs & Comparative Genomics)
METAP1D
DLX1
DLX2
CNCC H3K4me3
CNCC H3K27ac
CNCC H3K4me1
CNCC p300
CNCC ATAC
SCAP H3K4me3
SCAP H3K27ac
SCAP H3K4me1
SCAP p300
CNCC e SCAP H3K4me3
CNCC e SCAP H3K27ac
CNCC e SCAP H3K4me1
CNCC e SCAP p300
DNaseI Hypersensitivity Clusters in 125 cell types from ENCODE (V3)
DNase Clusters
100 vertebrates Basewise Conservation by PhyloP
100 Vert. Cons
DLX1/2-C1


B
Scale
chr2:
1 kb
hg19
172,957,500
172,958,000
172,958,500
172,959,000
172,959,500
172,960,000
Tooth Agenesis GWAS SNPs
Top 5000 Active Enhancers
UCSC Genes (RefSeq, GenBank, CCDS, Rfam, tRNAs & Comparative Genomics)
CNCC H3K4me3
64.1
0
No data
CNCC H3K27ac
No data
31.8
0
CNCC H3K4me1
11.6
0
CNCC p300
6.98
0
CNCC ATAC
4.01889
0
SCAP H3K4me3
2.1037
0
SCAP H3K27ac
No data
SCAP H3K4me1
No data
No data
SCAP p300
No data
CNCC e SCAP H3K4me3
CNCC e SCAP H3K27ac
CNCC e SCAP H3K4me1
CNCC e SCAP p300
DNase Clusters
4.88
100 Vert. Cons
-4.5
DNaseI Hypersensitivity Clusters in 125 cell types from ENCODE (V3)
100 vertebrates Basewise Conservation by PhyloP
DLX1/2-C1

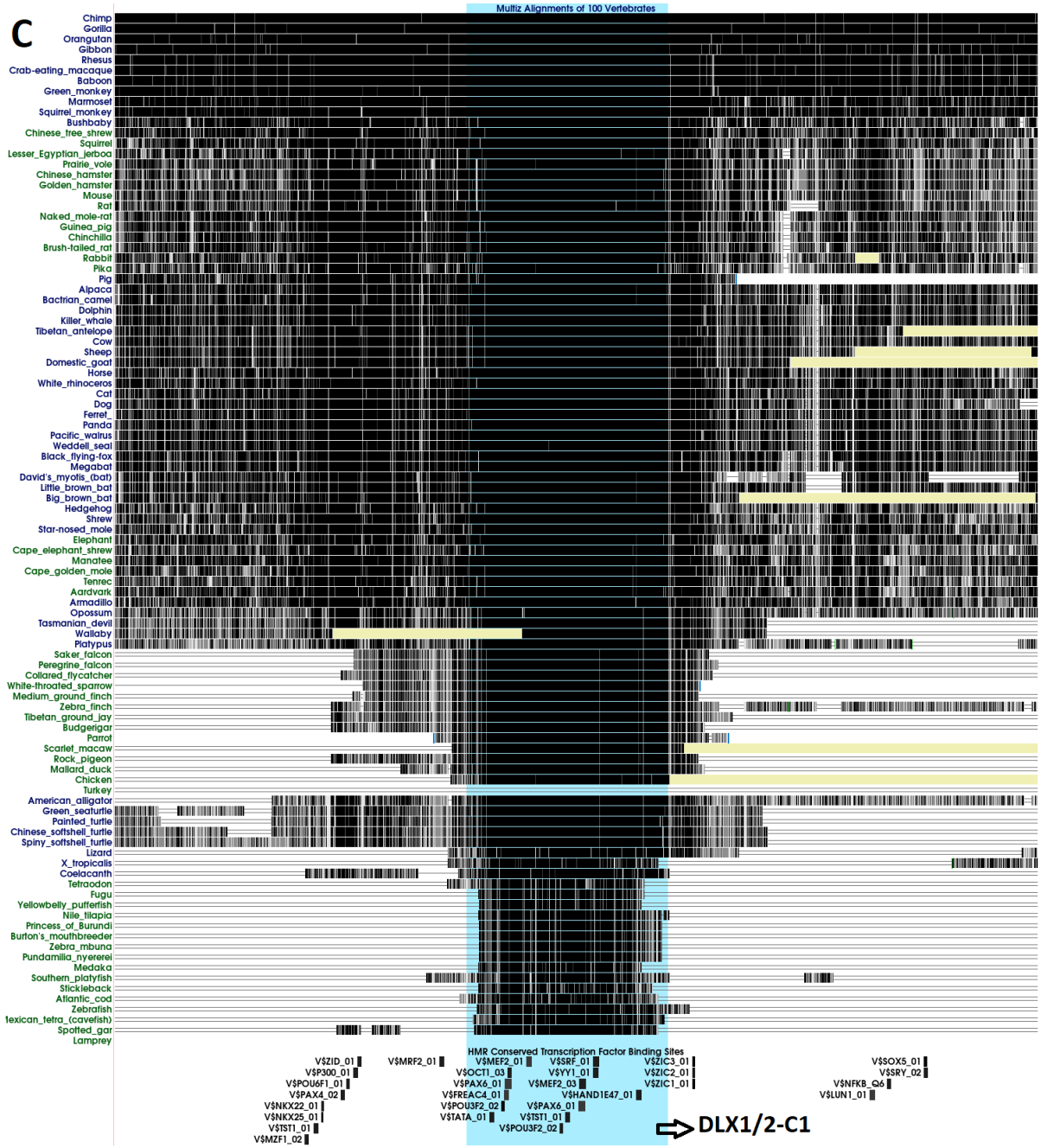


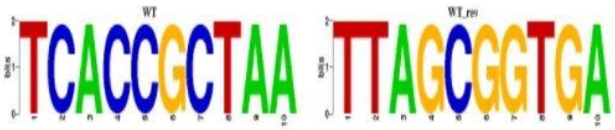
Figura 15. **A)** Mapa genômico contendo os genes DLX1 e DLX2, próximos ao gene METAP1D. A região grifada com a linha azul claro representa uma das regiões candidatas a *enhancer* destes dois genes (DLX1/2-C1) (chr2:172,958,325-172,958,945), a qual se encontra entre os mesmos. **B)** Mapa genômico com visualização ampliada para a região candidata a *enhancer* DLX1/2-C1. **C)** Conservação entre 100 espécies de vertebrados e sítios de ligação de fatores de transcrição da região ao redor do candidato a *enhancer* DLX1/2-C1. As imagens foram obtidas através do *UCSC Genome Browser.*

5.3 Investigação de polimorfismos provenientes de estudos GWAS

Como mencionado anteriormente, um estudo identificou um SNP (rs917412) cerca de 182 kbp de distância do gene LEF1, SNP o qual está presente em uma considerável parte da população da Islândia que apresenta agenesias dentárias. A presença dessa variante, aparenta diminuir a expressão de LEF1, e o efeito somado dessa menor expressão desde o nascimento, até a idade adulta dos indivíduos em questão tende a causar um fenótipo de agenesia dentária variado (Jonsson et al., 2018). Ao se analisar essa região do SNP, levando em consideração que a base afetada está no meio de uma região de 10 bp, a análise STAMP da variante mais comum demonstra afinidade para o fator de transcrição humano TBXT (*T-Box Transcription Factor T*) (Figura 16A), o qual afeta a transcrição de genes necessários na formação e diferenciação do mesoderma e também está presente em células derivadas da notocorda (Edwards et al., 1996). A presença do SNP rs917412, representada por uma mudança de citosina para guanina (C>T), aparenta não afetar a afinidade do motif para o fator de transcrição TBXT, porém possui uma afinidade aumentada com o fator de transcrição PAX5 (Figura 16B), não presente na sequência de variante mais comum. O fator de transcrição PAX5 está associado ao desenvolvimento do sistema nervoso central e dos testículos, e defeitos na expressão deste gene podem contribuir para a transformação neoplásica (Adams et al., 1992; Robichaud et al., 2004).

A

WT

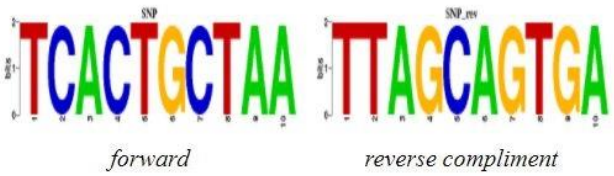


forward *reverse compliment*

Name	*E value*	*Alignment*	*Motif*
TBXT MA0009.1_T	1.0484e-04	-TCACCGCTAA TTCACACCTAG	
MA0242.1_run_Bgb	1.1227e-04	TTAGCGGTGA -TTGCGGTTW	
MA0372.1_RPH1	1.5778e-04	TTAGCGGTGA TTAGGGGN--	
MA0323.1_IXR1	4.3356e-04	TCACCGCTAA------ -CMCCGCTTCCGGYTT	
MA0054.1_myb.Ph3	1.0860e-03	TCACCGCTAA TAACNGTTW-	

B

SNP



forward *reverse compliment*

Name	*E value*	*Alignment*	*Motif*
TBXT MA0009.1_T	1.0484e-04	-TCACTGCTAA TTCACACCTAG	
PAX5 MA0014.1_Pax5	2.0529e-03	TTAGCAGTGA---------- NGNNNNNYGNWGCGKRRCSR	
MA0454.1_odd	3.6208e-03	TCACTGCTAA---- ---NTGCTACTGTN	
MA0127.1_PEND	4.0836e-03	--TTAGCAGTGA MATAAGAANT--	
MA0173.1_CG11617	5.0433e-03	TCACTGCTAA ---ATGTWAA	

Figura 16. Análise de alteração de motif para fatores de transcrição através da ferramenta STAMP. **A)** a sequência com a variante mais comum, TCAC**C**GCTAA. **B)** a sequência com o variante SNP rs917412, TCAC**T**GCTAA.

5.4 Busca ao redor de regiões com atividade de *enhancer* já demonstrada em camundongos transgênicos

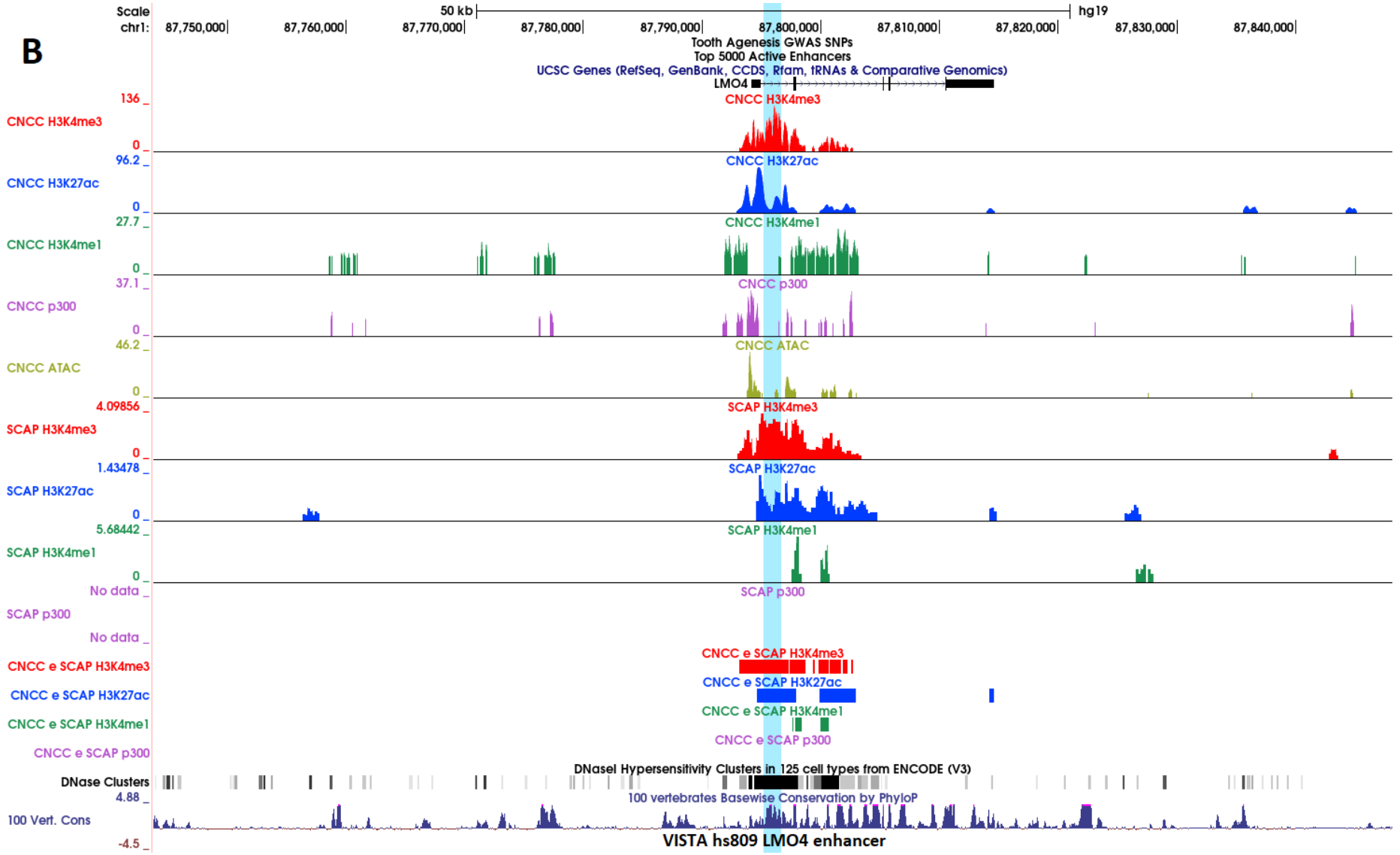
Partindo de uma abordagem contrária para explorar os dados gerados pelo presente trabalho, foram avaliados estudos prévios de atividade de *enhancer* que testaram sequências potencialmente regulatórias em camundongos transgênicos por meio de ensaio de Beta-galactosidase (LacZ). Esses dados foram obtidos da base de dados VISTA *Enhancer Browser*, e a partir desses, foram investigadas tais regiões ao redor de genes próximos tidos como potenciais alvos.

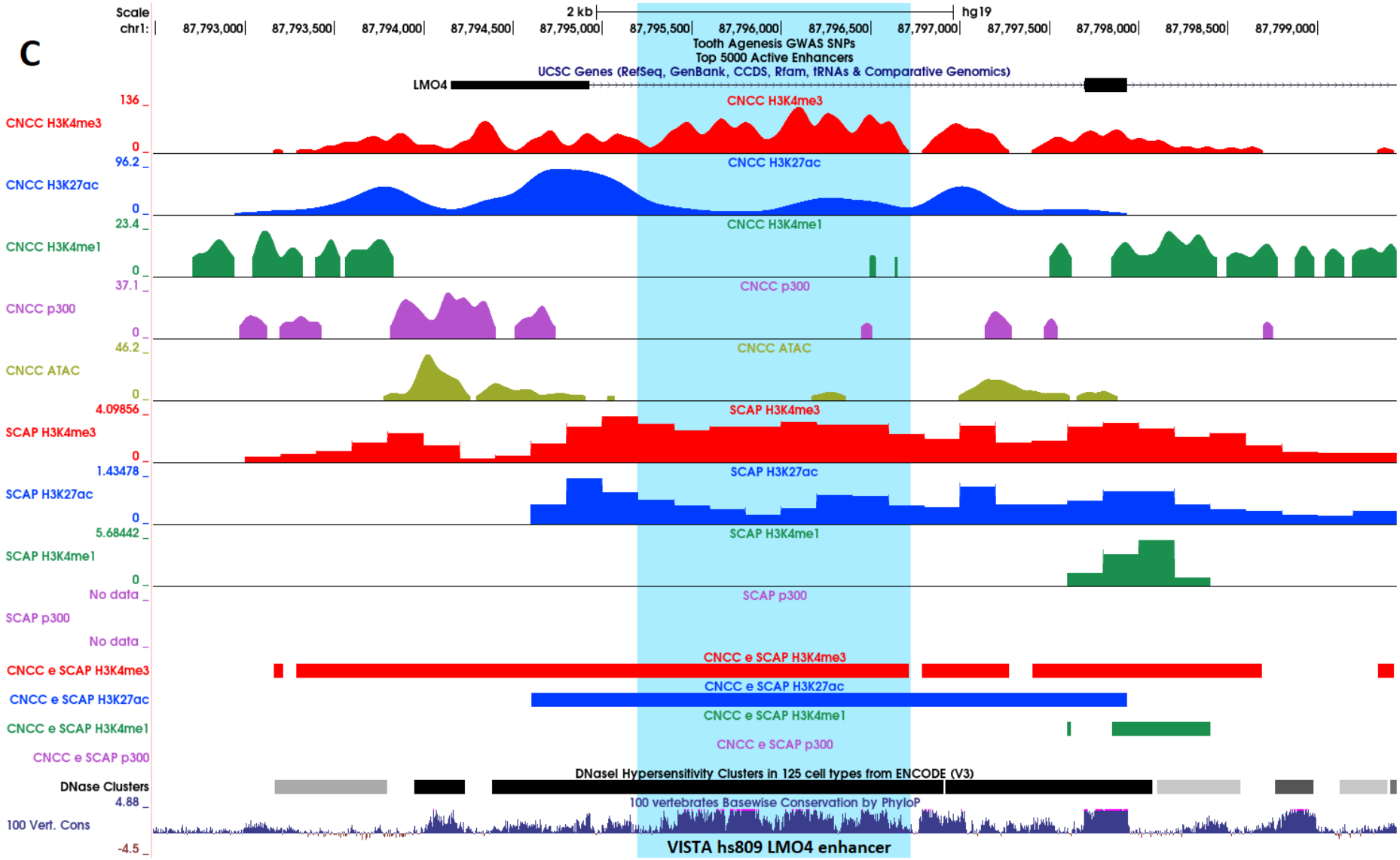
5.4.1 Enhancer próximo a LMO4

Foi encontrado um *enhancer* em camundongo transgênico, o qual se situa em uma região intrônica do gene LMO4, o qual demonstra ter participação no

desenvolvimento dentário (Zhang et al., 2005), sendo a identificação do camundongo **hs809** na base de dados VISTA (Figura 17A). Como regiões de *enhancer* também são expressas junto com os genes os quais elas regulam, foi identificada que essa região está ativa e é expressa nos arcos branquiais e nos nervos cranianos do camundongo, no estágio de embrião. Esse dado é interessante pois ao se investigar a região proposta na base de dados, há um forte enriquecimento das marcas de histona H3K4me3 e H3K27ac tanto para as CNCC quanto para as SCAP (Figuras 17B e 17C). Essa região apresenta uma grande sensibilidade à DNAseI, conservação entre vertebrados e diversos motifs para fatores de transcrição (Figura 17D).

Fonte: Lawrence Berkeley National Laboratory


B
Scale
chr1:
50 kb
hg19
87,750,000
87,760,000
87,770,000
87,780,000
87,790,000
87,800,000
87,810,000
87,820,000
87,830,000
87,840,000
Tooth Agenesis GWAS SNPs
Top 5000 Active Enhancers
UCSC Genes (RefSeq, GenBank, CCDS, Rfam, tRNAs & Comparative Genomics)
LMO4
136
CNCC H3K4me3
0
96.2
CNCC H3K27ac
0
27.7
CNCC H3K4me1
0
37.1
CNCC p300
0
46.2
CNCC ATAC
0
4.09856
SCAP H3K4me3
0
1.43478
SCAP H3K27ac
0
5.68442
SCAP H3K4me1
0
No data
SCAP p300
No data
CNCC e SCAP H3K4me3
CNCC e SCAP H3K27ac
CNCC e SCAP H3K4me1
CNCC e SCAP p300
DNase Clusters
DNaseI Hypersensitivity Clusters in 125 cell types from ENCODE (V3)
4.88
100 Vert. Cons
100 vertebrates Basewise Conservation by PhyloP
-4.5
VISTA hs809 LMO4 enhancer


C
Scale
chr1:
2 kb
hg19
87,793,000
87,793,500
87,794,000
87,794,500
87,795,000
87,795,500
87,796,000
87,796,500
87,797,000
87,797,500
87,798,000
87,798,500
87,799,000
Tooth Agenesis GWAS SNPs
Top 5000 Active Enhancers
UCSC Genes (RefSeq, GenBank, CCDS, Rfam, tRNAs & Comparative Genomics)
LMO4
CNCC H3K4me3
136
0
CNCC H3K27ac
96.2
0
CNCC H3K4me1
23.4
0
CNCC p300
37.1
0
CNCC ATAC
46.2
0
SCAP H3K4me3
4.09856
0
SCAP H3K27ac
1.43478
0
SCAP H3K4me1
5.68442
0
SCAP p300
No data
No data
CNCC e SCAP H3K4me3
CNCC e SCAP H3K27ac
CNCC e SCAP H3K4me1
CNCC e SCAP p300
DNase Clusters
DNaseI Hypersensitivity Clusters in 125 cell types from ENCODE (V3)
100 Vert. Cons
4.88
-4.5
100 vertebrates Basewise Conservation by PhyloP
VISTA hs809 LMO4 enhancer

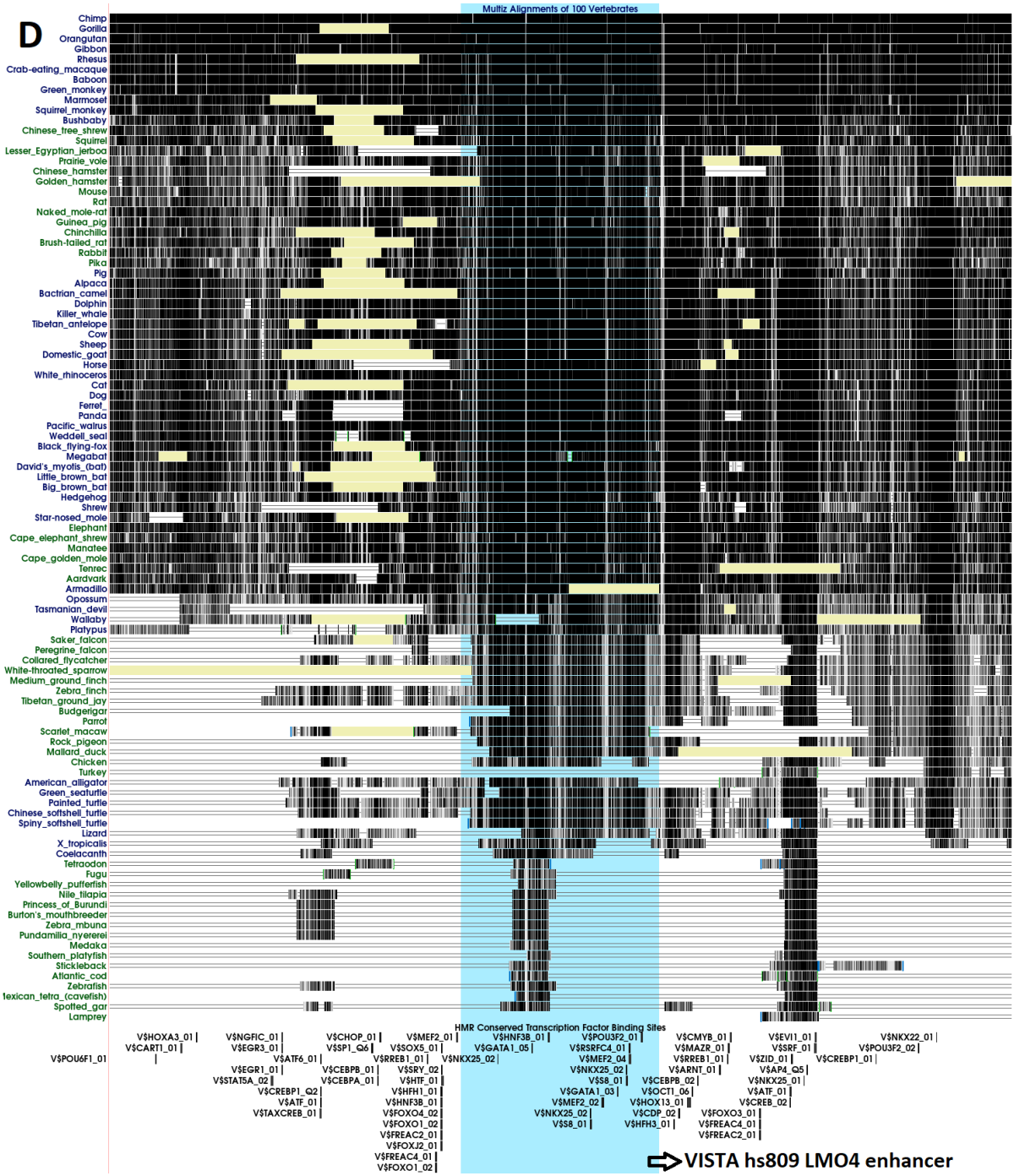
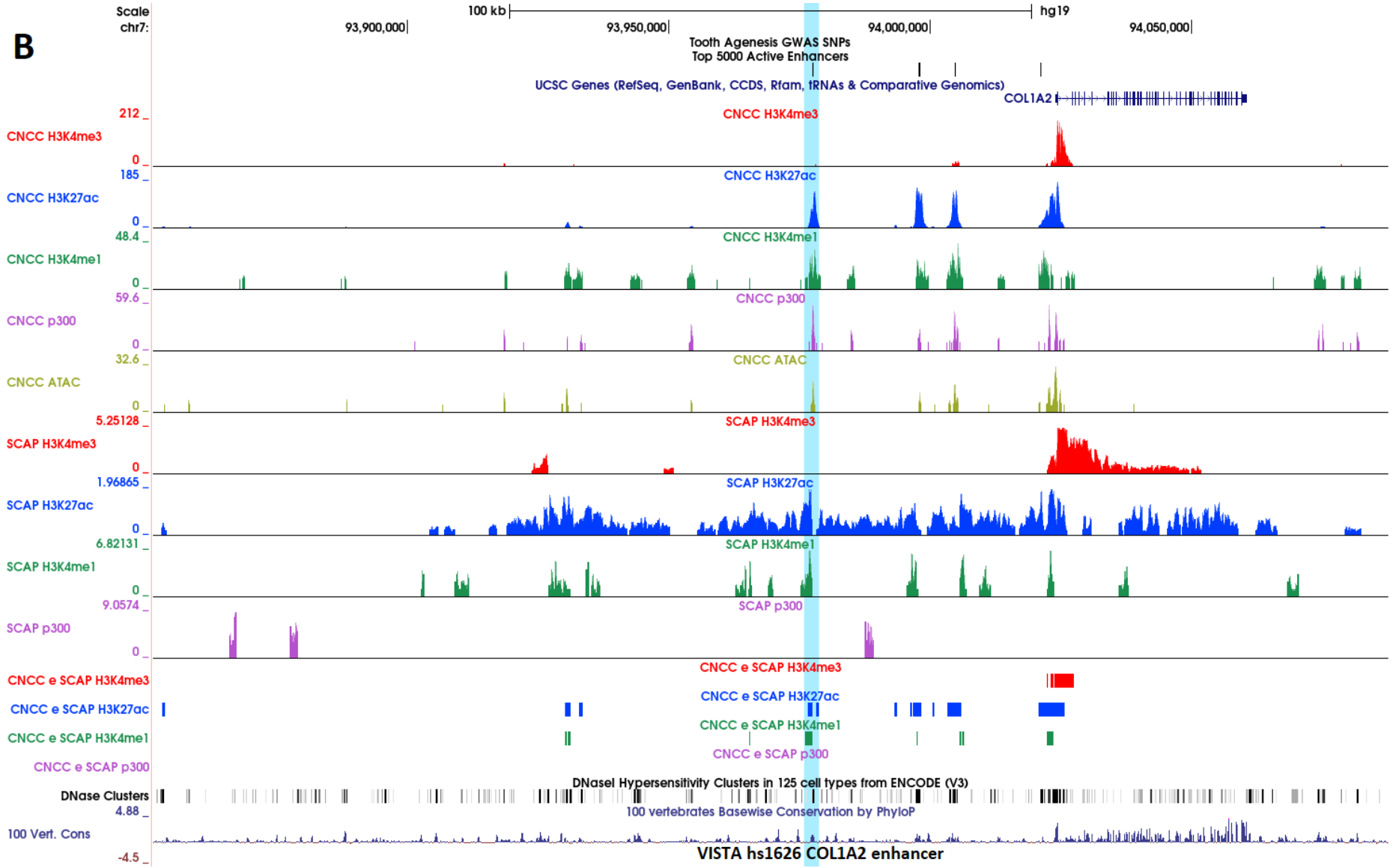


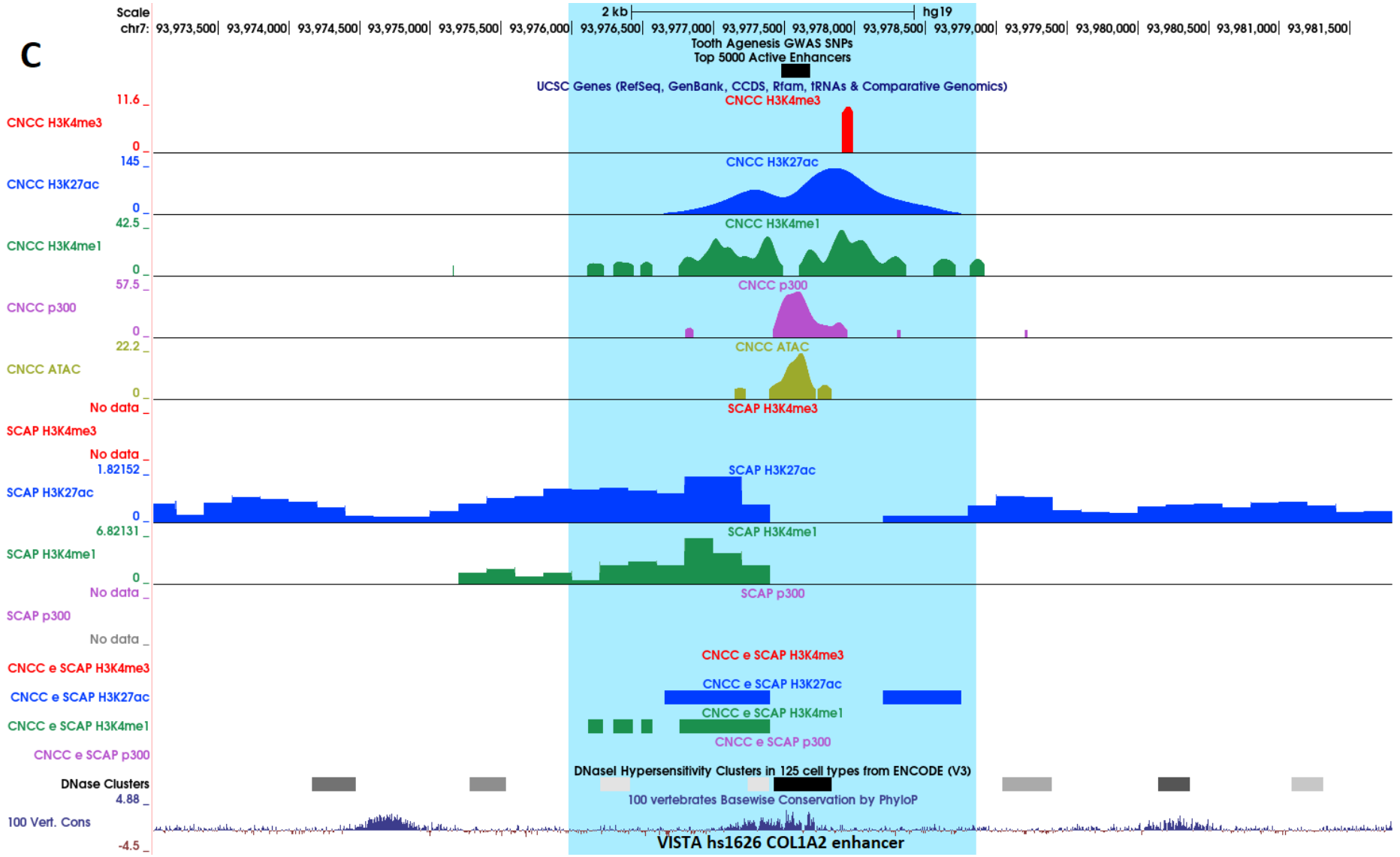
Figura 17. **A)** Embrião de camundongo transgênico hs809, em E11.5, com região do genoma humano chr1:87,795,192-87,796,737, conjugada à gene LacZ reporter, proveniente da base de dados VISTA (https://enhancer.lbl.gov/cgi-bin/imagedb3.pl?form=presentation&show=1&experiment_id=809&organism_id=1). **B)** Mapa genômico contendo o gene LMO4. A região grifada em azul é uma região com atividade de enhancer, demonstrada em camundongo transgênico hs809 como no item A, a qual se encontra no intron 1 do gene LMO4. **C)** Mapa genômico com visualização ampliada para a região de enhancer dentro do intron 1 de LMO4. **D)** Conservação entre 100 espécies de vertebrados e sítios de ligação de fatores de transcrição da região ao redor do *enhancer* encontrado dentro do intron 1 de LMO4. As imagens B, C e D foram obtidas através do *UCSC Genome Browser.*

5.4.2 Enhancer próximo a COL1A2

Outro *enhancer* encontrado próximo ao gene COL1A2 (45 kbp de distância), demonstrou alta atividade na região orofacial do embrião do camundongo, de identificação no VISTA **hs1626** (Figura 18A), e no mapa genômico, a região é altamente enriquecida para todas as marcas de histona em CNCC, e também possui certa acessibilidade da cromatina. Já nas SCAPs, há enriquecimento somente de H3K27ac e H3K4me1, o que não descarta o fato que essa região seja uma boa candidata para *enhancers* em ambos tipos celulares (Figuras 18B e 18C). Essa região possui grande conservação entre diferentes espécies de vertebrados, principalmente no meio da mesma, além de ser reconhecida por diversos fatores de transcrição como PAX4, OCT1 e TATA (TBP) (Figura 18D). Similarmente, a sensibilidade da região à DNAseI aparenta ser enriquecida somente na região mediana do enhancer (Figura 18D). Defeitos na expressão ou no gene COL1A2 tendem a causar *dentinogenesis imperfecta* e *osteogenesis imperfecta* (Marini et al., 2017).

Fonte: Lawrence Berkeley National Laboratory


B
Scale
chr7:
100 kb
hg19
93,900,000
93,950,000
94,000,000
94,050,000
Tooth Agenesis GWAS SNPs
Top 5000 Active Enhancers
UCSC Genes (RefSeq, GenBank, CCDS, Rfam, tRNAs & Comparative Genomics)
COL1A2
CNCC H3K4me3
CNCC H3K27ac
CNCC H3K4me1
CNCC p300
CNCC ATAC
SCAP H3K4me3
SCAP H3K27ac
SCAP H3K4me1
SCAP p300
CNCC e SCAP H3K4me3
CNCC e SCAP H3K27ac
CNCC e SCAP H3K4me1
CNCC e SCAP p300
DNaseI Hypersensitivity Clusters in 125 cell types from ENCODE (V3)
DNase Clusters
100 vertebrates Basewise Conservation by PhyloP
100 Vert. Cons
VISTA hs1626 COL1A2 enhancer


C
Scale
chr7: 93,973,500 93,974,000 93,974,500 93,975,000 93,975,500 93,976,000 93,976,500 93,977,000 93,977,500 93,978,000 93,978,500 93,979,000 93,979,500 93,980,000 93,980,500 93,981,000 93,981,500
2 kb
hg19
Tooth Agenesis GWAS SNPs
Top 5000 Active Enhancers
UCSC Genes (RefSeq, GenBank, CCDS, Rfam, tRNAs & Comparative Genomics)
CNCC H3K4me3
11.6
0
CNCC H3K27ac
145
0
CNCC H3K4me1
42.5
0
CNCC p300
57.5
0
CNCC ATAC
22.2
0
SCAP H3K4me3
No data
1.82152
0
SCAP H3K27ac
6.82131
0
SCAP H3K4me1
No data
SCAP p300
No data
CNCC e SCAP H3K4me3
CNCC e SCAP H3K27ac
CNCC e SCAP H3K4me1
CNCC e SCAP p300
DNase Clusters
DNaseI Hypersensitivity Clusters in 125 cell types from ENCODE (V3)
4.88
100 Vert. Cons
-4.5
100 vertebrates Basewise Conservation by PhyloP
VISTA hs1626 COL1A2 enhancer

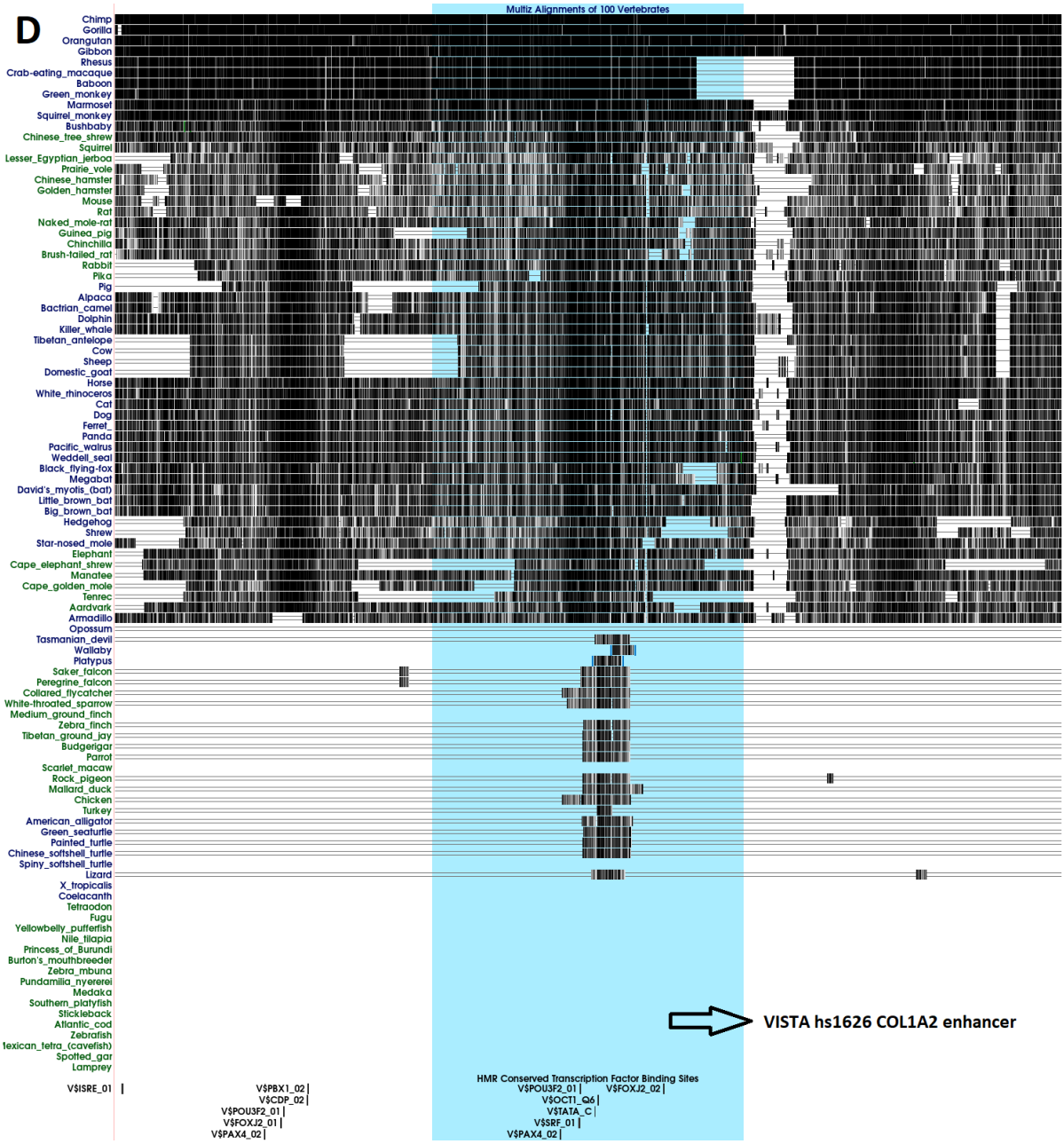
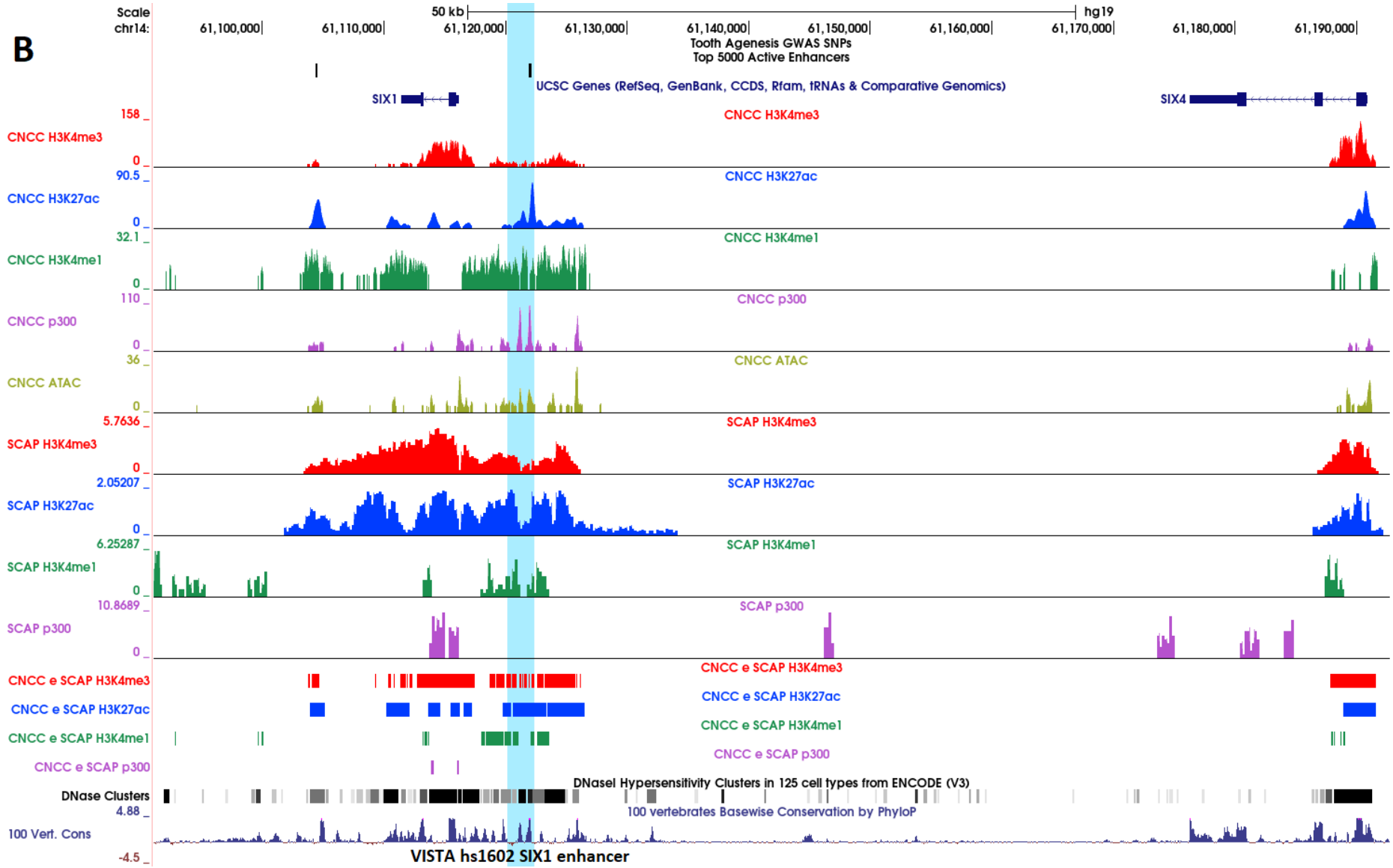


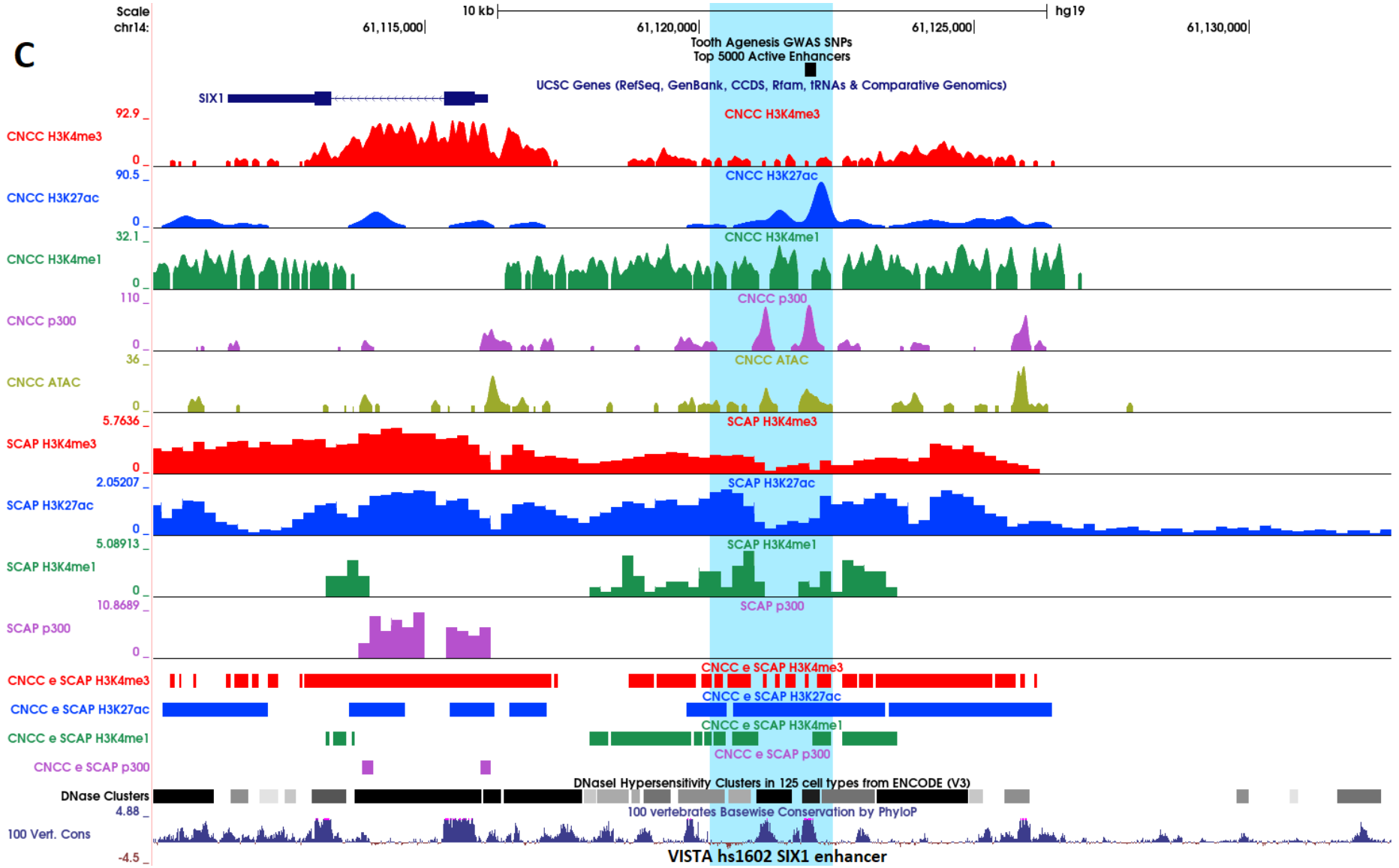
Figura 18. **A)** Embrião de camundongo transgênico hs1626, em E11.5, com região do genoma humano chr7:93,975,962-93,978,880, conjugada à gene LacZ reporter, proveniente da base de dados VISTA (https://enhancer.lbl.gov/cgi-bin/imagedb3.pl?form=presentation&show=1&experiment_id=1626&organism_id=1). **B)** Mapa genômico contendo o gene COL1A2. A região grifada em azul é uma região com atividade de enhancer, demonstrada em camundongo transgênico hs1626 como no item A, a qual se encontra a 45 kbp upstream do gene COL1A2. **C)** Mapa genômico com visualização ampliada para a região de enhancer próxima ao gene COL1A2. **D)** Conservação entre 100 espécies de vertebrados e sítios de ligação de fatores de transcrição da região ao redor do *enhancer* encontrado próximo ao gene COL1A2. As imagens B, C e D foram obtidas através do *UCSC Genome Browser.*

5.4.3 Enhancer próximo a SIX1

Um terceiro exemplo de enhancer identificado em camundongo transgênico é um encontrado próximo ao gene SIX1, o qual demonstrou atividade em regiões do arco branquial e do mesênquima facial do camundongo. O camundongo transgênico em questão é o de código VISTA **hs1602** (Figura 19A) e no mapa genômico, todas as marcas se mostram enriquecidas tanto em CNCC quanto SCAP, porém com ausência do cofator de transcrição p300 para SCAP somente (Figuras 19B e 19C, além de uma conservação entre espécies de vertebrados relativamente alta e ser alvo de diversos fatores de transcrição (Figura 19D). Estudos prévios comprovaram que o gene SIX1 é altamente expresso no mesênquima dentário no estágio de botão durante a formação do dente, e esse gene é necessário para a expressão de PAX9 no mesênquima dentário, o qual induz a invaginação das células da lâmina dentária para o mesênquima dentário, o que eventualmente tomará formato de botão e seguirá os estágios de formação dentário (Nonomura et al., 2010; Takahashi et al., 2020).

Fonte: Lawrence Berkeley National Laboratory


B
Scale
chr14:
50 kb
hg19
61,100,000
61,110,000
61,120,000
61,130,000
61,140,000
61,150,000
61,160,000
61,170,000
61,180,000
61,190,000
Tooth Agenesis GWAS SNPs
Top 5000 Active Enhancers
UCSC Genes (RefSeq, GenBank, CCDS, Rfam, tRNAs & Comparative Genomics)
SIX1
SIX4
158
CNCC H3K4me3
0
90.5
CNCC H3K27ac
0
32.1
CNCC H3K4me1
0
110
CNCC p300
0
36
CNCC ATAC
0
5.7636
SCAP H3K4me3
0
2.05207
SCAP H3K27ac
0
6.25287
SCAP H3K4me1
0
10.8689
SCAP p300
0
CNCC e SCAP H3K4me3
CNCC e SCAP H3K27ac
CNCC e SCAP H3K4me1
CNCC e SCAP p300
DNase Clusters
DNaseI Hypersensitivity Clusters in 125 cell types from ENCODE (V3)
4.88
100 Vert. Cons
100 vertebrates Basewise Conservation by PhyloP
-4.5
VISTA hs1602 SIX1 enhancer


C
Scale
chr14:
10 kb
hg19
61,115,000
61,120,000
61,125,000
61,130,000
Tooth Agenesis GWAS SNPs
Top 5000 Active Enhancers
UCSC Genes (RefSeq, GenBank, CCDS, Rfam, tRNAs & Comparative Genomics)
SIX1
92.9
CNCC H3K4me3
0
CNCC H3K4me3
90.5
CNCC H3K27ac
0
CNCC H3K27ac
32.1
CNCC H3K4me1
0
CNCC H3K4me1
110
CNCC p300
0
CNCC p300
36
CNCC ATAC
0
CNCC ATAC
5.7636
SCAP H3K4me3
0
SCAP H3K4me3
2.05207
SCAP H3K27ac
0
SCAP H3K27ac
5.08913
SCAP H3K4me1
0
SCAP H3K4me1
10.8689
SCAP p300
0
SCAP p300
CNCC e SCAP H3K4me3
CNCC e SCAP H3K4me3
CNCC e SCAP H3K27ac
CNCC e SCAP H3K27ac
CNCC e SCAP H3K4me1
CNCC e SCAP H3K4me1
CNCC e SCAP p300
CNCC e SCAP p300
DNaseI Hypersensitivity Clusters in 125 cell types from ENCODE (V3)
DNase Clusters
4.88
100 vertebrates Basewise Conservation by PhyloP
100 Vert. Cons
-4.5
VISTA hs1602 SIX1 enhancer

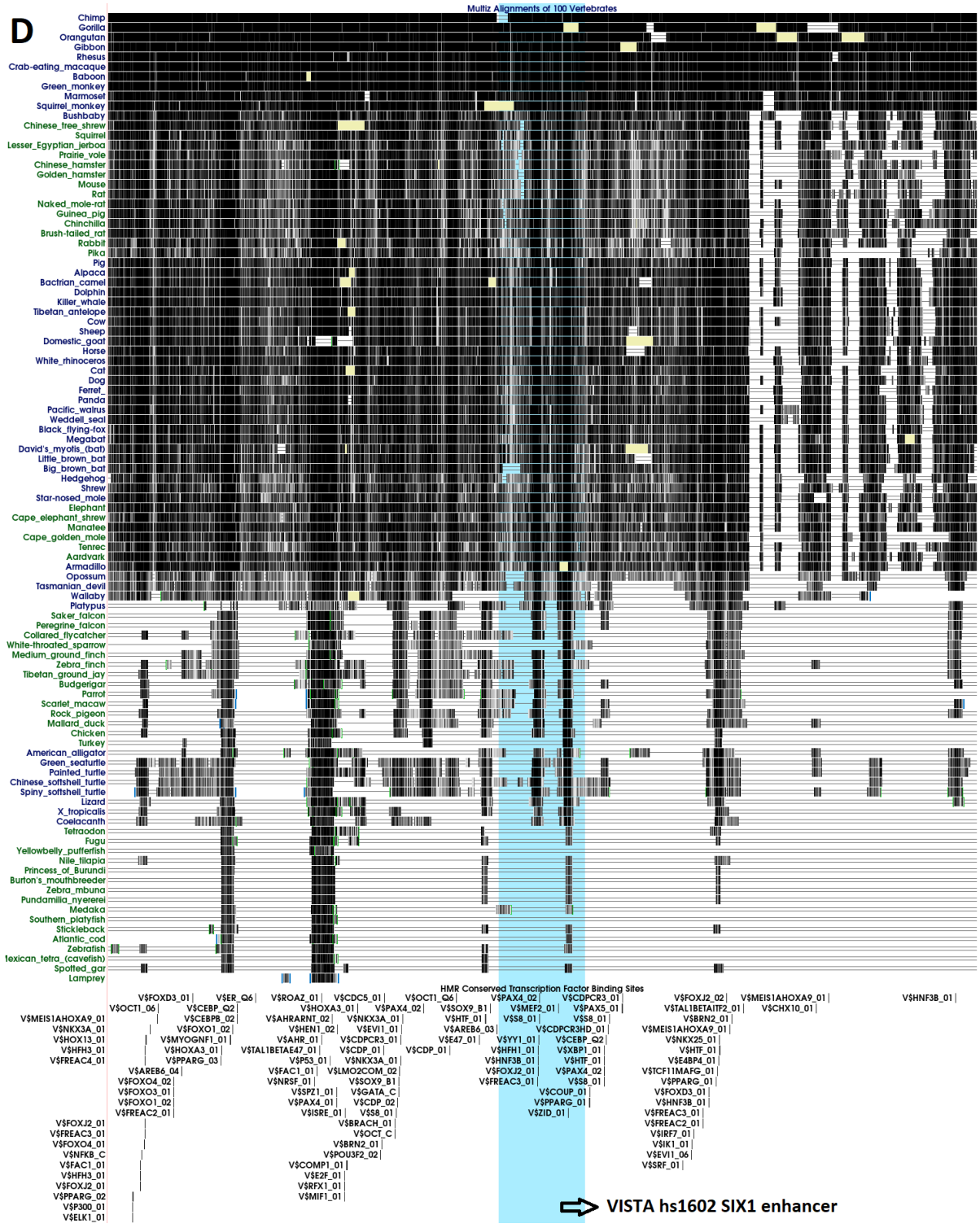


Figura 19. **A)** Embrião de camundongo transgênico hs1602, em E11.5, com região do genoma humano chr14:61,120,199-61,122,451, conjugada à gene LacZ reporter, proveniente da base de dados VISTA (https://enhancer.lbl.gov/cgi-bin/imagedb3.pl?form=presentation&show=1&experiment_id=1602&organism_id=1). **B)** Mapa genômico contendo o gene SIX1 e SIX4. A região grifada em azul é uma região com atividade de enhancer, demonstrada em camundongo transgênico hs1602 como no item A, a qual se encontra cerca de 5 kbp downstream do gene SIX1, e 53 kbp upstream do gene SIX4. **C)** Mapa genômico com visualização ampliada para a região de enhancer próxima ao gene SIX1. **D)** Conservação entre 100 espécies de vertebrados e sítios de

ligação de fatores de transcrição da região ao redor do *enhancer* encontrado próximo ao gene SIX1. As imagens B, C e D foram obtidas através do *UCSC Genome Browser.*

Enhancers adicionais com atividade na região craniofacial de camundongos também foram encontradas, as quais flanqueiam genes importantes no desenvolvimento dentário e outros, como BMP4, SOX9, MSX1 e BARX1.

6 DISCUSSÃO

6.1 Topologia das marcas de histona e de cofator de transcrição p300

De acordo com os nossos dados, a marca de H3K27ac tende a prevalecer em regiões genômicas que estão ativas transcricionalmente, e aparentam estar enriquecidas tanto em regiões promotoras como em regiões regulatórias (como introns, regiões intergênicas). Nas CNCCs e nas SCAPs, essa topologia se mostrou verdadeira, porém vale lembrar que as marcas de H3K27ac geralmente acompanham outras marcas de histona, dependendo da região onde se encontram, e denotam uma região aberta e ativa transcricionalmente.

A topologia das marcas de H3K4me1 em CNCC e SCAP se demonstrou presente em regiões não codificantes como regiões promotoras. Essa marca, geralmente em conjunto com H3K27ac pode indicar uma região cis-regulatória ativa, a qual se for identificada em região não codificante, pode representar um *enhancer*. H3K4me1 também pode estar enriquecida em regiões promotoras, já que essas também possuem um caráter regulatório, dependendo do contexto na qual a célula se encontra (Andersson e Sandelin, 2020).

A marca de H3K4me3 é regularmente associada à ativação da transcrição em regiões codificantes, principalmente quando encontradas junto das marcas de H3K27ac (Beacon et al., 2021). A topologia da marca de H3K4me3 está enriquecida em regiões promotoras das CNCCs e SCAPs. Entretanto, nas SCAPs, a maior porcentagem se encontra em regiões distais intergênicas e intrônicas, enquanto nas CNCCs, a maior porcentagem é encontrada em regiões promotoras como é normalmente esperado.

O cofator de transcrição p300, normalmente conjugado ao cofator CBP, são comumente encontrados em regiões com atividade transcricional, devido à função desse complexo de acetilar histonas e por fim facilitar a atividade da RNA Polimerase tipo II para transcrever uma dada região (Kalkhoven, 2004). Nas CNCCs, a presença do cofator de transcrição p300 se manteve em grande parte nas regiões promotoras, porém não exclusivamente, já que também foi demonstrado pela ferramenta ChIPSeeker aparecer em regiões distais intergênicas nesse tipo celular. Nas SCAPs, curiosamente, a maior parte do sinal do cofator de transcrição p300 esteve presente

em regiões intrônicas e distais intergênicas, sendo a presença do mesmo em regiões promotoras de genes muito menores.

6.2 Regiões com potencial de *enhancer* das CNCCs e SCAPs

Segundo a aquisição dos dados de marcações de histona das SCAP, em comparação com os dados já apresentados das CNCC (Prescott et al., 2015), partindo de uma varredura manual, foi possível agrupar regiões que flanqueiam genes importantes para o desenvolvimento dentário, e comparar o perfil dos dois tipos celulares nestas mesmas regiões.

Inicialmente, a varredura manual ao redor dos genes selecionados neste estudo proporcionou diversos candidatos a *enhancers*, seguida então de uma busca em bases de dados de artigos para verificar se algum candidato já havia sido previamente identificado. Os genes DLX1 e DLX2, conhecidos também como *Distal-less homeobox* 1 e 2, são genes encontrados no cromossomo 2, sendo a distância entre os dois de aproximadamente 10 kbp. São conhecidos por transcreverem fatores de transcrição que afetam diversos genes pertencentes à superfamília TGF-β, além de também exercerem papéis importantes na padronização e na formação craniofacial (Francis-West et al., 1998). Foram encontradas algumas regiões candidatas a *enhancer* que flanqueiam esses essa dupla de genes. Uma região entre os genes DLX1 e DLX2 foi previamente estudada, demonstrando atividade de *enhancer* quando na presença de fatores de transcrição FGF8 e FGF9 (Park et al., 2004). Os sinais encontrados nos nossos dados de SCAP, em comparação com os dados de CNCC, além da alta conservação da região, reforçam o fato de que a região entre a dupla de genes DLX1 e DLX2 é cis-regulatória para os mesmos, e que durante os estágios de CNCC, essa região já está preparada para se tornar um *enhancer* após diferenciações devido à presença da marcas H3K4me3 e H3K4me1 e ausência de H3K27ac. Nas SCAP, a marca de H3K27ac está presente nessa região intergênica, denotando que a região está ativa e com atividade de *enhancer*.

Também levantamos diversas regiões para o gene PAX9, normalmente expresso em regiões do arco branquial, e no mesênquima dentário em vertebrados. Esse gene também possui *enhancers* já previamente descritos, dando como exemplo outra região dentro do intron 7 do gene SLC25A21, observado em um estudo utilizando *zebrafish* (Chatterjee et al., 2011). Similar ao candidato PAX9-C4, encontrado no intron 1 do gene SLC25A21. Esses *enhancers* e candidatos a

enhancer reforçam o fato de que regiões intrônicas também contém potencial cis-regulatório, dependendo do tipo celular, onde certos genes possam estar silenciados, mas suas regiões não codificantes podem estar interagindo e regulando outros genes de maneira distal.

A diferenciação celular progride através de um processo de restrição de linhagem no qual o potencial celular é continuamente perdido ao longo do desenvolvimento do embrião. Nos estágios iniciais do embrião, as células pluripotentes embrionárias têm a capacidade de se diferenciar em todos os tipos de células somáticas, porém essa capacidade é logo restringida durante a formação das três camadas germinativas, a ectoderme, a mesoderme e a endoderme, cada uma dando origem a tipos celulares específicos e distintos ao longo do desenvolvimento.

Nos vertebrados, uma população de células similares à células tronco, chamadas células da crista neural contradizem a noção de que células mais especializadas possuem menor capacidade de diferenciação. Estas células estão localizadas na fronteira entre a placa neural e o ectoderma de superfície, onde são induzidas como um tipo de célula epitelial (Theveneau et al., 2012; Simões-Costa et al., 2015; Zalc et al., 2021) que subsequentemente sofre uma transição epitelial para mesenquimal, delaminando do epitélio dorsal e migrando pelo embrião para popular as regiões ventrais do mesmo, onde essa população de células então se diferencia em diversos tipos celulares (Theveneau et al., 2012; Simões-Costa et al., Soldatov et al., 2019). Células da crista neural que se originam da região rostral do embrião são chamadas de células da crista neural craniana (CNCCs), as quais além de dar origem à células derivadas do ectoderma como os neurônios e a glia, também dão origem à células que são normalmente de linhagem mesodérmica, como ossos, cartilagem e músculo liso (Le Douarin et al., 2004; Zalc et al., 2021). Sendo assim, essas células mesenquimais derivadas de células da crista neural craniana, as quais populam grande parte do esqueleto craniofacial, são comumente denominadas de “ectomesênquima”, a fim de diferenciar as mesmas dos tecidos derivados do mesoderma (Weston et al,. 2004; Zalc et al., 2021). Algumas regiões apresentaram enriquecimento mútuo para certas marcas de histona entre os dois tipos celulares, podendo ser indicativo de uma região que regula algum gene essencial para ambos tipos celulares, como genes de manutenção celular, como por exemplo no candidato PAX9-C1.

Outras regiões selecionadas como candidatas a *enhancer* puderam apresentar enriquecimento em um tipo celular e a ausência do sinal no outro tipo celular, o que é o caso do candidato LEF1-C4, que pode indicar o acontecimento de possíveis alterações epigenéticas durante processos de diferenciação celular. Um *enhancer* específico em um tipo celular pode não ser tão essencial no outro, sendo silenciado ao longo da diferenciação.

Notam-se também diversas regiões as quais marcas como H3K4me1 estavam altamente presentes, porém sem o enriquecimento de H3K27ac, como o candidato WNT5A-C8. Essa regiões tendem a ser conhecidas como regiões de *enhancer* "preparado", ou pré-programado, a qual pode vir a ter um potencial cis-regulatório após sofrer a acetilação de H3K27 no mesmo local. A acetilação pode ocorrer em algum instante da diferenciação para outro tipo de célula, tornando a região antes pré-programada em uma região ativa (Creyghton et al., 2010). Esse processo também pode ocorrer ao contrário, onde a desacetilação de H3K27ac inativa a atividade cis-regulatória da região que já contém H3K4me1.

Certas regiões são quase garantidas de receber um certo tipo de enriquecimento de uma marca de histona específica, como H3K4me3 para regiões promotoras. A avaliação da topologia dos dois tipos celulares estudados oferece uma visão mais abrangente, ao se investigar onde essas marcas se encontram ao longo dos genomas inteiros.

6.3 Regiões com polimorfismos já identificados em estudos GWAS

Nossos dados procuram reforçar a importância de regiões cis-regulatórias em um contexto biológico como um todo nesses dois tipos celulares, além de sua conservação entre espécies que pode ser preferencial para manter um fenótipo em comum, como a formação dos dentes, que acontece de maneira similar entre as espécies de vertebrados. Essas regiões podem ser reguladas de diferentes maneiras espaço-temporais no organismo, sendo ativas em específicos tecidos e inativas em outros, em diferentes estágios de diferenciação celular. Alterações no genoma, como deleções, duplicações, inversões e até polimorfismos de base única têm potencial de alterar não só um tipo celular, mas todos provenientes de diferenciação do mesmo, não só em regiões codificantes mas em regiões não codificantes, onde possa ocorrer uma falha na regulação cis. Tomando como exemplo a região do candidato a *enhancer* LEF1-C4, o qual contém um SNP rs917412 que causa agenesias dentárias

(Jonsson et al., 2018), possui um enriquecimento de sinais nas marcas de estudadas em CNCCs somente, e completa ausência em SCAPs. A presença do SNP diminui a expressão de LEF1, provavelmente por alterar a afinidade da região regulatória por certos fatores de transcrição, denotando a importância desta região em estágios que precedem a formação do dente.

Outro exemplo que demonstra o efeito de variantes em regiões com potencial regulatório é do SNP rs1884302 (T>C) cerca de 350 kbp de distância de BMP2, gene importante no desenvolvimento esquelético, avaliado em um caso de oligodontia não-sindrômica. Nesta variante, a expressão gene BMP2 é menor, denotando que esse SNP pode estar em uma região com efeito cis-regulatório no gene BMP2, porém a maneira como esse efeito surge não foi elucidada no trabalho (Coelho Neto et al., 2014). Este mesmo SNP foi relacionado a casos de craniossinostose sagital não sindrômica (Justice et al., 2012; Justice et al., 2017).

6.4 Regiões com atividade de enhancer já demonstrada em camundongos transgênicos

Partindo de uma análise semelhante em nossos dados, porém ao contrário, também encontramos algumas regiões do genoma humano já previamente identificadas com atividade de *enhancer*, que foram clonadas em camundongos transgênicos para avaliar a expressão de um gene repórter LacZ, principalmente na região da face e boca dos camundongos. Analisando essas regiões com nossos dados de ChIP-seq para SCAP, e os dados já depositados de CNCC usados como comparação, foi possível encontrar enriquecimento para as marcas investigadas, principalmente para H3K27ac, H3K4me1 e H3K4me3. Como exemplo, o enhancer próximo a SIX1 do camundongo transgênico demonstrou enriquecimento similar para todas as marcas, exceto p300 que não estava presente em SCAP, o que denota uma necessidade dessa região ser ativa em vários estágios de diferenciação celular na região da face, porém que não necessita mais ser acetilada pelo complexo protéico p300-CBP, que pode indicar uma perda de H3K27ac em subsequentes diferenciações das SCAP, onde esse enhancer possa não ser mais necessário.

7 CONCLUSÕES

Com tais dados de imunoprecipitação da cromatina em larga escala (genoma inteiro), aliadas aos dados de conservação entre espécies e dados pertinentes às regiões onde a cromatina tende a ser mais acessível, o que acontece de acordo com cada tipo celular, foi possível predizer uma gama de regiões candidatas à *enhancer* de genes importantes para o desenvolvimento dentário e craniofacial. Claramente, muitas dessas regiões podem representar falsos-positivos, porém ao se denominar tais regiões, é possível partir para investigações mais específicas em cada região para confirmar ou rebater a hipótese das mesmas possuírem atividade cis-regulatória.

Para obter dados mais precisos deste estudo exploratório, deverão ser replicados os ensaios de ChIP-seq em amostras de diversos pacientes, levando em consideração a alta variabilidade genética presente na população brasileira. Através do uso das ferramentas disponíveis aliadas aos ensaios de imunoprecipitação da cromatina, foi possível encontrar diversas regiões com possível potencial cis-regulatório, em regiões flanqueadoras de genes importantes no desenvolvimento dentário e craniofacial em células da papila apical. Utilizando também os dados disponíveis para células da crista neural, pôde-se comparar a topologia das marcas de histona entre estes dois tipos celulares, abrindo caminho para mais perguntas. A investigação da determinação da alteração das marcas de histona durante os estágios de diferenciação das CNCCs às células SCAPs, bem como sua topologia no genoma, são essenciais para a melhor compreensão da ação dessas sequências reguladoras no desenvolvimento de alterações dentárias.

A investigação da frequência de interação entre regiões do genoma de interesse precisa ser realizada, usando técnicas de captura de conformação da cromatina como 3C e seus derivados mais complexos como *Hi-C* e *CaptureC*. Essa abordagem elucidaria, por exemplo, a partir de uma região promotora, a identificação de quais regiões do genoma ela estaria próxima tridimensionalmente, o que também oferece um potencial de descobrir novas regiões cis-regulatórias distais não codificantes.

Regiões regulatórias não codificantes, como *enhancers*, quando funcionais tendem a ter alta conservação entre espécies, e pequenas variações nestas regiões podem responder por mudanças fenotípicas (normais ou não) quando tal variação

afeta principalmente motifs de fatores de transcrição. Sendo assim, a estratégia de mapear regiões regulatórias altamente conservadas, mas que apresentem pequenas divergências entre espécies, pode gerar indícios do papel destes *enhancers* na expressão gênica quando avaliada em uma perspectiva evolutiva. Portanto, após o mapeamento das regiões regulatórias de células odontogênicas, pretendemos também tentar associar pontos divergentes destas regiões entre espécies com o fenótipo dentário (número, tamanho, etc.).

Adicionalmente, identificar se existem variações genéticas já associadas a distúrbios de desenvolvimento dentário (má formações, agenesias, variações no número de dentes) que se encontram em *enhancers* identificados previamente. Os ensaios funcionais precisam ser conduzidos para confirmar a atividade *in vitro* dos *enhancers* identificados em relação aos supostos genes regulados, ou até ensaios *in vivo*, utilizando camundongos transgênicos para confirmar se a ausência das regiões encontradas pode afetar a estrutura craniofacial dos camundongos em questão.

REFERÊNCIAS[1]

Adams B, Dörfler P, Aguzzi A, et al. Pax-5 encodes the transcription factor BSAP and is expressed in B lymphocytes, the developing CNS, and adult testis. *Genes Dev*. 1992;6(9):1589-1607. doi:10.1101/gad.6.9.1589

Akiyama K, Chen C, Gronthos S, Shi S. Lineage differentiation of mesenchymal stem cells from dental pulp, apical papilla, and periodontal ligament. Methods Mol Biol. 2012;887:111-121. doi:10.1007/978-1-61779-860-3_11

Amano, T. Gene regulatory landscape of the sonic hedgehog locus in embryonic development. Develop Growth Differ. 2020; 62: 334– 342. https://doi.org/10.1111/dgd.12668

Andersson R, Sandelin A. Determinants of enhancer and promoter activities of regulatory elements. *Nat Rev Genet*. 2020;21(2):71-87. doi:10.1038/s41576-019-0173-8

Attanasio, C., Nord, A.S., Zhu, Y., Blow, M.J., Li, Z., Liberton, D.K., Morrison, H., Plajzer-Frick, I., Holt, A., Hosseini, R., et al. (2013). Fine tuning of craniofacial morphology by distant-acting enhancers. Science 342, 1241006. doi: 10.1126/science.1241006

Bailleul-Forestier I, Berdal A, Vinckier F, de Ravel T, Fryns JP, Verloes A. The genetic basis of inherited anomalies of the teeth. Part 2: syndromes with significant dental involvement. Eur J Med Genet. 2008;51(5):383-408. doi: 10.1016/j.ejmg.2008.05.003

Balic, A., & Thesleff, I. (2015). Tissue Interactions Regulating Tooth Development and Renewal. Craniofacial Development, 157–186. https://doi.org/10.1016/bs.ctdb.2015.07.006

Beacon TH, Delcuve GP, López C, *et al*. The dynamic broad epigenetic (H3K4me3, H3K27ac) domain as a mark of essential genes. *Clin Epigenetics*. 2021;13(1):138. Published 2021 Jul 8. doi:10.1186/s13148-021-01126-1

Bei, M. (2009). Molecular genetics of tooth development. Current Opinion in Genetics & Development, 19(5), 504–510. https://doi.org/10.1016/j.gde.2009.09.002

[1] De acordo com as normas da UNICAMP/FOP, baseadas na padronização do International Committee of Medical Journal Editors - Vancouver Group. Abreviatura dos periódicos em conformidade com o PubMed.

Bergqvist LP. The role of teeth in mammal history. *Braz J Oral Sci*. 2003;2(6):249–257

Carroll SB. Evo-devo and an expanding evolutionary synthesis: a genetic theory of morphological evolution. *Cell*. 2008;134(1):25-36. doi:10.1016/j.cell.2008.06.030

Chatterjee S, Bourque G, Lufkin T. Conserved and non-conserved enhancers direct tissue specific transcription in ancient germ layer specific developmental control genes. *BMC Dev Biol*. 2011;11:63. Published 2011 Oct 20. doi:10.1186/1471-213X-11-63

Chen L, Fish AE, Capra JA. Prediction of gene regulatory enhancers across species reveals evolutionarily conserved sequence properties. *PLoS Comput Biol*. 2018;14(10):e1006484. Published 2018 Oct 4. doi:10.1371/journal.pcbi.1006484

Cleves PA, Hart JC, Agoglia RM, et al. An intronic enhancer of Bmp6 underlies evolved tooth gain in sticklebacks. *PLoS Genet*. 2018;14(6):e1007449. Published 2018 Jun 14. doi:10.1371/journal.pgen.1007449

Coelho Neto OL, Reis MF, de Sabóia TM, Tannure PN, Antunes LS, Antonio AG. Clinical and genetic analysis of a nonsyndromic oligodontia in a child. *Case Rep Dent*. 2014;2014:137621. doi:10.1155/2014/137621

Courtney JM, Blackburn J, Sharpe PT. The Ectodysplasin and NFkappaB signalling pathways in odontogenesis. *Arch Oral Biol*. 2005;50(2):159-163. doi:10.1016/j.archoralbio.2004.11.019

Creyghton MP, Cheng AW, Welstead GG, *et al*. Histone H3K27ac separates active from poised enhancers and predicts developmental state. *Proc Natl Acad Sci U S A*. 2010;107(50):21931-21936. doi:10.1073/pnas.1016071107

Davidson, E. H. (1993). Later embryogenesis: Regulatory circuitry in morphogenetic fields. Development, 118(3), 665–690. https://doi.org/10.1242/dev.118.3.665

Davie K, Jacobs J, Atkins M, *et al*. Discovery of transcription factors and regulatory regions driving in vivo tumor development by ATAC-seq and FAIRE-seq open chromatin profiling. *PLoS Genet*. 2015;11(2):e1004994. Published 2015 Feb 13. doi:10.1371/journal.pgen.1004994

Driesen RB, Gervois P, Vangansewinkel T, Lambrichts I. Unraveling the Role of the Apical Papilla During Dental Root Maturation. *Front Cell Dev Biol*. 2021;9:665600. Published 2021 May 6. doi:10.3389/fcell.2021.665600

Edwards YH, Putt W, Lekoape KM, et al. The human homolog T of the mouse T(Brachyury) gene; gene structure, cDNA sequence, and assignment to chromosome 6q27. *Genome Res*. 1996;6(3):226-233. doi:10.1101/gr.6.3.226

Egusa H, Sonoyama W, Nishimura M, Atsuta I, Akiyama K. Stem cells in dentistry--part I: stem cell sources. *J Prosthodont Res*. 2012;56(3):151-165. doi:10.1016/j.jpor.2012.06.001

Fehrmann C, Dörfer CE, Fawzy El-Sayed KM. Toll-like Receptor Expression Profile of Human Stem/Progenitor Cells Form the Apical Papilla. *J Endod*. 2020;46(11):1623-1630. doi:10.1016/j.joen.2020.08.017

Fleischmannova J, Matalova E, Tucker AS, Sharpe PT. Mouse models of tooth abnormalities. Eur J Oral Sci. 2008;116(1):1-10. doi: 10.1111/j.1600-0722.2007.00504.x.

Francis-West P, Ladher R, Barlow A, Graveson A. Signalling interactions during facial development. *Mech Dev*. 1998;75(1-2):3-28. doi:10.1016/s0925-4773(98)00082-3

Fraser GJ, Hulsey CD, Bloomquist RF, Uyesugi K, Manley NR, Streelman JT. An ancient gene network is co-opted for teeth on old and new jaws. *PLoS Biol*. 2009;7(2):e31. doi:10.1371/journal.pbio.1000031

Gammill LS, Bronner-Fraser M. 2003 Neural crest specification: migrating into genomics. Nat. Rev. Neurosci. 4, 795– 805. doi:10.1038/nrn1219

Gaudin A, Tolar M, Peters OA. Lipoxin A4 Attenuates the Inflammatory Response in Stem Cells of the Apical Papilla via ALX/FPR2. *Sci Rep*. 2018;8(1):8921. Published 2018 Jun 11. doi:10.1038/s41598-018-27194-7

Gompel, N., Prud'homme, B., Wittkopp, P.J., Kassner, V.A., and Carroll, S.B.(2005). Chance caught on the wing: cis-regulatory evolution and the origin ofpigment patterns in Drosophila. Nature 433, 481–487. https://doi.org/10.1038/nature03235

Hovorakova M, Lesot H, Peterka M, Peterkova R. Early development of the human dentition revisited. *J Anat*. 2018;233(2):135-145. doi:10.1111/joa.12825

Huang GT, Sonoyama W, Liu Y, Liu H, Wang S, Shi S. The hidden treasure in apical papilla: the potential role in pulp/dentin regeneration and bioroot engineering. *J Endod*. 2008;34(6):645-651. doi:10.1016/j.joen.2008.03.001

Huang GT, Gronthos S, Shi S. Mesenchymal stem cells derived from dental tissues vs. those from other sources: their biology and role in regenerative medicine. *J Dent Res*. 2009;88(9):792-806. doi:10.1177/0022034509340867

Iwafuchi-Doi, M., & Zaret, K. S. (2014). Pioneer transcription factors in cell reprogramming. Genes & Development, 28(24), 2679–2692. doi:10.1101/gad.253443.114

Jernvall J, Thesleff I. Tooth shape formation and tooth renewal: evolving with the same signals. *Development*. 2012;139(19):3487-3497. doi:10.1242/dev.085084

Jonsson L, Magnusson TE, Thordarson A, et al. Rare and Common Variants Conferring Risk of Tooth Agenesis. *J Dent Res*. 2018;97(5):515-522. doi:10.1177/0022034517750109

Justice CM, Yagnik G, Kim Y, et al. A genome-wide association study identifies susceptibility loci for nonsyndromic sagittal craniosynostosis near BMP2 and within BBS9. *Nat Genet*. 2012;44(12):1360-1364. doi:10.1038/ng.2463

Justice CM, Kim J, Kim SD, et al. A variant associated with sagittal nonsyndromic craniosynostosis alters the regulatory function of a non-coding element. *Am J Med Genet A*. 2017;173(11):2893-2897. doi:10.1002/ajmg.a.38392

Kalkhoven E. CBP and p300: HATs for different occasions. *Biochem Pharmacol*. 2004;68(6):1145-1155. doi:10.1016/j.bcp.2004.03.045

Kapadia H, Mues G, D'Souza R. Genes affecting tooth morphogenesis. Orthod Craniofac Res. 2007;10(4):237-244. doi: 10.1111/j.1601-6343.2007.00395.x.

Kim TK, Shiekhattar R. Architectural and Functional Commonalities between Enhancers and Promoters. *Cell*. 2015;162(5):948-959. doi:10.1016/j.cell.2015.08.008

Koussoulakou DS, Margaritis LH, Koussoulakos SL. A curriculum vitae of teeth: evolution, generation, regeneration. Int J Biol Sci. 2009;5(3):226-243. doi: 10.7150/ijbs.5.226

Kurtén, B. (1963). Return of a lost structure in the evolution of the felid dentition. *Soc Sci Fenn Comm Biol* 4, 1–12.

Le Douarin NM, Creuzet S, Couly G, Dupin E. Neural crest cell plasticity and its limits. *Development*. 2004;131(19):4637-4650. doi:10.1242/dev.01350

Leche W. Zür Entwicklungsgeschichte des Zahnsystems des Säugethiere, zugleich ein Beitrag zur Stammesgeschichte dieser Thiergruppe. I. Ontogenie. *Zoologica* 1895;6, 1–160.

Letra A, Chiquet B, Hansen-Kiss E, Menezes S, Hunter E. Nonsyndromic Tooth Agenesis Overview. In: Adam MP, Ardinger HH, Pagon RA, et al., eds. *GeneReviews*®. Seattle (WA): University of Washington, Seattle; July 22, 2021.

Lettice, L.A., Heaney, S.J.H., Purdie, L.A., Li, L., de Beer, P., Oostra, B.A., Goode, D., Elgar, G., Hill, R.E., and de Graaff, E. (2003). A long-range Shh enhancer regulates expression in the developing limb and fin and is associated with preaxial polydactyly. Hum. Mol. Genet. 12, 1725–1735. doi: 10.1093/hmg/ddg180

Levine M. Transcriptional enhancers in animal development and evolution. *Curr Biol*. 2010;20(17):R754-R763. doi:10.1016/j.cub.2010.06.070

Line SRP. Variation of tooth number in mammalian dentition: connecting genetics, development, and evolution. Evol Devel. 2003;5(3):295–304. doi:10.1046/j.1525-142x.2003.03036.x.

Liu H, Gronthos S, Shi S. Dental pulp stem cells. Methods Enzymol. 2006;419:99-113. doi:10.1016/S0076-6879(06)19005-9

Liu XM, Liu Y, Yu S, Jiang LM, Song B, Chen X. Potential immunomodulatory effects of stem cells from the apical papilla on Treg conversion in tissue regeneration for regenerative endodontic treatment. *Int Endod J*. 2019;52(12):1758-1767. doi:10.1111/iej.13197

Long HK, Prescott SL, Wysocka J. Ever-Changing Landscapes: Transcriptional Enhancers in Development and Evolution. *Cell*. 2016;167(5):1170-1187. doi:10.1016/j.cell.2016.09.018

Mammoto, T., Mammoto, A., Torisawa, Y. S., Tat, T., Gibbs, A., Derda, R., et al. (2011). Mechanochemical control of mesenchymal condensation and embryonic tooth organ formation. Developmental Cell, 21(4), 758–769. https://doi.org/10.1016/j.devcel.2011.07.006

Marini JC, Forlino A, Bächinger HP, et al. Osteogenesis imperfecta. *Nat Rev Dis Primers*. 2017;3:17052. Published 2017 Aug 18. doi:10.1038/nrdp.2017.52

Martin A, Orgogozo V. The Loci of repeated evolution: a catalog of genetic hotspots of phenotypic variation. *Evolution*. 2013;67(5):1235-1250. doi:10.1111/evo.12081

Mina, M., & Kollar, E. J. (1987). The induction of odontogenesis in non-dental mesenchyme combined with early murine mandibular arch epithelium. Archives of Oral Biology, 32(2), 123–127. https://doi.org/10.1016/0003-9969(87)90055-0

Mostowska A, Kobielak A, Trzeciak WH. Molecular basis of non-syndromic tooth agenesis: mutations of MSX1 and PAX9 reflect their role in patterning human dentition. Eur J Oral Sci. 2003;111(5):365-370. doi: 10.1034/j.1600-0722.2003.00069.x

Mu Y, Tian R, Xiao L, et al. Molecular Evolution of Tooth-Related Genes Provides New Insights into Dietary Adaptations of Mammals. *J Mol Evol*. 2021;89(7):458-471. doi:10.1007/s00239-021-10017-1

Munroe RJ, Prabhu V, Acland GM, et al. Mouse H6 Homeobox 1 (Hmx1) mutations cause cranial abnormalities and reduced body mass. *BMC Dev Biol*. 2009;9:27. Published 2009 Apr 20. doi:10.1186/1471-213X-9-27

Nonomura K, Takahashi M, Wakamatsu Y, Takano-Yamamoto T, Osumi N. Dynamic expression of Six family genes in the dental mesenchyme and the epithelial ameloblast stem/progenitor cells during murine tooth development. *J Anat*. 2010;216(1):80-91. doi:10.1111/j.1469-7580.2009.01167.x

Owen R. Odontography; or a Treatise on the Comparative Anatomy of the Teeth; Their Physiological Relations, Mode of Development, and Microscopic Structure, in the Vertebrate Animals. *Br Foreign Med Rev*. 1845;20(40):376-406.

Park BK, Sperber SM, Choudhury A, et al. Intergenic enhancers with distinct activities regulate Dlx gene expression in the mesenchyme of the branchial arches. *Dev Biol*. 2004;268(2):532-545. doi:10.1016/j.ydbio.2004.01.010

Park JY, Jeon HJ, Kim TY, et al. Comparative analysis of mesenchymal stem cell surface marker expression for human dental mesenchymal stem cells. Regen Med. 2013;8(4):453-466. doi:10.2217/rme.13.23

Prescott SL, Srinivasan R, Marchetto MC, et al. Enhancer divergence and cis-regulatory evolution in the human and chimp neural crest. Cell. 2015;163(1):68-83. doi:10.1016/j.cell.2015.08.036

Rada-Iglesias A, Prescott SL, Wysocka J. 2013 Human genetic variation within neural crest enhancers: molecular and phenotypic implications. Phil Trans R Soc B 368: 20120360. https://doi.org/10.1098/rstb.2012.0360

Rada-Iglesias A. Genetic variation within transcriptional regulatory elements and its implications for human disease. Biol Chem. 2014;395(12):1453-1460. doi: 10.1515/hsz-2014-0109

Robichaud GA, Nardini M, Laflamme M, Cuperlovic-Culf M, Ouellette RJ. Human Pax-5 C-terminal isoforms possess distinct transactivation properties and are differentially modulated in normal and malignant B cells. *J Biol Chem*. 2004;279(48):49956-49963. doi:10.1074/jbc.M407171200

Sauka-Spengler T, Bronner-Fraser M. 2008 A gene regulatory network orchestrates neural crest formation. Nat. Rev. Mol. Cell Biol. 9, 557– 568. doi:10.1038/nrm2428

Schaffner W. Enhancers, enhancers - from their discovery to today's universe of transcription enhancers. *Biol Chem*. 2015;396(4):311-327. doi:10.1515/hsz-2014-0303

Shapiro, M.D., Marks, M.E., Peichel, C.L., Blackman, B.K., Nereng, K.S., Jónsson, B., Schluter, D.,and Kingsley, D.M. (2004). Genetic and developmental basis of evolutionary pelvic reduction in threespine sticklebacks. Nature 428,717–723. doi: 10.1038/nature02415

Simões-Costa M, Bronner ME. Establishing neural crest identity: a gene regulatory recipe. *Development*. 2015;142(2):242-257. doi:10.1242/dev.105445

Soldatov R, Kaucka M, Kastriti ME, et al. Spatiotemporal structure of cell fate decisions in murine neural crest. *Science*. 2019;364(6444):eaas9536. doi:10.1126/science.aas9536

Soukup V, Epperlein HH, Horácek I, Cerny R. Dual epithelial origin of vertebrate oral teeth. *Nature*. 2008;455(7214):795-798. doi:10.1038/nature07304

Stern DL, Orgogozo V. The loci of evolution: how predictable is genetic evolution?. *Evolution*. 2008;62(9):2155-2177. doi:10.1111/j.1558-5646.2008.00450.x

Stock DW. The genetic basis of modularity in the development and evolution of the vertebrate dentition. Philos Trans R Soc Lond B Biol Sci. 2001;356(1414):1633-1653. doi: 10.1098/rstb.2001.0917

Takahashi M, Ikeda K, Ohmuraya M, et al. Six1 is required for signaling center formation and labial-lingual asymmetry in developing lower incisors. *Dev Dyn*. 2020;249(9):1098-1116. doi:10.1002/dvdy.174

Thesleff I, Vaahtokari A, Partanen AM. Regulation of organogenesis. Common molecular mechanisms regulating the development of teeth and other organs. Int J Dev Biol. 1995;39(1):35-50.

Theveneau E, Mayor R. Neural crest migration: interplay between chemorepellents, chemoattractants, contact inhibition, epithelial-mesenchymal transition, and collective cell migration. *Wiley Interdiscip Rev Dev Biol*. 2012;1(3):435-445. doi:10.1002/wdev.28

Townsend G, Harris EF, Lesot H, Clauss F, Brook A. Morphogenetic fields within the human dentition: A new, clinically relevant synthesis of an old concept. Arch Oral Biol. 2008. doi: 10.1016/j.archoralbio.2008.06.011

Tubbs RS, Bosmia AN, Cohen-Gadol AA. The human calvaria: a review of embryology, anatomy, pathology, and molecular development. *Childs Nerv Syst*. 2012;28(1):23-31. doi:10.1007/s00381-011-1637-0

Volponi AA, Pang Y, Sharpe PT. Stem cell-based biological tooth repair and regeneration. *Trends Cell Biol*. 2010;20(12):715-722. doi:10.1016/j.tcb.2010.09.012

Weston JA, Yoshida H, Robinson V, Nishikawa S, Fraser ST, Nishikawa S. Neural crest and the origin of ectomesenchyme: neural fold heterogeneity suggests an alternative hypothesis. *Dev Dyn*. 2004;229(1):118-130. doi:10.1002/dvdy.10478

Wittkopp PJ, Kalay G. Cis-regulatory elements: molecular mechanisms and evolutionary processes underlying divergence. *Nat Rev Genet*. 2011;13(1):59-69. Published 2011 Dec 6. doi:10.1038/nrg3095

Wray GA. The evolutionary significance of cis-regulatory mutations. *Nat Rev Genet*. 2007;8(3):206-216. doi:10.1038/nrg2063

Yáñez-Cuna, J.O., Kvon, E.Z., and Stark, A. (2013). Deciphering the transcriptional cis-regulatory code. Trends Genet. TIG 29, 11–22. doi: 10.1016/j.tig.2012.09.007

Ye X, Attaie AB. Genetic Basis of Nonsyndromic and Syndromic Tooth Agenesis. *J Pediatr Genet*. 2016;5(4):198-208. doi:10.1055/s-0036-1592421

Zalc A, Sinha R, Gulati GS, et al. Reactivation of the pluripotency program precedes formation of the cranial neural crest. *Science*. 2021;371(6529):eabb4776. doi:10.1126/science.abb4776

Zhang L, Hua F, Sun ZJ, Zhang Q, Fan MW, Chen Z. *Zhonghua Kou Qiang Yi Xue Za Zhi*. 2005;40(5):398-401.

Zhang Y, Sun Z, Jia J, et al. Overview of Histone Modification. *Adv Exp Med Biol*. 2021;1283:1-16. doi:10.1007/978-981-15-8104-

APÊNDICE

Tabela S1. Tabela das regiões investigadas e quantificação da cobertura e enriquecimento das marcas de H3K4me3, H3K27ac, cofator de transcrição p300 e H3K4me1 nas SCAPs e CNCCs, além de dados de conservação e hipersensibilidade à DNAseI já integrados ao *UCSC Genome Browser*. Os símbolos representam: (+++) = Cobertura total (>95%) da região selecionada; (++) = Cobertura parcial (~50%) da região selecionada; Ausente = Sem sinal.

			SCAP				CNCC					Genome B.	
Gene Alvo	**Posição**	**Candidato #**	**H3K4me3**	**H3K27ac**	**P300**	**H3K4me1**	**H3K4me3**	**H3K27ac**	**P300**	**H3K4me1**	**ATAC**	**Conservação**	**DNAseI**
PAX9	chr14:37,133,061-37,133,510	PAX9-C1	(+++)	(+++)			(+++)	(+++)	(+++)	(+++)		(++)	(+++)
PAX9	chr14:37,136,120-37,136,787	PAX9-C2	(+++)	(+++)	(+++)					(++)		(++)	(+++)
PAX9	chr14:37,117,044-37,117,546	PAX9-C3	(+++)			(+++)	(+++)			(++)	(++)	(++)	(+++)
PAX9	chr14:37,384,062-37,384,950	PAX9-C4						(++)			(++)	(++)	
PAX9	chr14:37,437,475-37,438,313	PAX9-C5		(+++)								(++)	
PAX9	chr14:37,458,553-37,459,391	PAX9-C6		(+++)		(+++)						(++)	
PAX9	chr14:37,504,432-37,505,550	PAX9-C7		(+++)		(++)						(++)	(++)
PAX9	chr14:37,558,870-37,559,615	PAX9-C8		(+++)						(++)		(+++)	(++)
			SCAP				CNCC						
Gene Alvo	**Posição**	**Candidato #**	**H3K4me3**	**H3K27ac**	**P300**	**H3K4me1**	**H3K4me3**	**H3K27ac**	**P300**	**H3K4me1**	**ATAC**	**Conservação**	**DNAseI**
BMP4	chr14:54,414,039-54,414,751	BMP4-C1		(+++)						(++)		(++)	(+++)
BMP4	chr14:54,414,416-54,415,128	BMP4-C2		(+++)								(++)	(+++)
BMP4	chr14:54,401,944-54,402,471	BMP4-C3		(+++)				(+++)	(+++)		(+++)	(+++)	
BMP4	chr14:54,363,187-54,364,055	BMP4-C4		(+++)		(+++)						(++)	
BMP4	chr14:54,357,203-54,357,975	BMP4-C5		(+++)		(+++)						(++)	(++)
BMP4	chr14:54,465,643-54,466,286	BMP4-C6		(+++)		(+++)						(+++)	(++)
BMP4	chr14:54,468,197-54,468,840	BMP4-C7		(+++)								(++)	(++)
BMP4	chr14:54,469,347-54,469,990	BMP4-C8		(+++)		(+++)						(++)	(++)
BMP4	chr14:54,638,265-54,639,696	BMP4-C9		(+++)								(++)	
BMP4	chr14:54,644,339-54,645,399	BMP4-C10		(+++)								(++)	(++)
BMP4	chr14:54,654,085-54,654,870	BMP4-C11		(+++)				(++)		(++)		(++)	(++)

BMP4	chr14:54,686,814-54,687,504	BMP4-C12	(+++)	(+++)	(++)		(++)	(+++)	(+++)	(++)	(+++)	(+++)	(+++)
			SCAP				CNCC						
Gene Alvo	**Posição**	**Candidato #**	**H3K4me3**	**H3K27ac**	**P300**	**H3K4me1**	**H3K4me3**	**H3K27ac**	**P300**	**H3K4me1**	**ATAC**	**Conservação**	**DNAseI**
MSX1	chr4:4,862,698-4,863,072	MSX1-C1	(+++)	(+++)	(+++)	(+++)	(+++)	(+++)				(++)	(+++)
MSX1	chr4:4,859,560-4,860,685	MSX1-C2	(+++)	(+++)						(++)		(++)	(++)
MSX1	chr4:4,858,301-4,859,050	MSX1-C3	(+++)	(+++)			(++)	(++)		(++)	(++)	(++)	(+++)
MSX1	chr4:4,857,117-4,857,866	MSX1-C4	(+++)	(+++)		(++)		(+++)		(++)		(++)	(++)
MSX1	chr4:4,856,174-4,857,006	MSX1-C5	(++)	(+++)		(++)		(+++)		(++)	(++)	(++)	(++)
MSX1	chr4:4,821,954-4,822,508	MSX1-C6		(++)									(++)
MSX1	chr4:4,764,871-4,765,692	MSX1-C7		(+++)		(+++)		(+++)	(+++)	(++)	(+++)	(++)	(+++)
MSX1	chr4:4,761,375-4,761,983	MSX1-C8		(+++)				(+++)	(++)	(++)			(++)
MSX1	chr4:4,760,774-4,761,374	MSX1-C9		(+++)				(+++)		(+++)		(++)	(++)
MSX1	chr4:4,595,423-4,596,081	MSX1-C10		(+++)				(+++)	(++)	(++)	(++)	(++)	(++)
MSX1	chr4:4,592,027-4,592,685	MSX1-C11		(+++)		(+++)		(+++)	(+++)	(++)	(++)		(++)
MSX1	chr4:4,576,519-4,577,983	MSX1-C12	(+++)	(+++)						(++)	(++)		(++)
MSX1	chr4:4,576,896-4,577,546	MSX1-C13	(+++)	(+++)						(++)	(++)		(+++)
			SCAP				CNCC						
Gene Alvo	**Posição**	**Candidato #**	**H3K4me3**	**H3K27ac**	**P300**	**H3K4me1**	**H3K4me3**	**H3K27ac**	**P300**	**H3K4me1**	**ATAC**	**Conservação**	**DNAseI**
MSX2	chr5:174,162,831-174,163,524	MSX2-C1	(+++)	(+++)			(++)			(++)			(++)
MSX2	chr5:173,959,399-173,960,023	MSX2-C2		(+++)		(+++)				(++)			(++)
MSX2	chr5:173,960,628-173,961,252	MSX2-C3		(+++)		(++)							(++)
MSX2	chr5:173,811,533-173,812,157	MSX2-C4		(+++)		(+++)		(++)	(++)	(++)		(+++)	(++)
MSX2	chr5:173,754,175-173,754,799	MSX2-C5		(++)					(++)	(++)			(++)
			SCAP				CNCC						
Gene Alvo	**Posição**	**Candidato #**	**H3K4me3**	**H3K27ac**	**P300**	**H3K4me1**	**H3K4me3**	**H3K27ac**	**P300**	**H3K4me1**	**ATAC**	**Conservação**	**DNAseI**
DLX1/DLX2	chr2:172,958,325-172,958,945	DLX1/2-C1	(+++)	(+++)			(+++)			(+++)		(+++)	(+++)
DLX1/DLX2	chr2:172,956,404-172,957,093	DLX1/2-C2	(+++)	(+++)		(+++)			(++)	(++)		(+++)	(++)

DLX1/ DLX2	chr2:173,095,158-173,095,925	DLX1/2-C3		(+++)	(+++)	(+++)							(++)
DLX1/ DLX2	chr2:173,099,592-173,100,359	DLX1/2-C4	(+++)	(+++)			(+++)			(++)		(++)	(+++)
DLX1/ DLX2	chr2:172,016,510-172,017,140	DLX1/2-C5	(+++)	(+++)			(+++)	(+++)			(++)	(++)	(+++)
			SCAP				**CNCC**						
Gene Alvo	**Posição**	**Candidato #**	**H3K4 me3**	**H3K27 ac**	**P300**	**H3K4me 1**	**H3K4m e3**	**H3K27 ac**	**P300**	**H3K4 me1**	**ATA C**	**Conser vação**	**DNAse I**
AXIN2	chr17:63,511,907-63,512,682	AXIN2-C1		(+++)		(+++)						(+++)	(++)
AXIN2	chr17:63,580,280-63,581,061	AXIN2-C2		(+++)		(+++)						(++)	(++)
AXIN2	chr17:63,584,141-63,584,922	AXIN2-C3		(+++)									(++)
AXIN2	chr17:63,779,979-63,780,621	AXIN2-C4		(+++)		(+++)							(++)
AXIN2	chr17:63,787,677-63,788,641	AXIN2-C5		(+++)								(+++)	
AXIN2	chr17:63,690,846-63,691,488	AXIN2-C6		(+++)								(++)	
AXIN2	chr17:63,693,536-63,694,606	AXIN2-C7		(++)						(++)			
			SCAP				**CNCC**						
Gene Alvo	**Posição**	**Candidato #**	**H3K4 me3**	**H3K27 ac**	**P300**	**H3K4me 1**	**H3K4m e3**	**H3K27 ac**	**P300**	**H3K4 me1**	**ATA C**	**Conser vação**	**DNAse I**
INHBA (ACTIVIN A)	chr7:41,613,150-41,613,918	INHBA-C1		(++)								(+++)	(++)
INHBA (ACTIVIN A)	chr7:41,619,113-41,620,266	INHBA-C2		(+++)									(++)
INHBA (ACTIVIN A)	chr7:41,622,888-41,623,656	INHBA-C3		(+++)								(++)	
INHBA (ACTIVIN A)	chr7:41,697,033-41,698,070	INHBA-C4		(++)								(++)	
			SCAP				**CNCC**						
Gene Alvo	**Posição**	**Candidato #**	**H3K4 me3**	**H3K27 ac**	**P300**	**H3K4me 1**	**H3K4m e3**	**H3K27 ac**	**P300**	**H3K4 me1**	**ATA C**	**Conser vação**	**DNAse I**
BARX1	chr9:96,792,561-96,793,300	BARX1-C1	(+++)	(++)			(+++)	(+++)	(++)		(++)		(+++)
BARX1	chr9:96,928,048-96,929,034	BARX1-C2	(++)	(+++)			(++)	(+++)	(+++)		(+++)		(++)
BARX1	chr9:96,902,069-96,902,799	BARX1-C3				(++)		(+++)	(++)	(+++)	(++)		(+++)
			SCAP				**CNCC**						

Gene Alvo	Posição	Candidato #	H3K4 me3	H3K27 ac	P300	H3K4me 1	H3K4m e3	H3K27 ac	P300	H3K4 me1	ATA C	Conser vação	DNAse I
BMP2	chr20:6,561,033-6,561,843	BMP2-C1		(+++)		(++)						(++)	
			SCAP				CNCC						
Gene Alvo	Posição	Candidato #	H3K4 me3	H3K27 ac	P300	H3K4me 1	H3K4m e3	H3K27 ac	P300	H3K4 me1	ATA C	Conser vação	DNAse I
BMP6	chr6:8,180,109-8,180,883	BMP6-C1		(+++)				(+++)	(++)		(++)	(++)	(++)
BMP6	chr6:8,194,346-8,195,120	BMP6-C2							(+++)			(++)	(++)
			SCAP				CNCC						
Gene Alvo	Posição	Candidato #	H3K4 me3	H3K27 ac	P300	H3K4me 1	H3K4m e3	H3K27 ac	P300	H3K4 me1	ATA C	Conser vação	DNAse I
WNT5A	chr3:55,518,967-55,519,741	WNT5A-C1	(+++)	(+++)			(+++)	(+++)	(++)		(++)		(+++)
WNT5A	chr3:55,519,996-55,520,770	WNT5A-C2	(+++)	(+++)			(+++)	(+++)	(++)		(++)	(++)	(+++)
WNT5A	chr3:55,525,310-55,526,170	WNT5A-C3		(+++)		(+++)						(+++)	(++)
WNT5A	chr3:55,558,111-55,559,067	WNT5A-C4		(+++)								(++)	(++)
WNT5A	chr3:55,561,832-55,562,788	WNT5A-C5		(++)								(++)	
WNT5A	chr3:55,613,213-55,614,169	WNT5A-C6		(+++)								(++)	(++)
WNT5A	chr3:55,654,449-55,654,979	WNT5A-C7		(+++)				(++)				(+++)	(+++)
WNT5A	chr3:55,664,629-55,665,415	WNT5A-C8		(+++)				(+++)	(+++)	(+++)	(++)	(++)	
WNT5A	chr3:55,698,226-55,699,012	WNT5A-C9		(+++)								(++)	(++)
			SCAP				CNCC						
Gene Alvo	Posição	Candidato #	H3K4 me3	H3K27 ac	P300	H3K4me 1	H3K4m e3	H3K27 ac	P300	H3K4 me1	ATA C	Conser vação	DNAse I
LEF1	chr4:108204537-108205469	LEF1-C1						(+++)		(++)	(++)	(++)	(++)
LEF1	chr4:108951219-108952004	LEF1-C2		(+++)									(++)
LEF1	chr4:108963957-108965984	LEF1-C3						(+++)		(++)			
LEF1	chr4:109271626-109271983	LEF1-C4						(+++)	(+++)		(+++)	(+++)	(+++)
LEF1	chr4:109450887-109453033	LEF1-C5						(+++)		(++)			(++)
LEF1	chr4:109453786-109456067	LEF1-C6						(+++)		(++)			
LEF1	chr4:109810999-109811929	LEF1-C7		(+++)				(+++)	(++)	(++)	(++)		(++)
LEF1	chr4:109819027-109820108	LEF1-C8						(++)		(++)			

LEF1	chr4:109903375-109905025	LEF1-C9						(+++)		(++)		(++)	(++)
LEF1	chr4:110067440-110068969	LEF1-C10						(+++)		(++)		(++)	
LEF1	chr4:110081210-110082632	LEF1-C11						(+++)		(++)			
LEF1	chr4:108661034-108662308	LEF1-C12		(+++)		(++)							
LEF1	chr4:109317873-109318913	LEF1-C13		(+++)	(++)								(++)
LEF1	chr4:109353032-109356463	LEF1-C14		(+++)		(++)							
LEF1	chr4:109387150-109389437	LEF1-C15		(+++)		(++)		(++)					(++)

ANEXOS

Anexo 1 - Verificação de originalidade e prevenção de plágio

ESTUDO DE ELEMENTOS CIS-REGULATÓRIOS DE GENES ENVOLVIDOS COM O DESENVOLVIMENTO DENTÁRIO

RELATÓRIO DE ORIGINALIDADE

19%	17%	17%	10%
ÍNDICE DE SEMELHANÇA	FONTES DA INTERNET	PUBLICAÇÕES	DOCUMENTOS DOS ALUNOS

FONTES PRIMÁRIAS

1	coek.info Fonte da Internet	2%
2	Www.ijbs.com Fonte da Internet	1%
3	www.frontiersin.org Fonte da Internet	1%
4	www.wjgnet.com Fonte da Internet	1%
5	Alvaro Rada-Iglesias. "Genetic variation within transcriptional regulatory elements and its implications for human disease", Biological Chemistry, 2014 Publicação	1%
6	academic.oup.com Fonte da Internet	1%
7	pubmed.ncbi.nlm.nih.gov Fonte da Internet	1%

ouci.dntb.gov.ua

Anexo 2 - Parecer Consubstanciado do Comitê de Ética em Pesquisa - FOP

UNICAMP - FACULDADE DE ODONTOLOGIA DE PIRACICABA DA UNIVERSIDADE DE CAMPINAS - FOP/UNICAMP

PARECER CONSUBSTANCIADO DO CEP

DADOS DO PROJETO DE PESQUISA

Título da Pesquisa: ESTUDO DE ELEMENTOS REGULATÓRIOS DE GENES ENVOLVIDOS COM O DESENVOLVIMENTO DENTÁRIO

Pesquisador: Marcelo Rocha Marques

Área Temática: Genética Humana:
(Trata-se de pesquisa envolvendo Genética Humana que não necessita de análise ética por parte da CONEP;);
Projetos de pesquisa que envolvam organismos geneticamente modificados (OGM), células-tronco embrionárias e organismos que representem alto risco coletivo, incluindo organismos relacionados a eles, nos âmbitos de: experimentação, construção, cultivo, manipulação, transporte, transferência, importação, exportação, armazenamento, liberação no meio ambiente e descarte;

Versão: 2

CAAE: 47869121.2.0000.5418

Instituição Proponente: Faculdade de Odontologia de Piracicaba - Unicamp

Patrocinador Principal: FUND COORD DE APERFEICOAMENTO DE PESSOAL DE NIVEL SUP

DADOS DO PARECER

Número do Parecer: 4.872.304

Apresentação do Projeto:

O parecer inicial é elaborado com base na transcrição editada do conteúdo do registro do protocolo na Plataforma Brasil e dos arquivos anexados à Plataforma Brasil. Os pareceres de retorno, emendas e notificações são elaborados a partir dos dados e arquivos da última versão apresentada.

A EQUIPE DE PESQUISA citada na capa do projeto de pesquisa inclui MARCELO ROCHA MARQUES (Cirurgião Dentista, Docente na Área de Histologia e Embriologia da FOP/Unicamp, Pesquisador responsável), DOUGLAS VICTORINO ESPOSITO (Graduação em Bioquímica, Mestrando no PPG em Biologia Buco-Dental da FOP/UNICAMP), LARISSA LOPES RODRIGUES (Cirurgiã Dentista, Doutoranda no PPG em Biologia Buco-Dental da FOP/UNICAMP), o que é confirmado na declaração dos pesquisadores e na PB.

UNICAMP - FACULDADE DE ODONTOLOGIA DE PIRACICABA DA UNIVERSIDADE DE CAMPINAS - FOP/UNICAMP

Continuação do Parecer: 4.872.304

Declaração de Manuseio Material Biológico / Biorepositório / Biobanco	61AnexoBiorrepositorio.pdf	09/06/2021 15:40:51	Douglas Victorino Esposito	Aceito
TCLE / Termos de Assentimento / Justificativa de Ausência	4TCLE.pdf	09/06/2021 15:37:32	Douglas Victorino Esposito	Aceito
Folha de Rosto	folhaDeRosto.pdf	09/06/2021 11:19:44	Douglas Victorino Esposito	Aceito
Projeto Detalhado / Brochura Investigador	2Projeto.pdf	08/06/2021 17:00:48	Douglas Victorino Esposito	Aceito
Declaração de Pesquisadores	51DeclaraPesquisadores.pdf	08/06/2021 17:00:42	Douglas Victorino Esposito	Aceito
Declaração de Instituição e Infraestrutura	52DeclaraInstituicao.pdf	08/06/2021 17:00:34	Douglas Victorino Esposito	Aceito

Situação do Parecer:
Aprovado

Necessita Apreciação da CONEP:
Sim

PIRACICABA, 28 de Julho de 2021

Assinado por:
jacks jorge junior
(Coordenador(a))