



UNICAMP

UNIVERSIDADE ESTADUAL DE CAMPINAS
INSTITUTO DE BIOLOGIA

FERNANDA CRISTINA DA VEIGA

ANÁLISE DA REGULAÇÃO GÊNICA E EXPRESSÃO DE DACT1 NO
DESENVOLVIMENTO DA MUSCULATURA ESQUELÉTICA DE
AMNIOTOS

*ANALYSIS OF DACT1 GENE REGULATION AND EXPRESSION IN THE DEVELOPMENT
OF THE AMNIOTES SKELETAL MUSCLE*

CAMPINAS

2020

FERNANDA CRISTINA DA VEIGA

ANÁLISE DA REGULAÇÃO GÊNICA E EXPRESSÃO DE DACT1 NO
DESENVOLVIMENTO DA MUSCULATURA ESQUELÉTICA DE AMNIOTOS

*ANALYSIS OF DACT1 GENE REGULATION AND EXPRESSION IN THE DEVELOPMENT
OF THE AMNIOTES SKELETAL MUSCLE*

Tese apresentada ao Instituto de
Biologia da Universidade Estadual de
Campinas como parte dos requisitos
exigidos para a obtenção do título de
Doutora em Biologia Celular e
Estrutural, na área de Biologia Tecidual.

Thesis presented to the Institute of
Biology of the University of Campinas
in partial fulfillment of the
requirements for the degree of Doctor,
in the area of Tissue Biology.

ORIENTADORA: PROF^A. DR^A. LUCIA ELVIRA ALVARES

ESTE TRABALHO CORRESPONDE À VERSÃO
FINAL DA TESE DEFENDIDA PELA ALUNA
FERNANDA CRISTINA DA VEIGA, E ORIENTADA
PELA PROF^A. DR^A. LUCIA ELVIRA ALVARES.

CAMPINAS

2020

Ficha catalográfica
Universidade Estadual de Campinas
Biblioteca do Instituto de Biologia
Mara Janaina de Oliveira - CRB 8/6972

V533a Veiga, Fernanda Cristina da, 1987-
Análise da regulação gênica e expressão de Dact1 no desenvolvimento da musculatura esquelética de amniotos / Fernanda Cristina da Veiga. – Campinas, SP : [s.n.], 2020.

Orientador: Lúcia Elvira Alvares.
Tese (doutorado) – Universidade Estadual de Campinas, Instituto de Biologia.

1. Genes Dapper. 2. Elementos reguladores de transcrição. 3. Expressão gênica. 4. Regulação da expressão gênica. 5. Músculo esquelético. I. Alvares, Lúcia Elvira, 1968-. II. Universidade Estadual de Campinas. Instituto de Biologia. III. Título.

Informações para Biblioteca Digital

Título em outro idioma: Analysis of Dact1 gene regulation and expression in the development of the amniotes skeletal muscle

Palavras-chave em inglês:

Dapper genes
Regulatory elements, Transcriptional
Gene expression
Gene expression regulation
Skeletal muscle

Área de concentração: Biologia Tecidual

Titulação: Doutora em Biologia Celular e Estrutural

Banca examinadora:

Lúcia Elvira Alvares [Orientador]
Mariana de Freitas Nery
Elisa Maria Costa e Silva de Paiva
Leonardo dos Reis Silveira
Luiz Lehmann Coutinho

Data de defesa: 27-11-2020

Programa de Pós-Graduação: Biologia Celular e Estrutural

Identificação e informações acadêmicas do(a) aluno(a)

- ORCID do autor: <https://orcid.org/0000-0002-3575-8992>
- Currículo Lattes do autor: <http://lattes.cnpq.br/1916032278155051>

COMISSÃO EXAMINADORA

Profa. Dra. Lucia Elvira Alvares (Presidente)

Profa. Dra. Mariana de Freitas Nery

Dra. Elisa Maria Costa e Silva de Paiva

Prof. Dr. Leonardo dos Reis Silveira

Prof. Dr. Luiz Lehmann Coutinho

Os membros da Comissão Examinadora acima assinaram a Ata de Defesa, que se encontra no processo de vida acadêmica do aluno.

A Ata da defesa com as respectivas assinaturas dos membros encontra-se no SIGA/Sistema de Fluxo de Dissertação/Tese e na Secretaria do Programa de Pós-Graduação em Biologia Celular e Estrutural do Instituto de Biologia.

Agradecimentos

Agradeço, primeiramente à minha família pelo grande apoio, compreensão e carinho durante esta jornada e sempre.

Agradeço também meus queridos amigos do Departamento de Bioquímica e Biologia Tecidual, que se tornaram amigos para vida. Agradeço a todos imensamente pela companhia, carinho e respeito dispendidos durante todos esses anos. Em especial, agradeço às colegas do Laboratório de Biologia do Desenvolvimento, Carolina Mantovani e Renata Contriciani, e à Carolina Frandsen, que me ajudou com as ilustrações, por me presentarem com amizade e companheirismo.

À Profa. Dra. Lucia Elvira Alvares, minha orientadora, por ter me recebido no Laboratório de Biologia do Desenvolvimento e por me guiar neste caminho profissional. Sempre com muita delicadeza, respeito e afeto me ensinou muito durante todos esses anos como sua aluna.

Aos membros da banca examinadora, pela disponibilidade e pelas contribuições para o aprimoramento deste trabalho.

Aos Professores do Departamento de Bioquímica e Biologia Tecidual, Dr. Henrique Marques-Souza, Dr. Sílvio Roberto Consonni e Dr. Luis Violin Dias Pereira por compartilharem os laboratórios de Biologia Molecular e de Cultura Celular, o Laboratório de Citoquímica e Imunocitoquímica e o Laboratório de Tecnologia Instrucional em Embriologia e Biologia Tecidual, permitindo que todos os experimentos pudessem ser realizados.

À Thaís, técnica de laboratório, pela dedicação dispendida em manter nosso ambiente de trabalho sempre organizado.

À Universidade Estadual de Campinas e ao Programa de Pós-graduação em Biologia Celular e Estrutural, pela formação profissional proporcionada.

Aos funcionários da secretaria de pós-graduação do Instituto de Biologia, pelo suporte e orientação dados durante todo curso.

O presente trabalho foi realizado com apoio da Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brasil (CAPES) - Código de Financiamento 001 e do Conselho Nacional de Desenvolvimento Científico e Tecnológico – CNPq (Bolsa de Doutorado, processo número 141441/2018-6).

O gene *Dact1* é um importante membro de uma família gênica que codifica proteínas multifuncionais, as quais desempenham papéis no desenvolvimento embrionário e no adulto. Sua função mais conhecida é de modulador da via de sinalização Wnt/ β -catenina, que participa de múltiplos processos biológicos.

Neste contexto, esta tese visou caracterizar a expressão gênica e proteica de *Dact1* durante a miogênese esquelética de mioblastos murinos (C2C12) e também investigar *in silico* a modulação de *Dact1* em miopatias no humano. Nossos achados revelaram ampla expressão gênica e proteica de *Dact1* durante a miogênese em células C2C12, sugerindo que esta molécula possa atuar como um modulador bivalente da via Wnt/ β -catenina. Também foi demonstrado que a transcrição de *Dact1* é modificada em doenças humanas, como as distrofias, o que prevê um possível papel fisiopatológico de *Dact1* em doenças musculares.

Esta tese também buscou ampliar a compreensão da regulação transcricional de *Dact1*, identificando e caracterizando CNEs (*Conserved Non-coding Elements*), que potencialmente atuam como elementos cis-reguladores deste gene. Análises de genômica comparativa revelaram a presença de oito CNEs no intervalo genômico *KIAA0586-Dact1-Daam1* conservados entre humano, camundongo e galinha. Seis destes CNEs correspondem a elementos reguladores do *Ensembl Regulatory Build* no genoma humano. A análise de dados de atividade e repressão destes elementos no *Ensembl* permitiu verificar que há uma correlação entre os padrões de expressão de *Dact1* no ser humano com os perfis de atividade/repressão destes elementos em diferentes contextos. A maioria destes CNEs apresenta sítios de ligação para fatores de transcrição (TFBSs) conservados evolutivamente, os quais geram redes de interação proteína-proteína de alta significância, possibilitando mapear interações físicas e/ou funcionais entre estas proteínas e associá-las à atividade de *Dact1* em diversos contextos celulares e patológicos no humano.

Esta análise sistemática apontou para o CNE4 como possível *enhancer* de *Dact1* no contexto miogênico. A interação entre o CNE4 e o promotor de *Dact1* foi confirmada por dados de *Hi-C* em músculo humano, disponíveis no *Gene Expression Omnibus* (GEO). Em adição, a

atividade de *enhancer* do CNE4 foi validada *in vitro* por ensaio com o gene repórter *Luciferase* em células C2C12.

O detalhamento da conservação evolutiva do CNE4 foi realizado comparando a sequência humana com aquelas correspondentes em organismos-modelo, animais domésticos e também de interesse econômico. A análise de TFBSs e de interação proteína-proteína do CNE4 para primatas e bovídeos revelou que as redes de fatores de transcrição geradas possivelmente regulam atividades distintas de *Dact1* nestes grupos, indicando divergência funcional.

Em conjunto, os resultados apresentados nesta tese abrem novos caminhos para investigações futuras sobre a função de *Dact1* na miogênese esquelética e miopatias humanas, bem como apresenta um panorama inédito sobre a regulação transcricional de *Dact1* em diferentes contextos, abrindo perspectivas para estudos funcionais e avaliação de SNPs (*Single Nucleotide Polymorphisms*) nestes elementos cis-reguladores que estejam associados a doenças ou processos evolutivos.

Dact1 is an important member of a gene family that encodes multifunctional proteins, which play roles in embryonic and adult development. *Dact1* is well known as modulator of Wnt/ β -catenin signaling pathway, which participates in multiple biological processes.

In this context, this thesis aimed to characterize *Dact1* gene and protein expression during the skeletal myogenesis of murine myoblasts (C2C12) and also to investigate *in silico* *Dact1* modulation in human myopathies. Our findings revealed wide *Dact1* gene and protein expression during myogenesis in C2C12 cells, suggesting that this molecule may act as a bivalent modulator of Wnt/ β -catenin pathway. It has also been shown that *Dact1* transcription is modified in human diseases, such as dystrophies, which predicts a possible pathophysiological role of *Dact1* in muscle diseases.

This thesis also sought to broaden the understanding of the transcriptional regulation of *Dact1*, by identifying and characterizing CNEs (Conserved Non-coding Elements), which potentially act as cis-regulatory elements of this gene. Comparative genomic analyzes revealed the existence of eight CNEs in the *KIAA0586-Dact1-Daam1* genomic range conserved between human, mouse, and chicken. Six of these CNEs correspond to regulatory elements of the Ensembl Regulatory Build in human genome. The analysis of activity and repression data of these elements in Ensembl allowed us to verify that there is a correlation between *Dact1* expression patterns in humans and the activity/repression profiles of these elements in different contexts. Most of these CNEs have transcription factors binding sites (TFBSs) evolutionarily conserved, which generate protein-protein interaction networks with high significance, making it possible to map physical and/or functional interactions between these proteins and associate them with *Dact1* activity in several cellular and pathological contexts in human.

This systematic analysis pointed to CNE4 as a possible *Dact1* enhancer in the myogenic context. The interaction between CNE4 and *Dact1* promoter was confirmed by Hi-C data in human muscle, available on Gene Expression Omnibus (GEO). In addition, CNE4's enhancer activity was validated *in vitro* by *Luciferase* reporter gene assay on C2C12 cells.

Further investigation on the evolutionary conservation of CNE4 were carried out by comparing the human sequence with those corresponding to model organisms, domestic animals and also those with economic interest. The TFBSs and CNE4 protein-protein interaction analysis of primates and bovines revealed that the transcription factors networks obtained possibly regulate distinct *Dact1* activities within these groups, indicating functional divergence.

Taken together, the results presented in this thesis open new perspectives for future investigations on *Dact1* role in skeletal myogenesis and human myopathies, as well as shed a light on unprecedented scenario of *Dact1* transcriptional regulation in different contexts, opening possibilities for functional assays and SNPs evaluation (Single Nucleotide Polymorphisms) in these cis-regulatory elements that can be associated with diseases or evolutionary processes.

Figura 1: Principais tipos de elementos cis-reguladores.

Figura 2: Diagrama esquemático das principais técnicas de detecção da conformação da cromatina.

Figura 3: Arquitetura dos blocos genômicos reguladores.

Figura 4: Representação esquemática dos principais órgãos do corpo humano associados às doenças com significativa alteração na expressão de *Dact1*.

Capítulo 2

Figura 1: Vetor pGL2-TK-Luc utilizado para clonagem.

Figura 2: Diagrama resumido da metodologia empregada no capítulo 2.

Figura 3: Elementos não codificadores conservados (CNEs) em torno do locus de *Dact1* humano.

Figura 4: Gráfico de expressão do gene *Dact1* em amostras provenientes de órgãos e tecidos de humanos adultos saudáveis (*GTEx Portal*).

Figura 5: Diagrama dos picos de conservação e dos TFBSs presentes nas regiões correspondentes aos CNEs.

Figura 6: Rede de interações proteína-proteína gerada pelo *STRING* para o CNE1

Figura 7: Rede de interações proteína-proteína gerada pelo *STRING* para o CNE2.

Figura 8: Rede de interações proteína-proteína gerada pelo *STRING* para o CNE4.

Figura 9: CNE4 *Virtual 4C* gerado pelo dado de *Hi-C* proveniente do músculo Psoas.

Figura 10: Gráficos da atividade do gene repórter da *Luciferase* dirigido pelo CNE4 murino.

Figura 11: Comparação das sequências ortólogas do CNE4 entre as 12 espécies de animais estudadas e a sequência humana.

Figura 12: Análise dos sítios de ligação para TFs conservados do CNE4.

Figura 13: Análise das redes de fatores do CNE4 humano *vs.* chimpanzé (*Primates*).

Figura 14: Análise das redes de fatores do CNE4 vaca *vs.* ovelha *vs.* cabra (*Bovidae*).

Tabela 1: Padrão de expressão de *Dact1* durante o desenvolvimento embrionário.

Capítulo 2

Tabela 1: Localizações genômicas dos CNEs de acordo com o banco de dados do *Ensembl* e elementos reguladores humanos correspondentes que estão próximos ou sobrepostos a cada CNE.

Tabela 2: Elementos reguladores descritos que têm como alvo a região genômica em humanos que contém os CNEs identificados.

Tabela 3: Linhagens celulares, tecidos e órgãos com marcadores de atividade identificados no banco de dados *Ensembl Regulatory Build*.

Tabela 4: Linhagens celulares, tecidos e órgãos com marcadores de repressão identificados no banco de dados *Ensembl Regulatory Build*.

Tabela 5: TFBSs conservados entre humano, camundongo e galinha, identificados pelo *MultiTF*.

Tabela 6: Localização genômica das sequências correspondentes ao CNE4 identificadas em 12 animais distintos e no humano.

LISTA DE ABREVIATURAS E SIGLAS

3C: *Chromosome Conformation Capture*

4C: *Chromosome Conformation Capture-on-Chip*

5C: *Chromosome Conformation Capture Carbon Copy*

ABD-B: *Hox abdominal-B*

APC: *Adenomatous polyposis coli*

AQM: *Acute Quadriplegic Myopathy*

bHLHZ: *basic Helix-Loop-Helix Leucine Zipper*

BLV: *Bovine Leukemia Virus*

ChIA-PET: *Chromatin Interaction Analysis by Paired-End Tag Sequencing*

CK1: *Casein Kinase 1*

CLP: *Common Lymphoid Progenitor*

CNE: *Conserved Non-coding Element*

CTCF: *CCCTC-Binding Factor*

Daam1: Dishevelled Associated Activator of Morphogenesis 1

Dact: Dapper antagonist of catenin (gene)

Dact: Dapper antagonist of catenin (proteína)

Dapper: Dishevelled-associated antagonist of β -catenin

DBF4: *DumbBell Forming-4*

DMD: *Duchenne Muscular Dystrophy*

DNA: *Ácido Desoxirribonucleico*

DRD: *Dystrophin-related Domain*

Dvl: *Dishevelled*

ENCODE: *Encyclopedia of DNA Elements*

ESCC: *Esophageal Squamous Cell Carcinoma*

FANTOM: *Functional Annotation of the Mouse/Mammalian Genome*

FOXOs: *Forkhead box O*

FZD: *Frizzled*

GEO: *Gene Expression Omnibus*

GIST: *Gastrointestinal Stromal Tumor*

GRB: *Genomic Regulatory Block*

GSK-3 β : *Glycogen Synthase Kinase-3 β*

HCC: *Hepatocellular Carcinoma*

HDAC1: *Histone Deacetylase 1*

HNF4: *Hepatocyte Nuclear Factor 4*

HSC: *Hematopoietic Stem Cells*

HTLV: *Human T-Lymphotropic Virus*

IDRs: *Intrinsically Disordered Regions*

KEGG: *Kyoto Encyclopedia of Genes and Genomes*

KSHV: *Kaposi's Sarcoma-associated Herpesvirus*

LEF/TCF: *Lymphoid Enhancing Factor/T Cell Factor*

LGMD-2: *Limb-Girdle Muscular Dystrophy 2*

MEF2: *Myocyte Enhancer Factor 2*

MRFs: *Myogenic Regulatory Factors*

MYOD: *Myogenic Differentiation Factor*

MYOG: *Myogenin*

NES: *Nuclear Export Signal*

NK: *Natural Killer*

NLS: *Nuclear Localization Sequence*

NPC: *Neural Progenitor Cells*

NR: *Nuclear Receptor*

NSCLC: *Nonsmall Cell Lung Cancer*

PCP: *Planar Cell Polarity*

PDZ: *Post synaptic density-95/Discs large/Zonula occludens-1*

PET: *Pair-Ended Tag*

PGC-1 α : *Peroxisome proliferator-activated receptor Gamma Coactivator - 1 α*

PKA: *Protein Kinase A*

PKC: *Protein Kinase C*

PPAR: *Peroxisome Proliferator-activated Receptor*

PPI: *Protein-protein interaction*

PSM: *Presomitic Mesoderm*

qPCR: *quantitative Polimerase Chain Reaction*

RMS: *Rhabdomyosarcoma*

RNA: *Ácido Ribonucleico*

RTN1: *Reticulon 1*

sFRP: *Secreted Frizzled Related Protein*

SHP: *Small Heterodimer Partner*

SIRT1: *Sirtuin 1*

SNC: *Sistema Nervoso Central*

SNP: *Single Nucleotide Polymorphism*

SREBP1: *Sterol Regulatory Element-binding Protein 1*

STLV: *Simian T-Lymphotropic Virus*

TBP: *TATA-Binding Protein*

TF: *Transcription Factor*

TFBS: *Transcription Factor Binding Site*

TGF- β : *Transforming Growth Factor- β*

TPM: *Transcripts Per Million*

TSS: *Transcription Start Site*

Vangl2: *Van Gogh-like 2*

Wnt: *Wingless-type MMTV integration site*

ZIP4: *Zinc Transporter 4*

ZPA: *Zone of Polarizing Activity*

RESUMO GRÁFICO.....	17
1. INTRODUÇÃO GERAL.....	18
2. REVISÃO DE LITERATURA.....	21
3. OBJETIVOS.....	41
4. CAPÍTULO 1 – CARACTERIZAÇÃO DA EXPRESSÃO GÊNICA E PROTEICA DE DACT1	43
4.1 INTRODUÇÃO	43
4.2 RESULTADOS	46
4.3 DISCUSSÃO.....	85
4.4 CONCLUSÕES	88
5. CAPÍTULO 2 – IDENTIFICAÇÃO E CARACTERIZAÇÃO DE ELEMENTOS CIS-REGULADORES	89
5.1 INTRODUÇÃO	89
5.2 MATERIAIS E MÉTODOS	91
5.3 RESULTADOS	99
5.4 DISCUSSÃO.....	136
5.5 CONCLUSÕES	150
6. CONCLUSÕES GERAIS E PERSPECTIVAS SOBRE AS APLICAÇÕES DO ESTUDO.....	152
7. REFERÊNCIAS BIBLIOGRÁFICAS	154
8. APÊNDICES.....	172
8.1 Apêndice 1	172
8.2 Apêndice 2	174
8.3 Apêndice 3	180
8.4 Apêndice 4	193
8.5 Apêndice 5	194
8.7 Apêndice 7	200
8.8 Apêndice 8	201
9. ANEXOS.....	202
9.1 Artigos publicados em colaboração	202
9.2 Capítulo de livro publicado	203
9.3 Declaração de bioética e biossegurança.....	204
9.4 Declaração de direitos autorais	205

RESUMO GRÁFICO

Introdução

- 1 Os genes *Dact* codificam proteínas adaptadoras multifuncionais. A função biológica mais bem caracterizada de *Dact1* é a sua capacidade de modular a sinalização Wnt/ β -catenina através da ligação com Dvl e outras quinases.
- 2 *Dact1* é um gene essencial para o desenvolvimento dos vertebrados. Ele atua na adipogênese e é expresso em células precursoras do músculo esquelético. Em humanos está associado a vários tipos de câncer, como gastrointestinal, de ovário e de fígado.
- 3 Nesta tese, inicialmente, caracterizamos a expressão gênica e proteica de *Dact1* em C2C12 e no músculo fetal de camundongo. Além disso, investigamos sua expressão gênica em várias miopatias humanas por análises de bioinformática.
- 4 Em adição, realizamos uma análise sistemática do intervalo gênico *KIAA0586-Dact1-Daam1*, utilizando ferramentas de bioinformática, para identificar e caracterizar possíveis elementos cis-reguladores de *Dact1* em diversos contextos, incluindo o muscular.

Objetivos

- A
 - Avaliar a expressão gênica e a distribuição da proteína *Dact1* em células da linhagem miogênica (C2C12).
 - Investigar a modulação de *Dact1* em miopatias humanas.
- B
 - Ampliar a compreensão da regulação gênica de *Dact1*, identificando e caracterizando CNEs no seu locus, que possam atuar como seus possíveis reguladores.
 - Testar o papel regulador de um CNE específico em dois contextos celulares onde *Dact1* atua (adipogênico e miogênico).

Metodologia



Conclusões

- *Dact1/Dact1* é amplamente expresso durante toda miogênese em cultura de células C2C12 e deve atuar como um modulador bivalente da via Wnt/ β -catenina.
- *Dact1* tem sua expressão alterada em miopatias humanas.
- Existem oito CNEs localizados no intervalo genômico *KIAA0586-Dact1-Daam1* e são potenciais cis-reguladores de *Dact1*.
- Os CNEs apontam para modulação de *Dact1* na miogênese, adipogênese e em diversos tipos de câncer.

1. INTRODUÇÃO GERAL

Os genes da família *Dact* (*Dapper/Frodo*) codificam proteínas adaptadoras multifuncionais, que desempenham múltiplas funções no desenvolvimento embrionário e na manutenção da homeostase na fase pós-natal. Genes *Dact* foram identificados no genoma de mamíferos, bem como na galinha, em *Xenopus* e no peixe-zebra, sendo alvo de vários estudos ao longo das últimas duas décadas, dada sua capacidade de modular e possivelmente de integrar as cascatas de sinalização Wnt e TGF- β . Esta habilidade para modular vias de sinalização específicas depende principalmente da interação física entre as proteínas Dact com moléculas-chave destas vias de sinalização. No genoma dos mamíferos são encontrados três membros desta família – *Dact1*, *Dact2* e *Dact3*, sendo *Dact1* o membro mais estudado e o gene protagonista das pesquisas incluídas nesta tese.

A função biológica mais bem caracterizada de *Dact1* é a sua capacidade de modular a sinalização Wnt/ β -catenina através de interações diretas com Dishevelled (Dvl), uma molécula chave no controle de diferentes vias de sinalização Wnt (Cheyette et al., 2002; Gloy et al., 2002). Além de interagir com Dvl, *Dact1* é capaz de ligar-se com várias quinases, como PKA, PKC e caseína 1 δ/ϵ (CK1 δ/ϵ). Outra característica de seu modo de ação é a regulação da transcrição gênica pela ligação direta aos fatores de transcrição LEF/TCF bem como às desacetilases de histonas (HDAC) (Gao et al., 2008; Hikasa and Sokol, 2004). *Dact1* também interage com a quinase regulatória de divisão celular DBF4 (Brott and Sokol, 2005a) e, por fim, é capaz de se ligar e regular os níveis de Vangl2, uma proteína transmembranar que participa da via Wnt/PCP, regulando também esta via (Suriben et al., 2009).

Algumas evidências apontam para uma possível participação de *Dact1* na miogênese esquelética de vertebrados. Por exemplo, durante a somitogênese, *Dact1* é expresso na placa segmentar mesodérmica de embriões de camundongo, em fase com a expressão do inibidor de Wnt Axin2 (Suriben et al., 2006). Esta expressão ocorre imediatamente antes da ativação dos fatores PAX3 e PAX7 nos somitos recém-formados. Além disso, *Dact1* é expresso no dermomiótomo central de embriões de galinha (Alvares et al., 2009), uma região que é conhecida por gerar uma população de progenitores musculares que dá origem aos mioblastos

fetais, bem como às células satélites adultas (Gros et al., 2005). Em adição ao seu possível envolvimento na miogênese esquelética, *Dact1* atua como um importante gene pró-adipogênico, justamente modulando a via de sinalização molecular Wnt/ β -catenina (Lagathu et al., 2009). Por ser um modulador desta via, *Dact1* favorece a diferenciação dos pré-adipócitos em adipócitos maduros, mediante estímulos nutricionais. Além disto, um acúmulo crescente de informações revela que desregulações na atividade de *Dact1*/*Dact1*, causadas por modificações epigenéticas ou mutações estão associadas a tumores malignos de fígado, de pulmão, de ovário, tumores gastrointestinais, esôfago e cólon (Astolfi et al., 2010; Li et al., 2017b; Yang et al., 2010; Yau et al., 2005).

Embora o papel de *Dact1* venha sendo estudado, ainda há muito a ser explorado sobre suas funções, principalmente porque as vias de sinalização Wnt, reguladas por *Dact1*, são essenciais para múltiplos processos do desenvolvimento, bem como para a homeostase pós-natal, estando intimamente associadas à progressão de doenças humanas (Logan and Nusse, 2004; Moon et al., 2004; Wang and Wynshaw-Boris, 2004). Dado que são escassas as pesquisas sobre o papel de *Dact1* na miogênese esquelética de vertebrados e, considerando que são ainda mais raros os estudos sobre a regulação transcricional deste gene, as pesquisas realizadas nesta tese tiveram como foco estes dois aspectos relacionados ao gene *Dact1*. Dentro destes contextos, os ensaios laboratoriais e análises de bioinformática realizados ao longo do desenvolvimento desta tese possibilitaram:

a) caracterizar a expressão gênica e distribuição da proteína *Dact1* em células da linhagem miogênica de camundongos (C2C12), dado que em galinha essa proteína parece ter papel importante na miogênese e sarcomerogênese durante a formação dos músculos esqueléticos, um papel de *Dact1* não caracterizado previamente.

b) ampliar a compreensão da regulação gênica de *Dact1*, identificando e caracterizando sequências de DNA conservadas ao longo do tempo evolutivo (do inglês, *Conserved Non-coding Elements*, CNEs) no locus deste gene, que podem atuar como elementos cis-reguladores da transcrição.

c) identificar um *enhancer* de *Dact1*, o qual modula a atividade deste gene na miogênese esquelética.

Esta tese foi organizada em dois capítulos, que serão apresentados a seguir. O capítulo 1, intitulado “*Caracterização da Expressão Gênica e Proteica de Dact1*”, contém o manuscrito “*Dact1 is expressed during chicken and mouse skeletal myogenesis and modulated in human muscle diseases*”. Neste manuscrito, constam análises de expressão gênica e proteica de *Dact1* em células mio gênicas da linhagem C2C12 durante o processo de diferenciação destas células, bem como análises *in silico* para identificar a modulação de *Dact1* em doenças musculares no humano. Estes ensaios e análises foram realizados ao longo do desenvolvimento desta tese e integrados ao manuscrito, que conta também com outros autores. Como relatado no manuscrito, a ampla expressão de *Dact1*/Dact1 durante a miogênese sugere que ele possa atuar como um modulador bivalente (inibidor e ativador) da via Wnt/ β -catenina e que possui um possível papel fisiopatológico em distúrbios musculares. O capítulo 2, intitulado “*Identificação e Caracterização de Elementos Cis-reguladores*”, apresenta a identificação e a caracterização de CNEs no locus gênico de *Dact1* de vertebrados amniotos, que possivelmente atuam como cis-reguladores deste gene. Em adição, é apresentada a validação funcional de um dos CNEs com potencial regulatório em linhagens celulares. Finalmente, nesta tese também serão discutidas as perspectivas futuras para aplicações da pesquisa e possíveis desdobramentos que este trabalho pode gerar.

2. REVISÃO DE LITERATURA

2.1 *A regulação da expressão gênica para o desenvolvimento e homeostase pós-natal*

O desenvolvimento embrionário e a manutenção da homeostase pós-natal envolvem complexos mecanismos de controle da expressão gênica, que garantem a ativação dos genes nos momentos e locais corretos, bem como em níveis adequados, em um processo que depende da ação de elementos reguladores da transcrição gênica. Estes elementos reguladores incluem segmentos de DNA específicos, denominados elementos cis-reguladores, bem como proteínas especializadas, categorizadas como fatores trans-reguladores. As interações entre os elementos cis-reguladores (DNA) com os elementos trans-reguladores (proteínas que agem como fatores de transcrição) são célula-específicos, levando à ativação de genes responsáveis pela construção do corpo do indivíduo e promovendo respostas à presença de sinais externos, por exemplo, hormônios (Sakabe et al., 2012), assim coordenando complexas cascatas de sinalização molecular que regem o desenvolvimento dos organismos.

Elementos cis-reguladores da transcrição podem estar localizados *upstream*, *downstream* ou mesmo dentro das regiões codificadoras de um gene. Estes elementos reguladores podem atuar como promotores, *enhancers*, silenciadores e isoladores – os quais são necessários para controlar onde e quando um gene em particular é transcrito (Gilbert, 2010). Promotores são segmentos de DNA onde a RNA polimerase e fatores de transcrição se ligam para iniciar a transcrição, promovendo níveis basais de expressão do gene. Por sua vez, *enhancers* são sequências não codificadoras no DNA que controlam a eficiência e a taxa de transcrição de um promotor específico. Os *enhancers* são elementos definidos pela sua capacidade de ativação da transcrição de um ou mais genes, independentemente da sua distância e orientação em relação aos promotores destes (Bulger and Groudine, 2010). Assim, eles podem ativar promotores presentes no mesmo cromossomo (i.e., ligados em cis) e, por isso, são chamados de elementos cis-reguladores. Estes elementos geralmente ativam a transcrição remodelando a cromatina a fim de expor a região promotora, facilitando a ligação desta à RNA polimerase por

meio da estabilização dos fatores associados à TBP¹ (Gilbert, 2010). A ligação dos remodeladores de cromatina dependem de grupos químicos presentes nos nucleossomos – as histonas, proteínas constitutivas dos nucleossomos. As histonas podem ser dinamicamente modificadas, isto é, acetiladas, metiladas ou fosforiladas em diferentes resíduos. A maquinaria celular reconhece o código de elementos de DNA baseada na combinação dessas modificações (Jenuwein and Allis, 2001) e cada uma destas combinações resulta em um estado de ativação ou inativação daquela região. Outro tipo de modulador da transcrição gênica são os silenciadores, que compreendem as regiões do DNA onde se ligam fatores de transcrição repressores da transcrição de um gene, impedindo o acesso de fatores ativadores à região do promotor por meio de mudanças na conformação da cromatina através do acionamento de uma rede de fatores que levam ao estado de heterocromatina (Raab and Kamakaka, 2010). Em alguns casos, elementos repressores agem bloqueando a ligação de um fator de transcrição ativador. Em outros casos, os repressores competem pelo mesmo sítio de ligação (Maston et al., 2006). Há, ainda, uma terceira categoria de elementos cis-reguladores, os isoladores. Estes elementos são complexos formados pela interação DNA-proteína que são capazes de bloquear a interação entre um *enhancer* e um promotor e também podem agir como barreiras contra os efeitos silenciadores do estado de heterocromatina. Em vertebrados, os isoladores de *enhancers* atuam, principalmente, por meio da interação com CTCF², uma proteína multifuncional envolvida na regulação gênica (Raab and Kamakaka, 2010). Os principais tipos de elementos cis-reguladores estão esquematizados na Figura 1.

No que tange à importância desses elementos cis-reguladores durante o desenvolvimento embrionário, é importante destacar que a maioria das proteínas reguladoras de padrões corporais estão em regiões que abrigam múltiplos elementos cis-reguladores com sítios de ligação para diversos fatores de transcrição e, assim, controlam a expressão gênica dentro de um domínio espacial específico durante o desenvolvimento. A descoberta desta organização modular dos elementos cis-reguladores afetou diretamente os estudos sobre

¹ do inglês, TATA-*Binding Protein*

² do inglês, CCCTC-*Binding Factor*

genética evolutiva, pois pôde propor um mecanismo para justificar como mudanças seletivas da expressão gênica e morfologia poderiam evoluir de maneiras distintas em diferentes partes do corpo do animal (Carroll, 1995; Prud'homme et al., 2007).

Embora estes elementos cis-reguladores venham sendo estudados há décadas, foi somente com o advento da técnica experimental de *ChIP*, ferramenta com a qual é possível localizar experimentalmente sítios no DNA onde proteínas específicas se ligam, tornando possível a identificação de milhares de elementos reguladores da transcrição putativos.

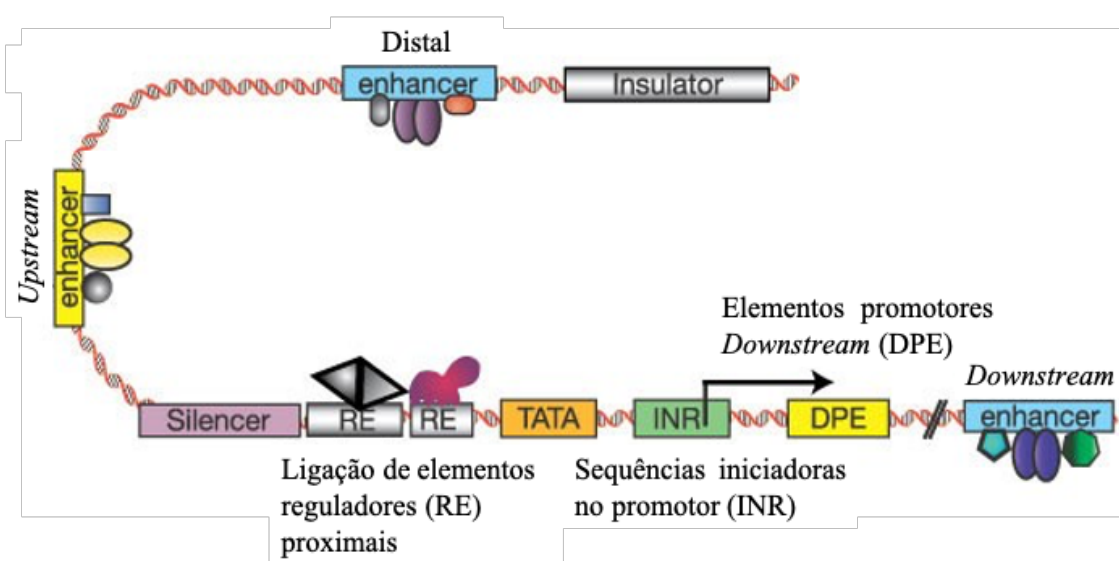


Figura 1. Principais tipos de elementos cis-reguladores. Organização de múltiplos módulos intensificadores (*enhancers*) intercalados com elementos silenciadores e isoladores que podem estar localizados *upstream* ou *downstream* de um promotor gênico (*TATA box* (*TATA*), sequências iniciadoras (*INR*) e elementos promotores *downstream* (*DPE*)). Adaptado de Levine and Tjian, 2003.

2.2 A arquitetura das interações na cromatina

Os cromossomos são algumas das estruturas mais complexas das células. A composição da cromatina é altamente diversificada ao longo de seu comprimento e apresenta uma conformação espacial bastante característica: dobras em três dimensões ao longo da sua extensão. Grandes esforços foram e ainda são dedicados ao mapeamento da estrutura da cromatina, analisando a presença de proteínas associadas ao DNA e suas modificações ao longo dos cromossomos. Tais estudos permitem a identificação de localizações genômicas de genes e

elementos reguladores que são ativos em um determinado tipo de célula e, com isso, possibilitam a identificação de muitos elementos funcionais do genoma humano, bem como do genoma de diversos organismos-modelo (Consortium, 2012; Dekker et al., 2013; Gerstein et al., 2010; Nègre et al., 2011).

Na última década foram desenvolvidas diversas abordagens moleculares para estudar a conformação cromossômica tridimensional com maior resolução e especificidade. Baseadas na captura da conformação do cromossomo (3C³), estas principais abordagens são: 3C, 4C, 5C, *Hi-C*, *ChIP-Loop* e *ChIA-PET*. Esses métodos permitem determinar a frequência com que qualquer par de *loci* no genoma está em proximidade suficiente (provavelmente na faixa de 10 a 100 nm) para formarem *crosslinks* entre si (Dekker et al., 2013).

Os passos iniciais das técnicas de 3C, 4C e 5C são semelhantes: primeiramente há a fixação das cromatinas em proximidade espacial por formaldeído, digestão dos *crosslinks* de cromatina com enzima de restrição, ligação da molécula de DNA híbrida e, finalmente, a reversão da ligação da molécula de DNA quimérica. Na técnica de 3C, são utilizados *primers* específicos para o locus de interesse para que, por qPCR⁴, seja detectado o par de interação. Após uma segunda digestão do produto de ligação, a técnica 4C⁵ gera um perfil de interação genômica em torno de um único locus por meio de PCR inversa, que permite que a região de interesse seja utilizada para amplificar todas as sequências desconhecidas que se ligam à ela, e são identificadas por sequenciamento ou *microarray*. Com a técnica de 5C⁶ é possível identificar múltiplos pares de interações com *primers* universais após a ligação da sequência oligo que são reconhecidos por estes *primers* e, então, esses pequenos fragmentos são identificados por sequenciamento ou *microarray* (Figura 2). Finalmente, a técnica de *Hi-C*, chamada de “todos vs. todos”, revela imparcialmente interações de quaisquer dois *loci* em todo o genoma. O método *Hi-C* usa o sequenciamento *high-throughput* para identificar todos os pares de sequências alinhados individualmente ao genoma determinando, assim, os fragmentos que sofrem ligação. Os

³ do inglês, *Chromosome Conformation Capture*

⁴ do inglês, *quantitative Polymerase Chain Reaction*

⁵ do inglês, *Chromosome Conformation Capture-on-Chip*

⁶ do inglês, *Chromosome Conformation Capture Carbon Copy*

fragmentos quiméricos são marcados com biotina, que pode ser capturada por *beads* de estreptavidina, a qual tem alta afinidade pela biotina, e então sequenciados. Dessa forma, todas as possíveis interações pareadas entre fragmentos são identificadas (Figura 2).

Já na técnica de *ChIP-Loop*, assim como no padrão 3C, a cromatina é fixada com formaldeído e digerida por enzimas de restrição. Os *crosslinks* de cromatina são então imunoprecipitados usando um anticorpo contra uma proteína específica de interesse e, posteriormente, a região putativa onde esta ligação da proteína ocorre é validada por PCR. Por fim, também com base na imunoprecipitação da cromatina e no sequenciamento de marcadores de extremidade emparelhados (PET⁷), o *ChIA-PET*⁸ permite a análise em todo o genoma das interações ligadas por proteínas específicas (Figura 2). Positivamente, todos esses métodos baseados em 3C passaram a gerar grandes quantidades de dados de interação em todo o genoma, o que possibilitou o desenvolvimento de outras ferramentas *in silico* que pudessem armazenar esses dados e compilá-los de maneira a permitir uma interface acessível desses bancos de dados, para que, assim, os pesquisadores possam ter acesso a esses experimentos e analisá-los de maneira a contribuir com pesquisas no mundo todo (Wang et al., 2019).

⁷ do inglês, *Pair-Ended Tag*

⁸ do inglês, *Chromatin Interaction Analysis by Paired-End Tag Sequencing*

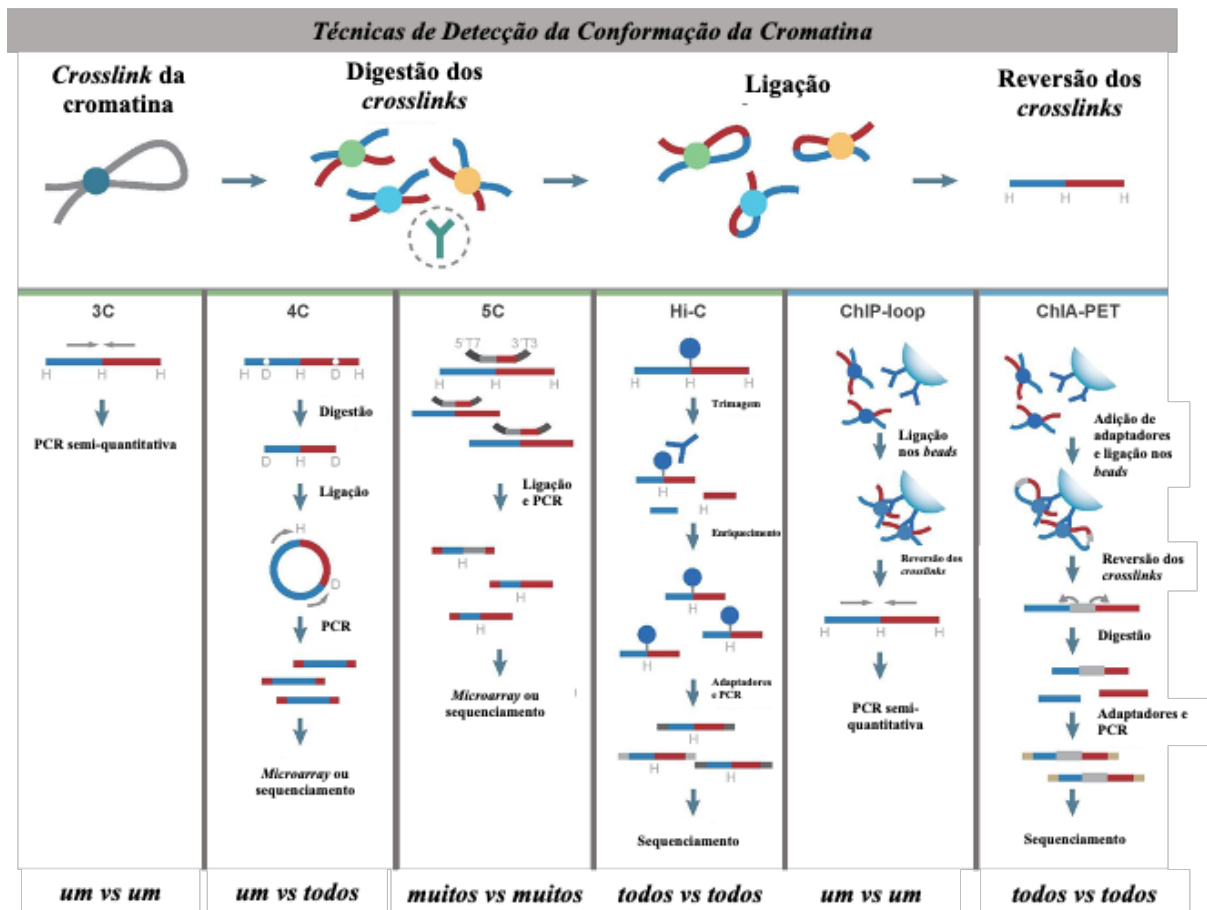


Figura 2. Diagrama esquemático das principais técnicas de detecção da conformação da cromatina. Adaptado de de Wit and de Laat, 2012.

2.3 Genômica comparativa e a identificação de elementos reguladores da transcrição

Devido à importância da compreensão sobre a regulação da expressão gênica, várias ferramentas de bioinformática foram desenvolvidas para prever com maior especificidade potenciais regiões que atuam como elementos cis-reguladores da transcrição, sendo uma destas a genômica comparativa (Frazer et al., 2004; Herrero et al., 2016; Karolchik et al., 2014; Ovcharenko et al., 2004).

A informação genética contida nos genomas de vertebrados é organizada em sequências codificadoras e sequências não codificadoras de proteínas, representadas por íntrons, elementos repetidos, RNAs não codificantes e elementos reguladores de genes. Enquanto as sequências codificadoras de proteínas representam cerca de 1,5 a 2% dos genomas dos vertebrados, a grande maioria das sequências não é codificadora (Lander et al., 2001).

Por via de regra, as sequências codificadoras de proteínas são conservadas filogeneticamente devido ao papel vital das proteínas no desenvolvimento e na homeostase pós-natal. Por outro lado, a maioria das sequências não codificadoras evolui rapidamente, conforme o esperado pela evolução neutra dessas sequências. No entanto, um subconjunto de sequências não codificadoras exibe um alto nível de conservação evolutiva, sendo, portanto, denominados elementos não codificadores conservados (CNEs⁹) (Polychronopoulos et al., 2017). A busca sistemática por CNEs revelou que existem centenas de milhares desses elementos distribuídos no genoma dos vertebrados, alguns dos quais têm sido conservados por mais de 400 milhões de anos (Bejerano et al., 2004; Woolfe et al., 2005). Nos últimos anos, vários CNEs foram identificados e caracterizados funcionalmente, demonstrando que muitos deles são importantes para o desenvolvimento embrionário e evolução de novas características nos animais, enquanto outros estão relacionadas às doenças humanas (Goode et al., 2011; Lickwar et al., 2017; Marcovitz et al., 2016; Parveen et al., 2013).

A disponibilidade de recursos *online* para a análise genômica comparativa permitiu a vários pesquisadores avaliar a conservação de regiões genômicas em torno de genes de interesse para identificar CNEs com possíveis papéis funcionais (Parveen et al., 2013). As regiões genômicas que envolvem os genes do desenvolvimento geralmente são enriquecidas em CNEs, e estes são distribuídos nos segmentos genômicos com baixa densidade gênica (Nobrega et al 2003). A explicação para essa distribuição é que muitos CNEs regulariam a expressão de genes do desenvolvimento, possivelmente coordenando a atividade de mais de um gene simultaneamente, e que regiões com menor densidade gênica alocariam múltiplos elementos reguladores (Polychronopoulos et al., 2017). De fato, ensaios com genes-repórter realizados utilizando importantes organismos-modelo, como peixe-zebra, galinha e camundongo, mostraram que os CNEs atuam, principalmente, como *enhancers* durante o desenvolvimento embrionário (Polychronopoulos et al., 2017). Embora nem todos os CNEs desempenhem um papel funcional, sabe-se que quanto maior o nível de conservação dos CNEs, maior a chance de eles desempenharem um papel funcional. A presença de vários sítios de ligação de fatores de

⁹ do inglês, *Conserved Non-coding Elements*

transcrição (TFBSs¹⁰) sobrepostos em CNEs específicos também é um forte indicador de funcionalidade (Visel et al., 2007).

2.4 *Blocos genômicos regulatórios coordenam sinergicamente os genes do desenvolvimento*

Os estudos de genômica comparada revelam que os mais diversos organismos possuem em seus genomas sequências não codificadoras de proteínas com alto grau de conservação entre diferentes espécies. No entanto, estas sequências denominadas CNEs, quando detectadas entre espécies filogeneticamente próximas (por exemplo, humanos e camundongos), podem não ser elementos funcionais, porém, uma grande maioria dos CNEs comuns entre espécies distantes têm chances maiores de serem funcionais (Polychronopoulos et al., 2017).

Assim como os outros elementos reguladores da expressão gênica, os CNEs podem estar presentes por todo o genoma, porém são comumente agrupados próximos a genes relacionados ao desenvolvimento embrionário (Polychronopoulos et al., 2017). Além disso, os CNEs podem estar organizados em grupos funcionais chamados de blocos genômicos reguladores (GRBs¹¹) (Polychronopoulos et al., 2017). Juntos, os CNEs são responsáveis por regular a expressão coordenada de diferentes genes-alvo e que geralmente possuem domínios de expressão semelhantes e/ou complementares. A arquitetura destes CNEs também apresenta uma particularidade: são encontrados em *clusters* distribuídos por regiões de baixa densidade gênica, ou seja, aparecem agrupados ao longo de longas regiões intergênicas e têm como alvo, geralmente, os genes que regulam o desenvolvimento e que se encontram flanqueadas por estas regiões intergênicas, como esquematizado na Figura 3 (Kim and Pritchard, 2007; Woolfe et al., 2005). Vários estudos já demonstravam que os CNEs atuam como elementos cis-reguladores que coordenam a expressão espaço-temporal de genes, especialmente durante o desenvolvimento embrionário, agindo, em sua vasta maioria, como *enhancers* (Aparicio et al., 1995).

¹⁰ do inglês, *Transcription Factor Binding Sites*

¹¹ do inglês, *Genomic Regulatory Blocks*

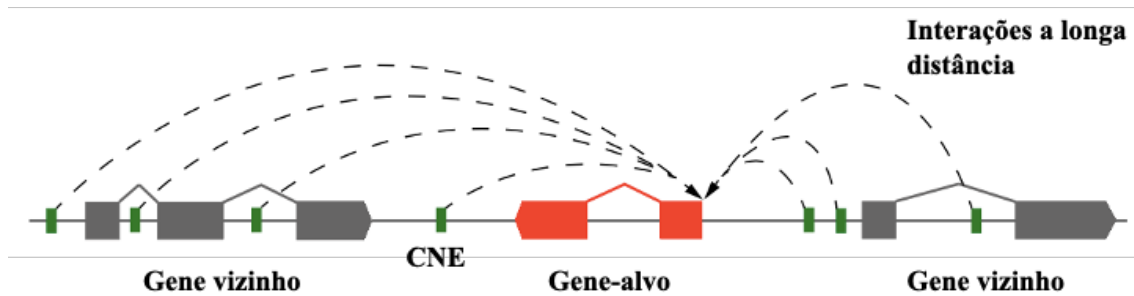


Figura 3. Arquitetura dos blocos genômicos reguladores. Em verde estão representados os elementos cis-reguladores, os quais, juntos, podem regular a expressão coordenada de diferentes genes-alvo que geralmente possuem domínios de expressão semelhantes e/ou complementares. Adaptado de Polychronopoulos et al., 2017.

2.5 Mudanças evolutivas dos elementos reguladores e a importância para a evolução

Elementos cis-reguladores podem sofrer mutações, duplicações ou deleções e serem modificados, gerando novos padrões de expressão dos seus genes-alvo. Isto ocorre, pois neste processo de modificação, estes elementos podem perder sítios de ligação para fatores de transcrição ou, então, novos sítios para outros fatores podem surgir dentro deste elemento. Além disso, mudanças ao longo do tempo evolutivo podem ocorrer não nos elementos cis-reguladores, mas sim no padrão de expressão dos fatores de transcrição que ali interagem, causando a ativação de um *enhancer* em um novo tipo celular, onde até então este último permanecia inativo (Strähle and Rastegar, 2008).

Várias mudanças evolutivas em características morfológicas foram atribuídas a alterações nos elementos cis-reguladores e não às mutações nas sequências codificadoras (Colosimo et al., 2005; Shapiro et al., 2006; Shapiro et al., 2004; Stern, 1998; Sucena et al., 2003). Isso inclui casos em que o ganho ou perda de sítios de ligação para fatores de transcrição altamente conservados culminam na alteração da função de elementos cis-reguladores e, por consequência, na atividade dos genes por eles regulados (Gompel et al., 2005; Jeong et al., 2006). Estas mudanças podem repercutir, inclusive, na seleção de características funcionalmente significativas e ecologicamente relevantes, afetando uma ampla gama de características morfológicas, fisiológicas e comportamentais e levando à grande diversidade taxonômica e, com

isso, terem um importante papel na diversificação morfológica (Bachner-Melman et al., 2005; Daborn et al., 2002; Hamblin and Di Rienzo, 2000; Shapiro et al., 2004; Trefilov et al., 2000; Wray, 2007).

Um exemplo deste fenômeno foi demonstrado por Jeong e colaboradores (2006) em diferentes espécies do gênero *Drosophila*. A pigmentação do abdômen posterior masculino é uma característica adquirida na linhagem de *Drosophila melanogaster*. O estudo mostrou que o fator de transcrição Hox Abdominal-B (ABD-B) ativa indiretamente a expressão do gene de pigmentação, chamado de *yellow*, nos segmentos posteriores. A regulação desta proteína evoluiu através do ganho do sítio de ligação para ABD-B em um elemento cis-regulador específico da regulação do gene *yellow*, em um ancestral comum das espécies com dimorfismo sexual. Dentre os subgrupos de *melanogaster*, que incluem *Drosophila bipectinata*, *Drosophila kikkawai* e *Drosophila santomea*, a pigmentação distinta nos machos foi subsequentemente perdida por mecanismos diferentes, sendo um deles a mutação no sítio de ligação para a proteína ativadora ABD-B (Jeong et al., 2006).

Em um outro estudo pioneiro, Tournamille e colaboradores (1995) demonstraram como uma mutação em um sítio de ligação para fatores de transcrição pode afetar fenótipos com repercussão na saúde humana. A membrana dos eritrócitos contém várias proteínas de superfície e transmembranares, sendo que grande parte destas é altamente polimórfica e dão origem aos chamados antígenos de grupos sanguíneos. Um destes sistemas de grupos sanguíneos é o Duffy, que constitui antígenos com importante significado clínico. Tournamille e colaboradores (1995) estudaram o mRNA do antígeno deste grupo sanguíneo, um receptor de eritrócito para o parasita causador da malária (*Plasmodium vivax*). Este mRNA composto por um único exon é silenciado em pessoas negativas para Duffy, pois carregam um alelo com uma substituição do tipo T-C no nucleotídeo -46. Esta mutação impede a atividade do promotor gênico nos eritrócitos, pois causa um rompimento no sítio de ligação para o fator de transcrição GATA1. A não expressão do receptor neste tipo celular, torna estas células resistentes à invasão pelo parasita causador da malária (Tournamille et al., 1995).

Importantemente, à medida que os métodos para identificação de mutações em elementos cis-reguladores, que geram efeitos funcionais, se tornem mais poderosos e,

principalmente, à medida que os pesquisadores concentrem esforços para identificá-las, mais numerosos se tornam os exemplos, o que pode contribuir amplamente para o entendimento da evolução fenotípica entre as espécies (Wray, 2007).

2.6 *Dact1 e seus múltiplos papéis*

2.6.1 *A família das proteínas adaptadoras Dact*

Os genes da família *Dact*¹² (também denominados *Dapper*¹³/*Frodo*) codificam proteínas multifuncionais que desempenham importantes funções na embriogênese e homeostase pós-natal. Estas proteínas intracelulares são importantes moduladoras das vias de sinalização Wnt¹⁴ e TGF- β ¹⁵, e são requeridas para regular os níveis de sinalização parácrina em diferentes contextos celulares. O primeiro membro desta família de genes, o *Dact1*, foi descoberto em exemplares de peixe-zebra (*Danio rerio*) (Cheyette et al., 2002). Hoje, sabe-se que os *Dacts* compõem uma família constituída por três genes (*Dact1*, *Dact2* e *Dact3*) em mamíferos, dois genes (*Dact1* e *Dact2*) em aves e nos anfíbios (*dact1* e *dact3*), e no peixe-zebra são encontrados seis membros dessa família gênica (*dact1*, *dact2*, *dact3a* e *dact3b*, *dact4*, *dact4r*) (Schubert et al., 2014).

As proteínas que compõem esta família têm tamanhos semelhantes (600 a 850 aminoácidos) e se distinguem por um motivo *leucine zipper* conservado próximo à porção N-terminal e um motivo de ligação para domínios PDZ¹⁶ na porção C-terminal (Cheyette et al., 2002; Fisher et al., 2006; Katoh et al., 2002; Katoh and Katoh, 2005), importante domínio encontrado em proteínas de organização do citoesqueleto e de transdução de sinal transmembrana (Ponting et al., 1997). Eles também compartilham alguns motivos curtos (4 a 8

¹² do inglês, *Dapper antagonist of catenin*

¹³ do inglês, *Dishevelled-associated antagonist of β -catenin*

¹⁴ do inglês, *Wingless-type MMTV integration site*

¹⁵ do inglês, *Transforming Growth Factor- β*

¹⁶ do inglês, *Post synaptic density-95/Discs large/Zonula occludens-1*

aminoácidos) idênticos distribuídos em outras partes de suas sequências primárias (Kivimäe et al., 2011).

Dact1, o primeiro membro da família *Dact* a ser isolado, foi identificado por meio do ensaio de sistema de duplo híbrido em levedura, buscando identificar novos parceiros da proteína Dvl¹⁷, a qual tem papel central na sinalização Wnt/ β -catenina ou canônica. Esta via é ativada pela família de glicoproteínas Wnt, com efeitos parácrinos e autócrinos na regulação da proliferação, sobrevivência, comprometimento e diferenciação celulares (Qin et al., 2010). Na via Wnt canônica, os membros desta família se ligam a receptores *frizzled* (FZD) e a co-receptores LRP5/6 levando à inativação do complexo de degradação da β -catenina, que inclui GSK-3 β ¹⁸, Axin e APC¹⁹, por meio de Dvl. Desta forma, na presença de Wnts, GSK-3 β deixa de bloquear a fosforilação da β -catenina. A hipofosforilação da β -catenina e a sua translocação para o núcleo celular leva à ligação com membros da família de fatores de transcrição LEF/TCF e a ativação de genes alvo da via Wnt (Christodoulides et al., 2009). Neste contexto, *Dact1* é capaz de se ligar à Dvl e encaminhá-la à degradação, em consequência disto, as moléculas de β -catenina também são degradadas, pois o complexo de degradação de β -catenina deixa de ser inativado por Dvl, culminando na inibição dos genes-alvo da via Wnt canônica. *Dacts* também podem atuar como moduladores negativos da via de sinalização Wnt/Polaridade Celular Planar (Wnt/PCP). Nesse contexto, *Dact* interage Dvl e Vangl2²⁰ bloqueando essa via (Kivimäe et al., 2011; Suriben et al., 2009).

Outros parceiros moleculares de *Dact1* e dos demais membros desta família foram descobertos posteriormente. Sabe-se hoje que estas proteínas também formam complexos com as CK1²¹ δ/ϵ , PKA²², PKC²³ e catenina p120 (Kivimäe et al., 2011). Além dessas interações, as proteínas *Dact* formam homo e heterodímeros entre si (Kivimäe et al., 2011). Em adição, *Dact1*

¹⁷ do inglês, *Dishevelled*

¹⁸ do inglês, *Glycogen synthase kinase 3*

¹⁹ do inglês, *Adenomatous polyposis coli*

²⁰ do inglês, *Van Gogh-like 2*

²¹ do inglês, *Casein Kinase 1*

²² do inglês, *Protein Kinase A*

²³ do inglês *Protein Kinase C*

apresenta também, na sua posição 120 N-terminal significativa similaridade com domínio *coiled-coiled* rico em leucina encontrado nas distrofinas (DRD²⁴) (Gloy et al., 2002), sugerindo um papel estrutural de Dact também em células musculares. Outras interações também foram reportadas entre Dacts e LEF/TCF²⁵, GSK-3 β , proteínas 14-3-3, HDAC1²⁶ e DBF4²⁷ (Brott and Sokol, 2005a; Brott and Sokol, 2005b; Cheyette et al., 2002; Gao et al., 2008; Hikasa and Sokol, 2004). Por fim, embora a maioria das interações que ocorrem com a família de proteínas Dact se dê com proteínas citoplasmáticas, é interessante ressaltar que já foi demonstrado que Dact1 contém uma sequência de localização nuclear (NLS²⁸) e um sinal de exportação nuclear (NES²⁹), permitindo que esta se transloque para o núcleo (Gao et al., 2008), o que levanta também a possibilidade de esta proteína atuar como um regulador transcricional ou mesmo desempenhe alguma outra função neste compartimento celular ainda a ser estabelecida.

2.6.2 Funções dos genes *Dact* no desenvolvimento dos vertebrados

Embora seja possível estabelecer mecanismos de ação comuns aos *Dacts*, algumas particularidades são observadas entre eles. Durante o desenvolvimento do peixe-zebra, *Dact2* regula os movimentos de extensão convergente, modulando a via de sinalização Wnt/PCP (Waxman et al., 2004). Quanto ao desenvolvimento em camundongos, sabe-se que *Dact2* age como um modulador negativo da via Wnt canônica durante a formação dos dentes, pois reprime a expressão de *Pitx2* (Li et al., 2013). Em 2017, Li e colaboradores demonstraram que *Dact2* tem a capacidade de suprimir o crescimento de células tumorais de câncer de mama em murinos através da inibição da via Wnt/ β -catenina e que a hipermetilação da região do promotor de *Dact2* está diretamente relacionada ao tamanho do tumor em humanos, pois esta metilação permite a ativação da via Wnt/ β -catenina (Li et al., 2017a). Por fim, foi demonstrado que *Dact2*

²⁴ do inglês, *Dystrophin-related Domain*

²⁵ do inglês, *Lymphoid Enhancing Factor/T Cell Factor*

²⁶ do inglês, *Histone Deacetylase 1*

²⁷ do inglês, *DumbBell Forming-4*

²⁸ do inglês, *Nuclear Localization Sequence*

²⁹ do inglês, *Nuclear Export Signal*

diminui a sinalização TGF- β /Nodal, facilitando a degradação lisossomal de receptores ALK4 e ALK5 durante a indução do mesoderma no desenvolvimento embrionário (Lee et al., 2010; Su et al., 2007; Zhang et al., 2004a).

O parálogo *Dact3* foi primariamente descrito apenas nos mamíferos (Fisher et al., 2006). No entanto, estudos recentes revelaram que dois parálogos desse gene (*dact3a* e *dact3b*) agem dinamicamente durante o desenvolvimento embrionário também em peixes-zebra (Mandal and Waxman, 2014). Em células tumorais colorretais humanas, a repressão epigenética de *Dact3* foi associada com o aumento da atividade da via Wnt canônica, sugerindo que *Dact3* age como um modulador negativo desta via e tem papel importante em patologias humanas (Jiang et al., 2008). Finalmente, no que diz respeito ao gene *Dact4*, foi demonstrado, *in silico*, que alterações de perda e ganho de motivos da proteína devem levar à incapacidade de Dact4 se dimerizar. Além disto, a ausência do domínio de ligação com Dvl neste parálogo sugere que esta proteína seja incapaz de modular a via Wnt ou que atua como uma forma dominante negativa dos demais parálogos desta família (Schubert et al., 2014). Assim, como alguns motivos de interação proteína-proteína se mantiveram conservados, sugere-se que Dact4 poderia atuar no sequestro de outras proteínas que interagem com os demais membros da família Dact, antagonizando suas funções (Schubert et al., 2014).

Foco das pesquisas apresentadas nesta tese, o gene *Dact1* é o membro mais extensamente estudado da família de proteínas Dact, sendo um importante gene do desenvolvimento embrionário e também de grande relevância para o equilíbrio pós-natal. Waxman e colaboradores (2004) demonstraram por ensaios de perda e ganho de função em embriões de peixes-zebra, que *Dact1* é um modulador positivo da via Wnt/ β -catenina, além de ativar os genes-alvo desta via através da interação com Dlv2 e outras quinases como CK1 ϵ , Par1 e CK2 (Waxman et al., 2004). Com relação ao padrão de expressão, nos estágios iniciais de desenvolvimento dos peixes-zebra, observa-se a expressão de *Dact1* por todo o blastoderma, sendo, mais tarde, especificamente expresso em domínios mesodérmicos e ectodérmicos (Waxman et al., 2004).

No modelo experimental de anfíbio, *Xenopus laevis*, transcritos de *Dact1/Frodo* estão restritos à placa neural anterior e às regiões dos olhos e dos arcos branquiais (Gloy et al.,

2002; Hunter et al., 2006). Conforme o desenvolvimento embrionário avança, outros grandes domínios de expressão também são observados, como no mesoderma pré-somítico (PSM) da cauda e também nos somitos (Hunter et al., 2006). Estudos de perda e ganho de função demonstraram que *Dact1/Frodo* é necessário para atividade indutora de eixo de Dvl, mas não de β -catenina (Gloy et al., 2002).

Nos modelos murinos e de aves (*Mus musculus* e *Gallus gallus*) *Dact1* é expresso majoritariamente nas estruturas de origem mesodérmica, por exemplo, o mesoderma paraxial não segmentado, somitos, além de, posteriormente, no mesênquima dos brotos dos membros. Contudo, sua expressão é também muito intensa em derivados ectodérmicos, como o sistema nervoso central (SNC) e pele (Alvares et al., 2009; Fisher et al., 2006). Esse padrão é similar ao observado nos outros modelos animais, o que reflete uma conservação evolutiva dos padrões de expressão de *Dact1* (Hunter et al., 2006). Os principais domínios de expressão durante o desenvolvimento embrionário dos modelos animais citados são sumarizados na Tabela 1.

Tabela 1. Padrão de expressão de *Dact1* durante o desenvolvimento embrionário.

Domínio de expressão	<i>Gallus gallus</i>	<i>Mus musculus</i>	<i>Danio rerio</i>	<i>Xenopus laevis</i>
Linha primitiva	+	+	+	+
Botão caudal	-	+	+	+
Processo cefálico/notocorda	+		+	
Mesoderma pré-somítico (PSM)	+	+	+	+
Somitos	+	+	+	+
Mesoderma lateral	+	+		+
Mesoderma intermediário	-			
Membros (mesoderma)	+	+		
Placóide neural	+			
Olhos	+		+	+
Placa neural	-	+	+	+
Tubo neural	-		+	+
Células da crista neural	+	+		+
Endoderma	+			

Referências: Cheyette et al., 2002; Fisher et al., 2006; Gillhouse et al., 2004; Gloy et al., 2002; Hunter et al., 2006; Suriben et al., 2006; Waxman et al., 2004.

Camundongos mutantes para *Dact1* exibem múltiplas anomalias, como agenesia das vértebras caudais, malformações anorretais, renais e da bexiga, causadas por falha de gastrulação na faixa primitiva posterior (Wen et al., 2010). Além disso, mutações *missense* no gene *Dact1* foram observadas também em humanos, por exemplo, em pacientes com defeitos do tubo neural (Shi et al., 2012), o que sugere que uma de suas principais funções é a regulação da via da polaridade celular planar (PCP) durante o desenvolvimento do tubo neural. Uma outra implicação da mutação de *Dact1* em humanos é a síndrome de Townes-Brocks, que causa o desenvolvimento esquelético anormal, incluindo malformação do polegar, espinha bífida oculta e escoliose (Webb et al., 2017).

2.6.3 Funções dos genes *Dact* na fase pós-natal

Dada a importância de *Dact1* para o desenvolvimento embrionário dos vertebrados, este gene passou a ser intensamente estudado também em outro aspecto da vida dos organismos – a homeostase no pós-natal. Estudos revelaram importantes correlações entre o desbalanço na expressão gênica de *Dact1* e suas repercussões em diversos contextos patológicos.

Alterações na atividade de *Dact1* estão relacionadas a diversas patologias que podem acometer os indivíduos na vida adulta. Lagathu e colaboradores (2009) demonstraram que o gene *Dact1* é expresso em pré-adipócitos de humanos e murinos e na fração vascular estromal do tecido adiposo branco. A expressão de *Dact1* confere potencial pró-adipogênico e sua presença nos pré-adipócitos é necessária para que haja diferenciação em adipócitos maduros. Foi também demonstrado que esse papel atribuído ao gene *Dact1* deve-se ao fato de ele atuar como um antagonista da via Wnt/ β -catenina nos pré-adipócitos, permitindo a cascata de reações que leva à degradação da β -catenina e que, portanto, permite a expressão de fatores de transcrição adipogênicos, como PPARG e CEBPA (Lagathu et al., 2009). Lagathu e colaboradores (2009) demonstraram, ainda, que sob condições de resistência à insulina, quando a adipogênese e a deposição de gordura atingem um platô, esta rede deixa de ocorrer, mas *Dact1* continua atuando, facilitando a expansão do tecido adiposo a fim de armazenar nutrientes baseado na situação nutricional e metabólica do indivíduo. A desregulação desta complexa rede pode ser a base patológica que leva a um balanço alterado entre o crescimento dos adipócitos e o recrutamento dos pré-adipócitos que, em última instância, leva à variação entre a hipertrofia e a hiperplasia neste tecido (Lagathu et al., 2009).

Guo e colaboradores (2017) demonstraram que a hipermetilação da região que contém o sítio de início da transcrição (TSS³⁰) pode ser um dos principais mecanismos para a expressão reduzida de *Dact1*, e também *Dact2*, no carcinoma de células escamosas do esôfago (ESCC³¹). A metilação simultânea de *Dact1* e *Dact2* pode desempenhar papéis importantes na progressão da ESCC e podem servir como biomarcadores prognósticos para pacientes com este tipo de câncer

³⁰ do inglês, *Transcription Start Site*

³¹ do inglês, *Esophageal Squamous Cell Carcinoma*

(Guo et al., 2017). A regulação negativa de *Dact1* foi observada também em carcinomas hepáticos (Yau et al., 2005) e gástricos (Astolfi et al., 2010), além de câncer de pulmão de células não-pequenas (NSCLC³²) (Yang et al., 2010) e em melanoma (Farmer, 2016). Já a superexpressão de *Dact1* foi detectada em câncer de cólon (Yuan et al., 2012).

Dact1 também foi relacionado com câncer epitelial de ovário tipo I. Nesta condição, a expressão de *Dact1* nas células epiteliais do ovário se mostrou reduzida em relação às células saudáveis (Li et al., 2017b). Utilizando cultura primária dessas células cancerosas provenientes de pacientes diagnosticadas, a superexpressão de *Dact1* por ensaio com lentivírus levou a uma redução do crescimento e do número de tumores em cultura (Li et al., 2017b). A este fato, Li e colaboradores (2017) atribuem a capacidade de *Dact1* ser um regulador negativo do câncer epitelial de ovário tipo I à sua ação como inibidor da via Wnt canônica, além de induzir a autofagia destas células tumorais. Este mesmo mecanismo de silenciamento de *Dact1* foi observado em casos de câncer de mama. Yin e colaboradores (2013) identificaram em um estudo epigenômico pioneiro que *Dact1* é um alvo de metilação em tumores malignos da mama. A expressão de *Dact1* estava silenciada em 88,9% das linhagens de câncer de mama, além de haver significativa redução da sua expressão proteica quando comparada ao tecido saudável que margeia os tumores. Em adição, estudos funcionais demonstraram que a expressão ectópica de *Dact1* pôde inibir a proliferação celular das células tumorais *in vivo* e *in vitro*, mais uma vez pela sua ação como inibidor da via Wnt/ β -catenina (Yin et al., 2013).

Em um outro exemplo da importância de *Dact1* em patologias, Zhang e colaboradores (2018) demonstraram em amostras provenientes dos pulmões de crianças com asma que os três membros da família *Dact* estavam superexpressos, e *Dact1* se mostrou ainda mais superexpresso quando comparado a *Dact2* e *Dact3*. Este estudo evidenciou que *Dact1* reprime a expressão e a translocação de β -catenina para o núcleo celular, inibindo a via Wnt, que é necessária para reduzir a resposta inflamatória local. Dessa forma, a ação exacerbada de *Dact1* possivelmente induz níveis mais altos de fatores pró-inflamatórios, característicos da doença (Zhang et al., 2018).

³² do inglês, *Non-small Cell Lung Cancer*

Recentemente, Hou e colaboradores (2019) apontaram o envolvimento de Dact1 na fibrilação atrial, uma arritmia cardíaca que apresenta considerável nível de morbidade e que pode levar à morte (Hou et al., 2019). Nesta patologia, a diminuição da expressão citoplasmática de Dact1 no miocárdio foi associada à fibrilação atrial e um alto grau de fibrose no coração (Hou et al., 2019). Nas células cardíacas, a β -catenina é essencial para manutenção da função celular, uma vez que esta se liga às caderinas formando um complexo com a actina do citoesqueleto, promovendo o rearranjo deste último (Hou et al., 2019). Assim, este estudo mostrou que, na presença de Dact1, a β -catenina não foi encontrada no núcleo, mas sim no citoplasma e na membrana celular, sugerindo que Dact1 pode facilitar o acúmulo extranuclear β -catenina, que, por sua vez, pode servir como uma proteína estrutural. Quando *Dact1* está diminuído, no caso da fibrilação atrial, isso não acontece, e então ele pode servir como um potencial marcador terapêutico para esta doença (Hou et al., 2019). A Figura 4 traz uma representação esquemática dos principais órgãos do corpo humano associados às doenças com relevante alteração na expressão de *Dact1*.



Figura 4. Representação esquemática dos principais órgãos do corpo humano associados às doenças com significativa alteração na expressão de Dact1.

3. OBJETIVOS

3.1 OBJETIVOS GERAIS

Caracterizar as expressões gênica e proteica de *Dact1* na miogênese de amniotos e ampliar os conhecimentos sobre a regulação transcricional deste gene em diversos contextos, incluindo o do músculo esquelético.

3.2 OBJETIVOS ESPECÍFICOS

3.2.1 CARACTERIZAÇÃO DA EXPRESSÃO GÊNICA E PROTEICA DE *DACT1*

- Caracterizar a expressão gênica e proteica de *Dact1* ao longo da diferenciação miogênica, utilizando o modelo *in vitro* de células C2C12.
- Investigar a existência de correlação da expressão de *Dact1* com miopatias no humano.

3.2.2 IDENTIFICAÇÃO E CARACTERIZAÇÃO DE ELEMENTOS CIS-REGULADORES

- Identificar CNEs presentes no entorno do locus gênico de *Dact1* por meio de uma abordagem de genômica comparativa.
- Caracterizar os CNEs identificados no locus gênico de *Dact1*, comparando-os com elementos já previamente descritos.
- Identificar os contextos celulares onde os CNEs possivelmente devem atuar, tendo como base os dados experimentais para os elementos previamente caracterizados nas mesmas regiões.
- Identificar os sítios de ligação para fatores de transcrição conservados evolutivamente presentes nos CNEs.

- Identificar e analisar as redes de interação proteína-proteína para as quais os CNEs apresentam sítios de ligação para fatores de transcrição preditos.
- Demonstrar a funcionalidade de um CNE do locus gênico de *Dact1* utilizando ensaios de gene-repórter nas linhagens celulares HEK293T, C2C12 e pré-adipócitos murinos imortalizados (9W).
- Analisar um CNE com potencial de regulação na miogênese, comparando a região correspondente em diversas espécies de amniotos.

4. CAPÍTULO 1 – CARACTERIZAÇÃO DA EXPRESSÃO GÊNICA E PROTEICA DE *DACT1*

4.1 INTRODUÇÃO

A musculatura esquelética nos vertebrados é responsável por funções essenciais para a manutenção da vida, como a respiração e a deglutição, além prover sustentação e movimento ao nosso corpo. Apesar disso, a perda da massa muscular esquelética, seja por situações patológicas, como na caquexia causada por alguns tipos de câncer ou pelo próprio envelhecimento, afeta a vida de milhares de pessoas. Por este motivo, estudos que identifiquem moléculas envolvidas na formação, homeostase e reparo do tecido muscular são importantes para o desenvolvimento de estratégias que possam aumentar a sobrevida e melhorar a qualidade de vida das pessoas.

A miogênese embrionária dá origem ao primeiro conjunto de fibras musculares multinucleadas nos miótomos e nos membros em desenvolvimento no embrião (Chal and Pourquié, 2017). Subsequentemente, a miogênese fetal ocorre para que haja crescimento e maturação dos músculos esqueléticos (Davis and Fiorotto, 2009). Finalmente, a miogênese no pós-natal é necessária para o reparo de músculos danificados e para hipertrofia muscular, envolvendo as células satélites, que são células mitoticamente quiescentes localizadas entre a lâmina basal e o sarcolema das fibras musculares (Le Grand and Rudnicki, 2007). Uma ferramenta amplamente utilizada para o estudo da miogênese é a cultura de mioblastos murinos C2C12. As células C2C12 são mioblastos derivados de células-satélites, cujo comportamento é semelhante ao da linhagem progenitora. Estas células são subclones de mioblastos C2 (Yaffe and Saxel, 1977a; Yaffe and Saxel, 1977b), que espontaneamente se diferenciam em cultura na ausência de soro (Blau et al., 1983). Estes mioblastos em diferenciação são comparáveis às células-satélites ativadas nas fibras musculares e, por isso, são comumente adotadas para o estudo dos processos miogênicos (Yoshida et al., 1998).

As vias Wnt/ β -catenina e Wnt não canônicas são conhecidas por também regular a miogênese esquelética. Além de coordenar a transcrição de fatores reguladores miogênicos (MRFs), como MYF5, MYOD, MRF4 e MYOG (Asfour et al., 2018; Tajbakhsh et al., 1998; Yu et

al., 2013), a sinalização Wnt canônica é necessária para que o somito gere o dermomiótomo e o miótomo e também para a delaminação das células progenitoras, as quais migram para o membro para compor os músculos apendiculares (Hutcheson et al., 2009). Dentro dos miótomos, a via não canônica Wnt/PCP é necessária para a orientação e o alongamento adequados dos miócitos (Gros et al., 2009). Na miogênese fetal, a via Wnt/ β -catenina modula a deposição de massa muscular, enquanto a especificação do tipo de fibra envolve a sinalização canônica e Wnt/ Ca^{2+} (Chin et al., 1998; Kuroda et al., 2013). Já na miogênese pós-natal, a via Wnt/PCP atua para equilibrar a autorrenovação e a diferenciação das células satélites, garantindo a manutenção de um reservatório de células-tronco satélites reparadoras no músculo danificado (Le Grand et al., 2009).

À luz de que *Dact1* é um importante modulador das vias de sinalização Wnt, que por sua vez atuam em diversos processos da miogênese, seja esta embrionária, fetal ou no adulto, buscamos caracterizar a expressão de *Dact1* neste contexto que, até então, ainda não havia sido caracterizado. Estudos recentes do nosso grupo de pesquisa investigaram o padrão de expressão gênica e proteica de *Dact1*/*Dact1* na cultura primária de mioblastos de galinha, o que motivou o interesse em avaliar se o mesmo é recapitulado na cultura de mioblastos murinos. Além disso, a caracterização adequada da expressão de *Dact1* ao longo da miogênese esquelética neste modelo é importante para entender a possível dinâmica de modulação deste gene neste contexto, inclusive no que tange à sua regulação em doenças que acometem o músculo esquelético no humano.

Neste capítulo iremos abordar os ensaios de caracterização da expressão gênica e distribuição da proteína *Dact1* ao longo da diferenciação das células miogênicas C2C12 em paralelo à caracterização em cultura primária de mioblastos de galinha – achados que compõem o manuscrito “*Dact1 is expressed during chicken and mouse skeletal myogenesis and modulated in human muscle diseases*”. O manuscrito originado destes ensaios também apresenta análises *in silico* empregadas para avaliar a expressão de *Dact1* em miopatias humanas, a partir de dados do *Gene Expression Omnibus* (GEO), e para identificar regiões intrinsecamente desordenadas (IDRs) na proteína *Dact1*.

Em suma, nossos resultados em conjunto indicam uma conservação dos padrões de expressão de *Dact1* em mamíferos e aves, sugerindo que este gene é necessário para o desenvolvimento da musculatura esquelética dos vertebrados amniotos e que *Dact1* tem sua expressão alterada em algumas distrofias humanas.

4.2 RESULTADOS

O manuscrito *Dact1 is expressed during chicken and mouse skeletal myogenesis and modulated in human muscle diseases*, apresentado abaixo em inglês, foi submetido à revista científica *Histochemistry and Cell Biology*.

Full-Length Original Research

***Dact1* is expressed during chicken and mouse skeletal myogenesis and modulated in human muscle diseases**

¹Renata Erbert Contriciani^a, ¹Fernanda Cristina da Veiga^a, Mariana Juliani do Amaral^b, Bianca Gazieri Castelucci^a, Lizandra Maia de Sousa^a, Marcelo Bispo de Jesus^a, Sílvio Roberto Consonni^a, Cláudia Mermelstein^c, Carla Beatriz Collares-Buzato^a, Lúcia Elvira Alvares^{a,*}

^aDepartment of Biochemistry and Tissue Biology
University of Campinas - UNICAMP
Campinas, SP – Brazil

^bFaculty of Pharmacy
Federal University of Rio de Janeiro - UFRJ
Rio de Janeiro, RJ - Brazil

^cBiomedical Sciences Institute
Federal University of Rio de Janeiro - UFRJ
Rio de Janeiro, RJ - Brazil.

*Corresponding author:
tel: +55 19 3521 6252
e-mail address: lealvare@unicamp.br

¹These authors contributed equally to this work

Abstract

A precise control of Wnt signaling is required for skeletal muscle development and repair. *Dact1* (Dapper/Frodo) is an important modulator of Wnt signaling, which interacts with key components of Wnt transduction pathways. Here, we characterized *Dact1* gene expression and protein distribution in chicken and mouse fetal muscles and during the differentiation of chick primary and mouse C2C12 myoblasts. We also performed *in silico* analysis to investigate *Dact1* gene expression in human myopathies and to evaluate features of *Dact1* protein. Our results showed that in chicken *Dact1* is expressed in the fetal breast muscle and in primary myogenic cultures. *In vitro*, *Dact1* transcription is upregulated from differentiation and the protein is found both in the cytoplasm and nuclei of myogenic cells, as discrete aggregates. This nucleo-cytoplasmic expression is observed from myoblasts to myofibers, in which the *Dact1* protein is organized in a sarcomeric pattern. In mice, similar but not identical expression patterns were found in fetal skeletal muscle *in situ* and in C2C12 cultures. In human, *Dact1* expression was found to be modulated in specific muscle pathologies, including muscle dystrophies. Our bioinformatic analysis of *Dact1* protein revealed long intrinsically disordered regions, similarly to other proteins involved in Wnt signaling. In addition, we found that *Dact1* has strong propensity for liquid-liquid phase separation, a feature that explains its ability to form nuclear aggregates. Taken altogether, our data revealed that *Dact1* is broadly expressed during skeletal myogenesis of amniotes and envisages a possible pathophysiological role for this protein in muscular diseases.

Keywords:

Dact1; Skeletal Myogenesis; Wnt signaling; Chicken primary myoblast; C2C12; Intrinsically Disordered Proteins; Muscular Diseases.

Introduction

Skeletal muscles form the largest tissue in the body of vertebrates and are involved in essential activities of daily living such as locomotion, breathing and feeding. In addition, they have a key role in energy metabolism and exert important effects on other tissues and organs by releasing myokines during muscular contractions (Hargreaves and Spriet 2020; Barbalho et al. 2020). Due to its importance, several aspects of skeletal muscle biology have been studied and the identification of new molecules involved in its development, repair or disease-related dysfunction is of interest in both basic and applied research.

Mesoderm progenitor cells undergo specification and progressively differentiate to generate contractile myofibers in the body of vertebrates. This process, known as myogenesis, occurs in sequential and temporally distinct phases throughout life. The embryonic or primary myogenesis gives rise to the first set of multinucleated muscle fibers in the myotomes and developing limbs of the embryo, setting up the basic pattern that allows the formation of adult muscles (Chal and Pourquié 2017). The fetal or secondary myogenesis as well as the perinatal myogenesis occur on enlarged scales and, therefore, are fundamental for growth and maturation of skeletal muscles (Davis and Fiorotto 2009). The adult myogenesis is required for the repair of damaged muscles and exercise-mediated muscle hypertrophy and involves the satellite cells, mitotically quiescent cells located between the basal lamina and the sarcolemma of muscle fibers (Le Grand and Rudnicki 2007). Although temporally distinct, the different phases of myogenesis rely on a common set of regulatory molecules, whose main components are the paired box transcription factors Pax3 and Pax7 as well as the Myogenic Regulatory Factors (MRFs), a small family of basic helix-loop-helix transcription factors composed by Myf5, MyoD, MRF4 and Myogenin (Asfour et al. 2018). The expression of these molecules is finally regulated along the different steps of skeletal myogenesis. Initially, muscle stem cells that are either Pax3+ (mouse) or Pax3+/Pax7+ (chicken) in the embryonic myogenesis or Pax7+ in the fetal, perinatal and adult myogenesis, generate Myf5+/MyoD+ proliferative myoblasts (Chal and Pourquié 2017). After several cycles of proliferation, the myoblasts withdraw from the cell cycle and differentiate into Myogenin+ mononucleated myocytes. The myocytes then elongate, producing mononucleated myotubes, which are Myosin Heavy Chain positive (MyHC+) and express

contractile proteins such as actin, myosin and elastic myofilaments to assemble the sarcomeres. Progressively, the single-nucleated myotubes fuse with nearby myocytes and further mature, producing multinucleated contractile myofibers (Chal and Pourquié 2017). These myofibers express different types of MyHC, generating the physiologically distinct slow-twitch Type I and fast-twitch Type II fibers.

The Wnt glycoproteins compose a family of secreted molecules, which are able to trigger different signaling pathways to exert their biological effects. The canonical or Wnt/ β -catenin pathway is the best known Wnt pathway and has β -catenin as its central component. In the absence of Wnts, β -catenin is constantly target for phosphorylation and is degraded by a destruction complex composed by Axin, Adenomatous Polyposis Coli (APC), Casein Kinase I (CKI) and Glycogen Synthase Kinase 3 β (GSK3 β). However, this process is interrupted when Wnts reach target cells and bind to the Frizzled (Fzd) receptors and Low-density lipoprotein receptor-related proteins 5/6 (LRP5/6) co-receptors. This occurs because the Wnt-Fzd interaction activates Dishevelled (Dvl), which blocks the activity of the β -catenin destruction complex. This allows β -catenin to accumulate and be translocated to the nucleus, where this protein interacts with the Lymphoid Enhancer Factor (LEF) and T-cell Factor (TCF) transcription factors, to activate specific Wnt target genes. Wnts can also activate non-canonical pathways, which are β -catenin-independent but still require Wnt-Fzd interaction as well as the transduction activity of Dvl. Among these pathways are the Planar Cell Polarity (Wnt/PCP) and the calcium-dependent (Wnt/ Ca^{2+}) pathways (Girardi and Le Grand 2018). The Wnt/PCP pathway has the Rho and Rac small GTPases as central participants, which promote Rho-associated protein kinase (ROCK) or c-Jun N-terminal kinases (JNKs) activation, leading to changes in cell adhesion/motility or tissue polarity (Girardi and Le Grand 2018). The Wnt/ Ca^{2+} pathway involves the activation of phospholipase C (PLC) and increase of intracellular calcium (Ca^{2+}) levels, activating PKC (Protein Kinase C) or CamKII (Calcium-Calmodulin-Dependent Kinase II) and CN (Calcineurin). Once activated, the Wnt/ Ca^{2+} pathway promotes changes in cell fate, migration or tissue polarity (Girardi and Le Grand 2018; von Maltzahn et al. 2012).

Both Wnt/ β -catenin and non-canonical Wnt pathways are known to regulate several processes of skeletal myogenesis. During embryonic myogenesis, Wnts are secreted by the dorsal

neural tube (Wnt1, Wnt3a and Wnt4) and surface ectoderm (Wnt7a and Wnt6), coordinating the spatial-temporal expression of *Pax3/7* as well as that of the MRFs (Otto et al. 2006; Tajbakhsh et al. 1998; Yu et al. 2013). Besides coordinating gene transcription, Wnt canonical signaling is required for the somite to generate the dermomyotome and myotome as well as for the delamination of progenitor cells that migrate into the limb to compose the appendicular muscles (Hutcheson et al. 2009). Within the myotomes, non-canonical Wnt/PCP pathway mediated by Wnt11 is required for the proper orientation and elongation of myocytes (Gros et al. 2009). During fetal myogenesis, the Wnt/ β -catenin pathway modulates muscle mass deposition while fiber type specification involves both canonical and Wnt/ Ca^{2+} signaling (Kuroda et al. 2013; Chin et al. 1998). In addition, non-canonical Wnt signaling is necessary during fetal myogenesis for the clustering of acetylcholine receptors in the center of muscle fibers during the establishment of the neuromuscular junctions (Kuroda et al. 2013). In adult myogenesis, the Wnt/PCP pathway triggered by Wnt7A acts to balance self-renewal and differentiation of satellite cells, ensuring the maintenance of a reservoir of reparative satellite stem cells in the damaged muscle (Le Grand et al. 2009). In turn, excessive Wnt/ β -catenin signaling was shown to promote fibrosis in aged muscle (Brack et al. 2007).

Dact1 (aka Dapper1 and Frodo1) is a Dvl binding protein that modulates the Wnt/ β -catenin and Wnt/PCP signaling pathways (Cheyette et al. 2002). This protein acts as a context-dependent regulator of the Wnt/ β -catenin pathway that can both inhibit and activate this pathway, depending on its phosphorylation status (Cheyette et al. 2002; Gloy et al. 2002; Gao et al. 2008; Teran et al. 2009). Concerning the Wnt/PCP pathway, Dact1 acts with Dvl2 and SEC14-like and spectrin-type domains 1 (Sestd1) protein to control the trafficking and post-transcriptional modifications of VANGL planar cell polarity protein 2 (Vangl2), a transmembrane molecule that plays a crucial role in this pathway (Suriben et al. 2009; Yang et al. 2013). Possibly due to its ability to balance signaling levels of different Wnt pathways, Dact1 participates in several processes during development, including gastrulation, morphogenetic movements, somitogenesis, notochord and head development, and heart formation (Cheyette et al. 2002; Suriben et al. 2006; Hikasa and Sokol 2004; Brott and Sokol 2005). Accordingly, a series of posterior malformations occur in mice mutant for *Dact1*, which are related to increased

activity of the PCP protein Vangl2 in the posterior primitive streak (Suriben et al. 2009; Wen et al. 2010). In addition, mutations in human *Dact1* gene have been associated to neural tube defects (Shi et al. 2012) and to the Townes-Brocks syndrome (Webb et al. 2017), while changes in *Dact1* transcriptional activity are related to several types of cancer (Yuan et al. 2012; Yin et al. 2013; Li et al. 2017).

Previous studies have shown that *Dact1* is expressed in the presomitic mesoderm of mouse embryos, in phase with the Wnt inhibitor Axin2 (Suriben et al. 2006). This expression occurs just before the activation of *Pax7* in the newly formed somites. In addition, *Dact1* is expressed in the central dermomyotome of chicken embryos (Alvares et al. 2009), a region that is known to generate a population of muscle progenitors that gives rise to fetal myoblasts as well as to adult satellite cells (Gros et al. 2005). Despite the evidences suggesting that *Dact1* is involved in skeletal muscle development, a detailed expression profile of this molecule has not been established during myogenesis. Therefore, in this work we characterized *Dact1* gene and protein expression in two well established *in vitro* systems of myogenesis: the primary culture of chick myoblasts (Hirst and Marcelle 2015; Mermelstein et al. 2007) and culture of C2C12 murine myoblasts, a cell lineage derived from adult satellite cells (Yaffe and Saxel 1977). We also compared *Dact1* gene expression levels between healthy and diseased human muscle using the Gene Expression Omnibus (GEO) datasets. Finally, we performed *in silico* analyses to search for intrinsically disordered regions (IDRs) and to evaluate *Dact1* protein propensity to undergo liquid-liquid phase transition, given the importance of these features for the biological activity of *Dact1*.

Our results showed a wide expression of *Dact1* gene and protein in fetal muscles and during *in vitro* skeletal myogenesis of chicken and mouse and revealed a modulation of *Dact1* expression in specific human muscle diseases. In addition, we found that the *Dact1* protein contains long IDRs, as other proteins involved in controlling Wnt signaling, besides having a strong propensity to undergo liquid-liquid phase separation, a feature that explain the ability of *Dact1* to form nuclear aggregates in myogenic cells.

Materials and methods

Animal Procedures

Fertilized eggs from the Yamaguishi Organic Farm - Brazil were incubated at 38.5°C in a humidified atmosphere to obtain the 11-day-old (E11) chicken embryos used in this work. The procedures used for embryos harvesting and euthanasia were evaluated by the Ethics Committee for Animal Care and Use in Scientific Research from the University of Campinas (CEUA/UNICAMP) and received the approval number: 4236-1.

C57BL/6 mice for breeding were obtained at the Multidisciplinary Center for Biological Investigation on Laboratory Animal Science (CEMIB/UNICAMP). The procedures for animals' care, mating and euthanasia of pregnant females for collection of E14.5 mouse embryos were evaluated by the CEUA/UNICAMP and received the approval number 3780-1.

Primary Culture of Chicken Embryonic Myoblasts

Primary cultures of myogenic cells were prepared from breast muscles of E11 chick embryos. The protocol used was adapted from (Mermelstein et al. 2007), with the following modification: the suspension of cells isolated from tissue was pre-plated in 100 mm diameter culture dish for 25 min in order to reduce the amount of fibroblasts that are co-isolated with myoblasts. Cells were grown under a humidified 5% CO₂ atmosphere at 37°C and fed daily with fresh culture medium (Minimum Essential Medium Eagle with the addition of 10% horse serum, 0.05% chick embryo extract, 1% L-glutamine and 100 IU/mL penicillin and 100 µg/mL streptomycin).

C2C12 Culture

C2C12 cells were cultured on poly-D-lysine (Sigma-Aldrich)-coated coverslips in Dulbecco's modified Eagle High Glucose Medium (DMEM) containing 10% fetal bovine serum (FBS), 100 IU/mL penicillin and 100 µg/mL streptomycin in a humidified 5% CO₂ atmosphere. To induce differentiation, C2C12 cells were grown to sub-confluence and then shifted to medium containing 2% horse serum. Harvesting of C2C12 cells was done during proliferation

and days 0 (first day of differentiation induction), 3 and 5. All cell culture reagents were purchased from Thermo Fisher Scientific.

Histological Preparations and Microscopy

Small fragments of muscle were dissected from E11 chicken embryos (breast) or E14.5 mouse embryos (gastrocnemius). Samples fixation was carried out in 4% paraformaldehyde overnight at 4 °C. Tissue processing was performed according to standard histological procedures (Castelucci et al. 2018) prior to embedding either in paraffin (Paraplast-Sigma) or historesin (Leica Microsystems, Germany). Ticker (7 µm) or thinner (3 µm) sections were prepared for *in situ* hybridization (ISH) or staining with Giemsa, respectively. Stained sections were analyzed and documented under a Nikon Eclipse E800 microscope (Nikon Corporation, Japan).

***In situ* Hybridization**

ISH on paraffin sections (7 µm) of E11 chicken breast muscle was carried out as previously described (Sensiate et al. 2014). Antisense RNA probes for chicken *Dact1* and *Myogenin* are those described in (Alvares et al. 2009) and (Berti et al. 2015), respectively. Sections were analyzed and photographed under a Nikon Eclipse E800 microscope on 100x lens (Nikon Corporation).

Immunofluorescence in Fetal Muscle Tissue and Antibodies

Immunofluorescences were carried out in E11 chicken breast muscle (8 µm cryosections) and mouse E14.5 gastrocnemius muscle (7 µm paraffin sections). After deparaffinization, the same protocol was used for chicken and mouse samples as follows. The slides were incubated with blocking buffer solution (1% BSA, 0.1% Triton-X 100, 50 mM Glycine and 0.1 M PBS) for 30 min at room temperature and then with the primary antibodies (*Dact1*, dilution 1:300, ab51260 Abcam®; *Desmin*, dilution 1:300, D93F5 XP® *Cell Signaling technology*) at 4°C overnight. As negative control of reactions, the incubation step with the primary antibody was omitted. After washing, sections were incubated with the secondary antibody conjugated with Alexa Fluor® 488 (dilution 1:400, ab150077 Abcam®). Nuclei were stained using DAPI

(dilution 1:1000, sc-3598, Santa Cruz) and mounted with VECTASHIELD mounting medium (Vector Labs). Sections were observed and photographed using a Leica DM5500B fluorescence microscope on 20x and 40x lenses.

Immunofluorescence in Cells

Primary chick myoblast and C2C12 cells on coverslips were fixed in 100% acetone on ice. The immunofluorescence protocol and Dact1 and Desmin antibodies used in these experiments were the same described above. In addition, we used the β -catenin antibody C-18 (dilution 1:300, sc-1496, Santa Cruz) to perform the immunofluorescence assays in primary chick myoblasts. Cells were mounted with VECTASHIELD mounting medium (Vector Labs) and observed using a Laser Confocal Microscope LSM 780-NLO (Carl Zeiss, Germany) or Leica DM5500B.

Quantitative PCR (qRT-PCR)

Gene expression levels of *Dact1*, *Myogenin*, and *Myomaker* were analyzed in chick myoblasts and C2C12 cells. Total RNA was extracted using TRIzol Reagent (Invitrogen), and cDNAs were synthesized using RevertAid H Minus First Strand cDNA Synthesis kit (Fermentas), according to manufacturer's instructions. qRT-PCR reactions were carried out using SYBR Green (Applied Biosystems) in an Applied Biosystems 7300 thermocycler. Gene-specific primers (Invitrogen) are presented in Table 1. The 18S rRNA and GAPDH were used as housekeeping genes for chick myoblasts and C2C12 cells, respectively. The mathematical model $2^{-\Delta\Delta Ct}$ was used to quantify gene expression, considering day 1 (chick myoblasts) and proliferation (C2C12) as normalizers. To validate the real-time PCR assays, efficiency and dissociation curves were determined for each gene, and cDNA concentration and primer adjustments were performed. Three independent experiments were performed, and measurements of gene expression levels were determined using technical duplicates.

Statistical Analyses

Dact1, *Myogenin* and *Myomaker* gene expression analyses over time were done by ANOVA for repeated measures followed by Tukey's post-test. GraphPad Prism 5.0 (LLC, San Diego, CA, USA) was used for statistical calculations and graph design. Statistical significance was defined as $p < 0.05$ (*), $p < 0.01$ (**), and $p < 0.001$ (***). Data were presented as means $\pm$ standard error of the mean SEM (N=3).

Analysis of *Dact1* Gene Expression in Human Skeletal Muscle

Genotype-Tissue Expression (GTEx) Portal (<https://www.gtexportal.org/>) was used to determine *Dact1* gene expression levels in non-disease human skeletal muscle samples (dbGaP accession number phs000424.vN. on 05/01/2019). For gene-level quantifications, read counts and TPM values were produced with RNA-SeQC v1.1.8 (DeLuca et al. 2012).

GeoProfiles Analysis

The shinyGEO application was used to analyze *Dact1* gene expression in the Omnibus dataset GSE3307, containing Affymetrix mRNA profiles from human patient muscle biopsies using HG-U133A microarrays (Bakay et al. 2006; Dumas et al. 2016). p-values were calculated using a two-sample t-test for differential expression. Statistical significance was defined as $p < 0.05$.

Bioinformatics Analysis of *Dact1* protein

The sequences used for bioinformatic predictions were Human (Uniprot ID Q9NYF0); Mouse (Uniprot ID Q8R4A3) and Chick (Uniprot ID Q1G7H4). The presence of intrinsically disordered regions in the *Dact1* protein was investigated by the following disorder predictors: PONDR-FIT (Predictor of Intrinsically Disordered Amino Acids) (Xue et al. 2010), PONDR-VLXT (Romero et al. 2001), IUPred-long (long regions of intrinsic disorder) and IUPred-short (short regions of intrinsic disorder) (Dosztanyi 2018), PONDR-VSL2 and PONDR-VL3 (Peng et al. 2005), and PrDOS (Ishida and Kinoshita 2007). PONDR-FIT is available at <http://original.disprot.org/pondr-fit.php>; PONDR-VLXT, PONDR-VL3, and PONDR-VSL2 at

<http://www.pondr.com/>; IUPred2-long disorder and IUPred2-short disorder at <https://iupred2a.elte.hu/plotand> and PrDOS at <http://prdos.hgc.jp/cgi-bin/top.cgi>. The average disorder profile was obtained by calculating the mean of disorder reports from the seven computational tools. A score > 0.5 refers to amino acid residues in disordered regions whereas scores from 0.2 to 0.5 indicate residues in flexible segments. CIDER (Classification of Intrinsically Disordered Ensemble Regions) server (<http://pappulab.wustl.edu/CIDER/analysis/>) generated the plot of net charge along primary sequence, the sliding window size was 5 (Holehouse et al. 2017)). CatGRANULE algorithm (http://s.tartagliab.com/new_submission/catGRANULES) predicted liquid-liquid phase separation based on the following features: primary sequence composition, structural disorder and nucleic acid binding propensities (Bolognesi et al. 2016). Controls proteins for catGRANULE analyses were hNPM1 (Uniprot ID P06748), hTau40 (Uniprot ID P10636-6), hAlb (Uniprot ID P02768) and hSmac/Diablo (Uniprot ID Q9NR28).

Results

Dact1 expression in fetal breast muscle of chicken

We started by investigating Dact1 expression in E11 chicken breast muscle by *in situ* hybridization (ISH) and immunofluorescence (IF). Morphological analysis of this tissue was also performed for comparisons. Since the developing breast muscle is composed by different cell types, we used *Myogenin* as a molecular marker for myogenic progenitor cells in the ISH assays. We choose this marker because it is specifically expressed in the nuclei of differentiating myogenic cells (Zammit 2017). Desmin, an intermediate filament protein that is found only in muscle and endothelial cells (Capetanaki et al. 1997), was used as a marker for skeletal muscle differentiation in the IF assays.

Our analyses revealed that at E11 the chicken breast muscle is composed mainly by myotubes surrounded by fibroblast-like cells (Fig. 1A), which are either myoblasts or muscle fibroblasts (Yablonka-Reuveni et al. 1988). Cells juxtaposed to the myotubes probably are myoblasts (Fig. 1A, arrow), which approach the myotubes to fuse into them to generate multinucleated myofibers. *Dact1* mRNAs were identified both in the myotubes (Fig. 1B, asterisk) and in the cells surrounding them (Fig. 1B, arrow) in the chicken developing fetal muscle. *Myogenin* transcripts were detected only in the nuclei of myotubes (Fig. 1C, arrowhead). Dact1 protein showed the same distribution observed for Dact1 transcripts, being found both in myotubes and in cells adjacent to them (Fig. 1D, asterisk and arrow, respectively). Desmin was specifically expressed in the developing myotubes (Fig. 1E). These initial results confirmed that Dact1 participates in the fetal myogenesis of chicken.

Dact1 expression during *in vitro* differentiation of chick primary myoblasts

To further detail Dact1 expression profiles during chicken myogenesis, we quantified its gene expression levels and established its protein distribution along the five days of differentiation (D1 to D5) of primary myoblasts in culture. Besides *Dact1*, *Myogenin* and *Myomaker* mRNA levels were also determined by qRT-PCR as markers for differentiation and myoblast-myotube fusion (Millay et al. 2013), respectively. Along Dact1 and Desmin, we also

evaluated β -catenin cellular distribution by IF during myogenesis as an indicative of the Wnt canonical signaling activity.

Our results revealed that *Dact1* gene expression increases progressively throughout myogenesis, reaching higher levels from D4 of culture differentiation (Fig. 2A). This induction of *Dact1* expression begins when *Myogenin* reaches its highest level at D4 and, therefore, coincides with the onset of the myogenic differentiation. However, while *Myogenin* expression drops rapidly, *Dact1* expression remains elevated until D5 (Fig. 2A). *Myomaker* gene expression was upregulated at D3, revealing that the highest fusion activity of myoblasts occurs at this time in culture, one day before *Dact1* upregulation (Fig. 2A).

Regarding Dact1 protein, a subtle immunolabeling was observed in the myogenic cells from D1 to D3 (Fig. 2B). In these cells, Dact1 protein was found in the cytoplasm (Fig. 2B, red arrowheads) but also in the nuclei as punctate spots (Fig. 2B, white arrowheads). At the late stages of myogenic differentiation (D4 to D5), it was possible to observe that the cytoplasmic distribution of Dact1 changes, with this protein gradually assuming the highly organized and striated arrangement of sarcomeres (Fig. 2B, red arrow). Besides this cytoplasmic pattern, Dact1 forms conspicuous aggregates in the nuclei of myofibers (Fig. 2B, D5, white arrowhead). Similar to Dact1, β -catenin was also expressed throughout myogenesis (Fig. 3). However, while from D1 to D3 this protein was mostly localized in the cytoplasm (Fig. 3, red arrowheads), between D4 and D5 β -catenin was translocated to the nuclei of myotubes (Fig. 3, D5, white arrowhead), besides being observed in the plasma membrane (Fig. 3, yellow arrowheads). The translocation of β -catenin to the nuclear compartment indicates that the Wnt canonical pathway is activated from D4, concomitantly with Dact1 transcriptional induction and with the broad distribution of this protein in myotubes and myofibers. Therefore, at this point in myogenesis the role of Dact1 is not compatible with its function as an inhibitor of the Wnt/ β -catenin pathway, suggesting that this molecule performs another function at this stage.

In summary, our findings confirmed that *Dact1* is expressed during chicken fetal myogenesis *in vitro*, with higher expression levels from the differentiation step. In addition, we found that Dact1 is a nucleocytoplasmic protein located in all cells of the myogenic lineage, from myoblasts to myofibers. Interestingly, *Dact1* reaches its higher gene expression levels and its

protein displays a sarcomeric arrangement concurrently with active Wnt signaling, as determined by the nuclear distribution of β -catenin at late myogenesis.

Dact1 expression during mouse skeletal myogenesis

After determining that *Dact1* gene and protein are expressed during chicken *in vivo* and *in vitro* myogenesis, we set out to investigate whether this molecule also participates in the skeletal myogenesis of mouse. For that, we used the murine C2C12 myoblast cell line as a model system to track *Dact1* gene expression and protein distribution from proliferation to different days post-induction of differentiation (D0, D3 and D5). In addition, IF was carried out in the E14.5 gastrocnemius muscle. This day was chosen because in mouse the embryonic myogenesis finishes at E13.5, and at E14.5 the fetal myogenesis is ongoing (Zammit et al. 2008; Chal and Pourquié 2017), as evaluated in chicken.

Our qRT-PCR results showed that *Dact1* gene expression levels increased progressively during C2C12 differentiation (Fig. 4A), reaching its highest levels in comparison to proliferation at D3. *Myogenin* and *Myomaker* were both upregulated from D3, indicating that differentiation and fusion of myoblasts to the forming myotubes occur with greater intensity from this day on. Since *Dact1* upregulation was observed concomitantly with the induction of these molecular markers, we conclude that higher levels of *Dact1* gene expression are associated with the late phases of mouse myogenesis. Regarding Dact1 protein, we observed that during proliferation, this protein is diffusely distributed over the cytoplasm (Fig. 4B, red arrowheads) and nuclei of myoblasts. As differentiation progresses (D0-D5, Fig.4B), Dact1 maintains its dispersed distribution throughout the cytoplasm while progressively accumulates in the nuclei generating protein aggregates (Fig. 4B, D5, white arrowhead). Importantly, Dact1 protein does not display the striated pattern in the myofibers generated by the differentiation of C2C12 cells, as observed in the chicken myofibers derived from primary myoblasts. In the mouse E14.5 gastrocnemius (Fig. 4C), Dact1 is evenly distributed in the developing muscle tissue, showing that in mammals this protein also participates in the fetal myogenesis, as observed in chicken.

Our findings showed that *Dact1* is also expressed during mouse myogenesis both *in vitro* and *in vivo*, indicating an evolutionary conserved role for this protein during skeletal muscle development of amniotes.

***Dact1* gene expression in human myopathies**

After identifying that *Dact1* is expressed during chicken and mouse myogenesis, we sought to investigate whether this gene is expressed in human skeletal muscle. This was done by analyzing the gene expression database available on Genotype-Tissue Expression (GTEx) Portal. Furthermore, we investigated whether the expression of this gene was altered in dystrophies and other muscle diseases in human patients. For that, differential expression analysis was performed in the GSE3307 series, which contains expression profiles by microarray for 12 different muscle diseases (HG-U133A array) obtained from Children's National Medical Center (Bakay et al. 2006), totalizing 121 human samples. These analyses were done using shinyGEO, a web-based application for analyzing gene expression omnibus datasets (Dumas et al. 2016).

Dact1 expression in human skeletal muscle was confirmed by data source GTEx Analysis Release V7 (dBGap Accession phs000424.v7. p2), presenting low levels of Transcripts Per Million (TPM=1.020) from non-diseased skeletal muscle samples. Regarding the analysis of *Dact1* expression in human myopathies, we found that this gene is up-regulated in patients with Duchenne Muscular Dystrophy (DMD, OMIM #310200) compared to normal muscle samples (Table 2) with a 1.73-fold-change (FC) ($P=0.001$). In addition, samples from patients with Limb-Girdle Muscular Dystrophy type 2A (LGMD-2A, OMIM #253600) display *Dact1* up-regulation (FC=1.45, $P=0.017$). On the other hand, in Acute Quadriplegic Myopathy (AQM) muscle biopsies, *Dact1* is downregulated when compared to control samples (FC=0.77, $P=0.018$). The analysis of *Dact1* in the other nine muscle diseases in this array showed no significant differences in gene expression profiles. Details over the sample groups and *Dact1* differential expression data were presented in Table 2.

Our findings showed that *Dact1* is expressed at low levels in normal adult human skeletal muscle while its expression is modulated in some muscular dystrophies and other muscular diseases.

Search for IDRs in Dact1 protein and evaluation of phase-separation propensity

The fact that Dact1 forms aggregates in the nuclei of myogenic cells prompted us to carry out bioinformatic analyses to understand characteristics of this protein that could be related to this ability. Given that biomolecular condensates are generally mounted by proteins presenting intrinsically disordered regions (IDRs) (Boeynaems et al. 2018), we started by analyzing the intrinsic disorder characteristics of the primary structure of human, mouse and chicken Dact1. Analysis by PONDR-FIT (Xue et al. 2010) identified long IDRs conserved among all species (Fig. 5A). Human Dact1 had 72% of residues predicted to be within natively unfolded segments (score > 0.5), followed by 66% (mouse) and 65% (chick) (Fig. 5A). The human Dact1 was subsequently analyzed using other six disorder predictors (Dosztanyi et al. 2005; Ishida and Kinoshita 2007; Peng et al. 2005; Peng et al. 2006; Romero et al. 2001). The mean average disorder profile obtained from these analyses confirmed that IDRs encompass most Dact1 residues (Fig. 5B, black curve). Since scores between 0.2 and 0.5 denote flexible regions, we can conclude that Dact1 is mostly intrinsically disordered (Fig. 5B, black curve). Generally, fully intrinsically disordered proteins contain several charged amino acids residues in conjunction with few hydrophobic residues. To better characterize Dact1 intrinsic disorder, we used the CIDER algorithm (Holehouse et al. 2017) showed that human Dact1 contains well-mixed positive and negative amino acids (Fig. 5C). This feature is observed in domains that adopt random coil or globular conformation depending on the net charge (Das et al. 2015). Based on sequence similarity by BLAST alignment tool, Dact1 was shown to contain a coiled-coil domain (residues 92-156) which is also predicted by MULTICOIL server (Wolf et al. 1997). All these structural features pointed us to investigate liquid-liquid phase separation propensity of Dact1, the principle behind higher-order assemblies formation (Brangwynne et al. 2009; Boeynaems et al. 2018) using catGRANULE algorithm (Bolognesi et al. 2016). This analysis revealed a strong tendency for Dact1 liquid-liquid phase separation (scores ca. 1.0). Well characterized phase behavior of proteins that interact with nucleic acids such as human NPM1 and hTau40 were used as positive controls for catGRANULE analysis and showed similar results to Dact1 (Fig. 5D). Proteins such as human Albumin and human Smac, which represent well-structured

globular proteins and as reported to date, do not phase separate, showed low scores for catGRANULE phase separation.

Overall, our bioinformatic analysis revealed that the Dact1 protein of human and other vertebrates contains long IDRs. In addition, we showed that the Dact1 protein has a strong propensity to liquid-liquid phase separation, a feature that can facilitate the organization of the Dact1 aggregates observed in the nuclei of myogenic cells.

Discussion

Dact1 is a key regulator of the Wnt/ β -catenin and Wnt/PCP signaling pathways. In this work, we aimed to investigate its gene and protein expression in fetal muscles as well as during skeletal myogenesis using *in vitro* culture systems for both chicken primary and murine C2C12 myoblasts. In addition, we performed *in silico* analyses to investigate *Dact1* transcriptional modulation in human muscle diseases as well as to evaluate features of the Dact1 protein that may have functional implications in its biological role.

The Wnt modulator Dact1 is expressed along skeletal myogenesis of amniotes

Several molecules involved in Wnt signaling play critical roles on skeletal myogenesis and usually display a specific spatiotemporal expression. In the case of *Dact1*, we observed increased expression levels concurrently with *Myogenin* and *Myomaker* upregulation both in chicken and mouse *in vitro* myogenesis, indicating a possible role for Dact1 during skeletal muscle differentiation and/or myoblasts fusion. However, Dact1 protein was present in the cytoplasm and nucleus of cells throughout myogenic differentiation; therefore, it was not possible to associate its expression/activity with specific steps of myogenesis. Thus, how to hypothesize about the functions that Dact1 may have during skeletal myogenesis, based on its expression profile? The answer to this question may lie in the fact that Dact1 is able to convert from a repressor to an activator of the Wnt canonical pathway, in a context dependent manner. Interestingly, a transition from a state of low levels to another of high levels of Wnt/ β -catenin signaling occurs during fetal and reparative myogenesis, mimicked in this work by the chicken primary and mouse C2C12 myoblasts culture, respectively. For instance, the early steps of fetal myogenesis occur under low levels of Wnt canonical signaling and a premature and sustained activation of the Wnt/ β -catenin pathway drastically reduces the mass of skeletal muscles in mouse embryos (Kuroda et al. 2013). In addition, during reparative myogenesis the early activation and proliferation of satellite cells is coordinated by Notch signaling; however, a transition to a state of increased levels of Wnt/ β -catenin signaling is demanded to promote the formation of fusion-competent myoblasts (Tsivitsse 2010). Accordingly, an inhibition of Wnt canonical signaling during the early reparative myogenesis reduces the number and size of

myotubes formed after injury (Tsivitse 2010). In this context, it is possible that Dact1 converts from an inhibitor (early myogenesis) to an activator of the Wnt/ β -catenin pathway (late myogenesis). In its Wnt inhibitory state, Dact1 possibly acts by promoting Dvl ubiquitination and autophagic degradation as previously described (Ma et al. 2015). Accordingly, the ablation of Dact1 in mice mutant for the *Islr* (immunoglobulin superfamily containing leucine-rich repeat) gene, which encodes a protein that interacts and protects Dvl from autophagy, recovers the impaired differentiation phenotype shown by the *Islr* mutants during reparative myogenesis (Zhang et al. 2018). In parallel, Dact1 may also inhibit Wnt signaling by disrupting the interaction between β -catenin and LEF1 in the nucleus, acting to further reduce Wnt-dependent gene activity (Gao et al. 2008). The conversion from this inhibitory to an activator state possibly occur once Dact1 is phosphorylated by the casein kinase I δ/ϵ (CKI δ/ϵ), which promotes its dissociation from Dvl, causing an increase in Wnt/ β -catenin signaling (Teran et al. 2009). In addition, Dact1 may further promote Wnt canonical signaling by stabilizing p120-catenin, allowing this protein to sequester the transcriptional inhibitor Kaiso from promoters of Wnt target genes, increasing Wnt mediated transcription (Park et al. 2006). Therefore, the broad expression of Dact1 along myogenesis suggests that this molecule may contribute to the fine adjustments of Wnt canonical signaling by working both as inhibitor and activator of the Wnt/ β -catenin pathway at different steps of this process.

It is also important to consider that Dact1 may also contribute to skeletal myogenesis through the Wnt/PCP signaling pathway. In this sense, it is known that Dact1 collaborates with *Sestd1* to control post-translational changes in the Vangl2 protein, a central molecule of the PCP pathway (Yang and Cheyette 2013). In addition, the *Dact1* and *Sestd1* genes are part of a repertoire of target genes whose promoters are synergistically driven by the transcription factors MyoD and Six, which are necessary for myogenic fate acquisition of mouse embryonic fibroblasts (Santolini et al. 2016).

In addition to the cytoplasmic and nuclear distribution throughout myogenesis, which, as discussed, may be associated with the modulation of the Wnt signaling pathways, we also observed a well-organized sarcomeric arrangement of Dact1 protein in chicken myotubes *in vitro* but not in C2C12 myotubes. This difference may be due to the fact that the chick myoblast

primary culture is a more robust *in vitro* model of myogenesis, since in this system both myoblasts and fibroblasts are present, reproducing an environment that is similar to the *in vivo* muscle tissue condition. In contrast, the C2C12 mouse myoblast cell line is a pure population of myoblasts that needs serum withdrawal to differentiate (Denes et al. 2019). Accordingly, C2C12 myofibers usually have between 2-30 nuclei and rarely form sarcomeres, whereas in chick primary myogenic cultures the myofibers formed are much larger, display well-striated myofibrils, and can contain thousands of nuclei in a single multinucleated cell (Mermelstein et al. 2007; Possidonio et al. 2014). Having made these reservations, it is possible to hypothesize that Dact1 participates in the myofibrillogenesis during the late stages of skeletal myogenesis, at least in chicken. In fact, different authors have suggested that Dact1 contributes to cellular cytoskeleton arrangement in different contexts, through a mechanism involving cytoplasmic β -catenin accumulation (Hou et al. 2011; Hou et al. 2019).

Finally, it is important to mention that Dact1 was shown to be expressed in fetal muscles of chicken (breast) and mouse (gastrocnemius) *in vivo*, revealing a conserved expression during the fetal myogenesis of amniotes. Since fetal myogenesis is critical to build the bulk of skeletal muscle mass of each individual, as well as to establish the functional link between skeletal muscles and the nervous system by the establishment of the neuromuscular junctions, many questions about the possible roles of this molecule during fetal myogenesis remain to be answered.

***Dact1* transcription is modulated in human muscle myopathies**

An important finding of our study was the modulation of *Dact1* gene expression in DMD and LGMD-2A muscle dystrophies, and AQM. The characteristic lack of Dystrophin seen in Duchenne Muscular Dystrophy patients and in the *mdx* mouse model was shown to be compensated by utrophin up-regulation, a structural protein with similarities with Dystrophin and a potential clinical target (Hirst et al. 2005; Fairclough et al. 2013). Given that Dact1 has a N-terminal region similar to the leucine-rich coiled-coil domain of Dystrophin and Utrophin (Gloy et al. 2002), it is possible to hypothesize that *Dact1* up-regulation in DMD is part of a compensatory mechanism in muscle cells to offset the absence of Dystrophin. In LGMD-2A

patients, the absence of calpain 3 is responsible for muscular dystrophy by altering the cleavage of substrates and impairing muscle homeostasis (Richard et al. 1995), which affect structural proteins balance. Given that calpain 3 has multiple roles in skeletal muscle and that the mechanisms that lead to muscular dystrophy in mutated individuals are complex and still not fully understood (Huang et al. 2008), it is not possible yet to establish a precise correlation between increased *Dact1* expression in patients with LGMD-2A dystrophy. AQM is an acquired myopathy that commonly affects bedridden patients and is characterized by severe muscle weakness due to the loss of the motor protein myosin and myosin-associated thick filament proteins and to alterations in muscle membrane excitability (Larsson et al. 2000). Interestingly, muscle wasting in AQM has been associated with the activation of different proteolytic pathways (Showalter and Engel 1997; Helliwell et al. 1998; Di Giovanni et al. 2004). Considering that *Dact1* is expressed at basal levels in muscle cells, it is possible that the proteolysis activation initiated present in AQM also affects *Dact1*, resulting in its downregulation through mechanisms of transcription inhibition resulting from degradation of other regulatory proteins, including *Dact1* regulators. However, the biological significance of these *in silico* findings need to be tested by functional assays, and also in animal models of muscular dystrophies.

***Dact1* is an intrinsically disordered protein with tendency to liquid-liquid phase transition**

In our work, we showed that the *Dact1* protein of human, mouse and chicken display long IDRs. IDRs are protein regions that lack persistent secondary or tertiary structure under native conditions, a feature known to play a crucial role in protein-protein interactions (Harnoš et al. 2018). Interestingly, IDRs are found in most components of the canonical Wnt pathway, including Axin, CKI-a, GSK-3b, APC, and β -catenin. For instance, Axin uses its long disordered region to assemble the various components of the β -catenin destruction complex into a single complex (Xue et al. 2012). In addition, proteins involved in autophagy, as is the case of *Dact1* that is capable of inducing Dvl autophagy, usually display IDRs (Ma et al. 2015; Popelka 2020). Therefore, it is possible that the *Dact1* IDRs contribute to its function in modulating the Wnt

pathway and promoting Dvl autophagy, possibly facilitating your interaction with specific molecules depending on the cellular context.

In addition, our bioinformatic analysis revealed that Dact1 has a strong propensity to liquid-liquid phase separation, a process involved in higher-order assemblies or membraneless organelles formation (Brangwynne et al. 2009; Boeynaems et al. 2018). This trend may be associated with the formation of nuclear granules observed in cells throughout myogenesis. As higher-order assemblies are very dynamic, rapidly assembling and disassembly upon cellular stimuli (Boeynaems et al. 2018), Dact1 may be a Wnt modulator that switches the “on-off” states of Wnt signaling by its phase separation property. The association between the structural features of Dact1 protein and the modulation of the Wnt/ β -catenin pathway deserves further investigation.

Conclusions

Dact1 is expressed in chicken and mouse fetal skeletal muscle as well as during chicken primary and murine C2C12 myoblasts differentiation. Dact1 protein is located both in the cytoplasm and nuclei of myoblasts, myocytes, myotubes and myofibers, as well as in the sarcomeres of myofibers. This broad expression in cells of the myogenic lineage indicates that Dact1 may exert different functions on specific steps of myogenesis, possibly acting as a bivalent modulator (inhibitor and activator) of the Wnt/ β -catenin pathway. Importantly, *Dact1* transcription is modified in DMD, LGMD-2A and AQM muscle diseases, which envisages a possible pathophysiological role of *Dact1* in muscular disorders. Despite the numerous hypothesis raised in this study about the mechanisms by which Dact1 modulates Wnt signaling during myogenesis, functional studies are necessary to determine how this molecule works during skeletal muscle formation, growth, and regeneration as well as to understand the participation of Dact1 in human muscle diseases.

Acknowledgments

This study was financed by the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior – CAPES (Finance Code 001) and the Conselho Nacional de Desenvolvimento Científico e Tecnológico - CNPq (grants 141441/2018-6 and 158979/2018-4). We also thank INFABIC UNICAMP laboratories, Dr. Henrique Marques Barbosa de Souza, Dr. Paulo Pinto Joazeiro, Dr. Hernandes Faustino de Carvalho, Dr. Alexandre Leite Rodrigues de Oliveira for providing access to their equipment and assistance.

Conflict of Interest

The authors have declared no conflict of interests.

References

- Alvares LE, Winterbottom FL, Sobreira DR, Xavier J, Schubert FR, Dietrich S (2009) Chicken Dapper Genes Are Versatile Markers for Mesodermal Tissues, Embryonic Muscle Stem Cells, Neural Crest Cells, and Neurogenic Placodes. *Dev Dyn* 238 (5):1166-1178. doi:10.1002/dvdy.21950
- Asfour HA, Allouh MZ, Said RS (2018) Myogenic regulatory factors: The orchestrators of myogenesis after 30 years of discovery. *Exp Biol Med (Maywood)* 243 (2):118-128. doi:10.1177/1535370217749494
- Bakay M, Wang Z, Melcon G, Schiltz L, Xuan J, Zhao P, Sartorelli V, Seo J, Pegoraro E, Angelini C, Shneiderman B, Escolar D, Chen YW, Winokur ST, Pachman LM, Fan C, Mandler R, Nevo Y, Gordon E, Zhu Y, Dong Y, Wang Y, Hoffman EP (2006) Nuclear envelope dystrophies show a transcriptional fingerprint suggesting disruption of Rb-MyoD pathways in muscle regeneration. *Brain* 129 (Pt 4):996-1013. doi:10.1093/brain/awl023
- Barbalho SM, Vieira Prado Neto E, de Alvares Goulart R, Bechara MD, Federighi Baisi Chagas E, Audi M, Guissoni Campos LM, Landgraf Guiger E, Leoni Buchain R, Buchain DV, Cressoni Araujo A (2020) Myokines: a descriptive review. *J Sports Med Phys Fitness*. doi:10.23736/S0022-4707.20.10884-3
- Berti F, Nogueira JM, Wohrle S, Sobreira DR, Hawrot K, Dietrich S (2015) Time course and side-by-side analysis of mesodermal, pre-myogenic, myogenic and differentiated cell markers in the chicken model for skeletal muscle formation. *J Anat* 227 (3):361-382. doi:10.1111/joa.12353
- Boeynaems S, Alberti S, Fawzi NL, Mittag T, Polymenidou M, Rousseau F, Schymkowitz J, Shorter J, Wolozin B, Van Den Bosch L, Tompa P, Fuxreiter M (2018) Protein Phase Separation: A New Phase in Cell Biology. *Trends Cell Biol* 28 (6):420-435. doi:10.1016/j.tcb.2018.02.004
- Bolognesi B, Lorenzo Gotor N, Dhar R, Cirillo D, Baldrighi M, Tartaglia GG, Lehner B (2016) A Concentration-Dependent Liquid Phase Separation Can Cause Toxicity upon Increased Protein Expression. *Cell Rep* 16 (1):222-231. doi:10.1016/j.celrep.2016.05.076
- Brack AS, Conboy MJ, Roy S, Lee M, Kuo CJ, Keller C, Rando TA (2007) Increased Wnt signaling during aging alters muscle stem cell fate and increases fibrosis. *Science* 317 (5839):807-810. doi:10.1126/science.1144090
- Brangwynne CP, Eckmann CR, Courson DS, Rybarska A, Hoege C, Gharakhani J, Julicher F, Hyman AA (2009) Germline P granules are liquid droplets that localize by controlled dissolution/condensation. *Science* 324 (5935):1729-1732. doi:10.1126/science.1172046
- Brott BK, Sokol SY (2005) A vertebrate homolog of the cell cycle regulator Dbf4 is an inhibitor of Wnt signaling required for heart development. *Dev Cell* 8 (5):703-715. doi:10.1016/j.devcel.2005.02.012
- Capetanaki Y, Milner DJ, Weitzer G (1997) Desmin in muscle formation and maintenance: knockouts and consequences. *Cell Struct Funct* 22 (1):103-116

Castelucci BG, Consonni SR, Rosa VS, Sensiate LA, Delatti PCR, Alvares LE, Joazeiro PP (2018) Time-dependent regulation of morphological changes and cartilage differentiation markers in the mouse pubic symphysis during pregnancy and postpartum recovery. *PLoS One* 13 (4):e0195304. doi:10.1371/journal.pone.0195304

Chal J, Pourquié O (2017) Making muscle: skeletal myogenesis in vivo and in vitro. *Development* 144 (12):2104-2122. doi:10.1242/dev.151035

Cheyette BNR, Waxman JS, Miller JR, Takemaru KI, Sheldahl LC, Khlebtsova N, Fox EP, Earnest T, Moon RT (2002) Dapper, a Dishevelled-associated antagonist of beta-catenin and JNK signaling, is required for notochord formation. *Developmental Cell* 2 (4):449-461. doi:10.1016/s1534-5807(02)00140-5

Chin ER, Olson EN, Richardson JA, Yang Q, Humphries C, Shelton JM, Wu H, Zhu W, Bassel-Duby R, Williams RS (1998) A calcineurin-dependent transcriptional pathway controls skeletal muscle fiber type. *Genes Dev* 12 (16):2499-2509. doi:10.1101/gad.12.16.2499

Das RK, Ruff KM, Pappu RV (2015) Relating sequence encoded information to form and function of intrinsically disordered proteins. *Curr Opin Struct Biol* 32:102-112. doi:10.1016/j.sbi.2015.03.008

Davis TA, Fiorotto ML (2009) Regulation of muscle growth in neonates. *Curr Opin Clin Nutr Metab Care* 12 (1):78-85. doi:10.1097/MCO.0b013e32831cef9f

DeLuca DS, Levin JZ, Sivachenko A, Fennell T, Nazaire MD, Williams C, Reich M, Winckler W, Getz G (2012) RNA-SeQC: RNA-seq metrics for quality control and process optimization. *Bioinformatics* 28 (11):1530-1532. doi:10.1093/bioinformatics/bts196

Denes LT, Riley LA, Mijares JR, Arboleda JD, McKee K, Esser KA, Wang ET (2019) Culturing C2C12 myotubes on micromolded gelatin hydrogels accelerates myotube maturation. *Skelet Muscle* 9 (1):17. doi:10.1186/s13395-019-0203-4

Di Giovanni S, Molon A, Broccolini A, Melcon G, Mirabella M, Hoffman EP, Servidei S (2004) Constitutive activation of MAPK cascade in acute quadriplegic myopathy. *Ann Neurol* 55 (2):195-206. doi:10.1002/ana.10811

Dosztanyi Z (2018) Prediction of protein disorder based on IUPred. *Protein Sci* 27 (1):331-340. doi:10.1002/pro.3334

Dosztanyi Z, Csizmok V, Tompa P, Simon I (2005) IUPred: web server for the prediction of intrinsically unstructured regions of proteins based on estimated energy content. *Bioinformatics* 21 (16):3433-3434. doi:10.1093/bioinformatics/bti541

Dumas J, Gargano MA, Dancik GM (2016) shinyGEO: a web-based application for analyzing gene expression omnibus datasets. *Bioinformatics* 32 (23):3679-3681. doi:10.1093/bioinformatics/btw519

Fairclough RJ, Wood MJ, Davies KE (2013) Therapy for Duchenne muscular dystrophy: renewed optimism from genetic approaches. In: Nat Rev Genet, vol 14. vol 6. England, pp 373-378. doi:10.1038/nrg3460

Gao X, Wen J, Zhang L, Li X, Ning YH, Meng AM, Chen YG (2008) Dapper1 Is a Nucleocytoplasmic Shuttling Protein That Negatively Modulates Wnt Signaling in the Nucleus. Journal of Biological Chemistry 283 (51):35679-35688. doi:10.1074/jbc.M804088200

Girardi F, Le Grand F (2018) Wnt Signaling in Skeletal Muscle Development and Regeneration. Prog Mol Biol Transl Sci 153:157-179. doi:10.1016/bs.pmbts.2017.11.026

Gloy J, Hikasa H, Sokol SY (2002) Frodo interacts with Dishevelled to transduce Wnt signals. Nat Cell Biol 4 (5):351-357. doi:10.1038/ncb784

Gros J, Manceau M, Thome V, Marcelle C (2005) A common somitic origin for embryonic muscle progenitors and satellite cells. Nature 435 (7044):954-958. doi:10.1038/nature03572

Gros J, Serralbo O, Marcelle C (2009) WNT11 acts as a directional cue to organize the elongation of early muscle fibres. Nature 457 (7229):589-593. doi:10.1038/nature07564

Hargreaves M, Spriet LL (2020) Skeletal muscle energy metabolism during exercise. Nat Metab. doi:10.1038/s42255-020-0251-4

Harnoš J, Ryneš J, Víšková P, Foldynová-Trantírková S, Bajard-Ešner L, Trantírek L, Bryja V (2018) Analysis of binding interfaces of the human scaffold protein AXIN1 by peptide microarrays. J Biol Chem 293 (42):16337-16347. doi:10.1074/jbc.RA118.005127

Helliwell TR, Wilkinson A, Griffiths RD, McClelland P, Palmer TE, Bone JM (1998) Muscle fibre atrophy in critically ill patients is associated with the loss of myosin filaments and the presence of lysosomal enzymes and ubiquitin. Neuropathol Appl Neurobiol 24 (6):507-517

Hikasa H, Sokol SY (2004) The involvement of Frodo in TCF-dependent signaling and neural tissue development. Development 131 (19):4725-4734. doi:10.1242/dev.01369

Hirst CE, Marcelle C (2015) The avian embryo as a model system for skeletal myogenesis. Results Probl Cell Differ 56:99-122. doi:10.1007/978-3-662-44608-9_5

Hirst RC, McCullagh KJ, Davies KE (2005) Utrophin upregulation in Duchenne muscular dystrophy. Acta Myol 24 (3):209-216

Holehouse AS, Das RK, Ahad JN, Richardson MO, Pappu RV (2017) CIDER: Resources to Analyze Sequence-Ensemble Relationships of Intrinsically Disordered Proteins. Biophys J 112 (1):16-21. doi:10.1016/j.bpj.2016.11.3200

Hou J, Li EM, Shen JH, Qing Z, Wu ZY, Xu XE, Wu JY, Qiao H, Shen J, Guo MZ, Xu LY (2011) Cytoplasmic HDPR1 is involved in regional lymph node metastasis and tumor development via beta-

catenin accumulation in esophageal squamous cell carcinoma. *J Histochem Cytochem* 59 (7):711-718. doi:10.1369/0022155411409941

Hou J, Yue Y, Hu B, Xu G, Su R, Lv L, Huang J, Yao J, Guan Y, Wang K, Wu Z (2019) DACT1 Involvement in the Cytoskeletal Arrangement of Cardiomyocytes in Atrial Fibrillation by Regulating Cx43. *Braz J Cardiovasc Surg* 34 (6):711-722. doi:10.21470/1678-9741-2019-0033

Huang Y, de Morree A, van Remoortere A, Bushby K, Frants RR, den Dunnen JT, van der Maarel SM (2008) Calpain 3 is a modulator of the dysferlin protein complex in skeletal muscle. *Hum Mol Genet* 17 (12):1855-1866. doi:10.1093/hmg/ddn081

Hutcheson DA, Zhao J, Merrell A, Halder M, Kardon G (2009) Embryonic and fetal limb myogenic cells are derived from developmentally distinct progenitors and have different requirements for beta-catenin. *Genes Dev* 23 (8):997-1013. doi:10.1101/gad.1769009

Ishida T, Kinoshita K (2007) PrDOS: prediction of disordered protein regions from amino acid sequence. *Nucleic Acids Res* 35 (Web Server issue):W460-464. doi:10.1093/nar/gkm363

Kuroda K, Kuang S, Taketo MM, Rudnicki MA (2013) Canonical Wnt signaling induces BMP-4 to specify slow myofibrogenesis of fetal myoblasts. *Skelet Muscle* 3 (1):5. doi:10.1186/2044-5040-3-5

Larsson L, Li X, Edstrom L, Eriksson LI, Zackrisson H, Argentini C, Schiaffino S (2000) Acute quadriplegia and loss of muscle myosin in patients treated with nondepolarizing neuromuscular blocking agents and corticosteroids: mechanisms at the cellular and molecular levels. *Crit Care Med* 28 (1):34-45

Le Grand F, Jones AE, Seale V, Scime A, Rudnicki MA (2009) Wnt7a activates the planar cell polarity pathway to drive the symmetric expansion of satellite stem cells. *Cell Stem Cell* 4 (6):535-547. doi:10.1016/j.stem.2009.03.013

Le Grand F, Rudnicki MA (2007) Skeletal muscle satellite cells and adult myogenesis. *Curr Opin Cell Biol* 19 (6):628-633. doi:10.1016/j.ceb.2007.09.012

Li RN, Liu B, Li XM, Hou LS, Mu XL, Wang H, Linghu H (2017) DACT1 Overexpression in type I ovarian cancer inhibits malignant expansion and cis-platinum resistance by modulating canonical Wnt signalling and autophagy. *Sci Rep* 7 (1):9285. doi:10.1038/s41598-017-08249-7

Ma B, Liu B, Cao W, Gao C, Qi Z, Ning Y, Chen YG (2015) The Wnt Signaling Antagonist Dapper1 Accelerates Dishevelled2 Degradation via Promoting Its Ubiquitination and Aggregate-induced Autophagy. *J Biol Chem* 290 (19):12346-12354. doi:10.1074/jbc.M115.654590

Mermelstein CS, Portilho DM, Mendes FA, Costa ML, Abreu JG (2007) Wnt/beta-catenin pathway activation and myogenic differentiation are induced by cholesterol depletion. *Differentiation* 75 (3):184-192. doi:10.1111/j.1432-0436.2006.00129.x

Millay DP, O'Rourke JR, Sutherland LB, Bezprozvannaya S, Shelton JM, Bassel-Duby R, Olson EN (2013) Myomaker is a membrane activator of myoblast fusion and muscle formation. *Nature* 499 (7458):301-305. doi:10.1038/nature12343

Otto A, Schmidt C, Patel K (2006) Pax3 and Pax7 expression and regulation in the avian embryo. *Anat Embryol (Berl)* 211 (4):293-310. doi:10.1007/s00429-006-0083-3

Park JI, Ji H, Jun S, Gu D, Hikasa H, Li L, Sokol SY, McCrea PD (2006) Frd1 links Dishevelled to the p120-catenin/Kaiso pathway: distinct catenin subfamilies promote Wnt signals. *Dev Cell* 11 (5):683-695. doi:10.1016/j.devcel.2006.09.022

Peng K, Radivojac P, Vucetic S, Dunker AK, Obradovic Z (2006) Length-dependent prediction of protein intrinsic disorder. *BMC Bioinformatics* 7:208. doi:10.1186/1471-2105-7-208

Peng K, Vucetic S, Radivojac P, Brown CJ, Dunker AK, Obradovic Z (2005) Optimizing long intrinsic disorder predictors with protein evolutionary information. *J Bioinform Comput Biol* 3 (1):35-60. doi:10.1142/s0219720005000886

Popelka H (2020) Dancing while self-eating: Protein intrinsic disorder in autophagy. *Prog Mol Biol Transl Sci* 174:263-305. doi:10.1016/bs.pmbts.2020.03.002

Possidonio AC, Soares CP, Portilho DM, Midlej V, Benchimol M, Butler-Browne G, Costa ML, Mermelstein C (2014) Differences in the expression and distribution of flotillin-2 in chick, mice and human muscle cells. *PLoS One* 9 (8):e103990. doi:10.1371/journal.pone.0103990

Richard I, Broux O, Allamand V, Fougerousse F, Chiannilkulchai N, Bourg N, Brenguier L, Devaud C, Pasturaud P, Roudaut C, et al. (1995) Mutations in the proteolytic enzyme calpain 3 cause limb-girdle muscular dystrophy type 2A. *Cell* 81 (1):27-40

Romero P, Obradovic Z, Li X, Garner EC, Brown CJ, Dunker AK (2001) Sequence complexity of disordered protein. *Proteins* 42 (1):38-48. doi:10.1002/1097-0134(20010101)42:1<38::aid-prot50>3.0.co;2-3

Santolini M, Sakakibara I, Gauthier M, Ribas-Aulinas F, Takahashi H, Sawasaki T, Mouly V, Concordet JP, Defossez PA, Hakim V, Maire P (2016) MyoD reprogramming requires Six1 and Six4 homeoproteins: genome-wide cis-regulatory module analysis. *Nucleic Acids Res* 44 (18):8621-8640. doi:10.1093/nar/gkw512

Sensate LA, Sobreira DR, Da Veiga FC, Peterlini DJ, Pedrosa AV, Rirsch T, Joazeiro PP, Schubert FR, Collares-Buzato CB, Xavier-Neto J, Dietrich S, Alvares LE (2014) Dact gene expression profiles suggest a role for this gene family in integrating Wnt and TGF-beta signaling pathways during chicken limb development. *Dev Dyn* 243 (3):428-439. doi:10.1002/dvdy.23948

Shi Y, Ding Y, Lei YP, Yang XY, Xie GM, Wen J, Cai CQ, Li H, Chen Y, Zhang T, Wu BL, Jin L, Chen YG, Wang HY (2012) Identification of novel rare mutations of DACT1 in human neural tube defects. *Hum Mutat* 33 (10):1450-1455. doi:10.1002/humu.22121

Showalter CJ, Engel AG (1997) Acute quadriplegic myopathy: analysis of myosin isoforms and evidence for calpain-mediated proteolysis. *Muscle Nerve* 20 (3):316-322. doi:10.1002/(sici)1097-4598(199703)20:3<316::aid-mus8>3.0.co;2-e

Suriben R, Fisher DA, Cheyette BN (2006) Dact1 presomitic mesoderm expression oscillates in phase with Axin2 in the somitogenesis clock of mice. *Dev Dyn* 235 (11):3177-3183. doi:10.1002/dvdy.20968

Suriben R, Kivimae S, Fisher DA, Moon RT, Cheyette BN (2009) Posterior malformations in Dact1 mutant mice arise through misregulated Vangl2 at the primitive streak. *Nat Genet* 41 (9):977-985. doi:10.1038/ng.435

Tajbakhsh S, Borello U, Vivarelli E, Kelly R, Papkoff J, Duprez D, Buckingham M, Cossu G (1998) Differential activation of Myf5 and MyoD by different Wnts in explants of mouse paraxial mesoderm and the later activation of myogenesis in the absence of Myf5. *Development* 125 (21):4155-4162

Teran E, Branscomb AD, Seeling JM (2009) Dpr Acts as a molecular switch, inhibiting Wnt signaling when unphosphorylated, but promoting Wnt signaling when phosphorylated by casein kinase Idelta/epsilon. *PLoS One* 4 (5):e5522. doi:10.1371/journal.pone.0005522

Tsivitsse S (2010) Notch and Wnt signaling, physiological stimuli and postnatal myogenesis. *Int J Biol Sci* 6 (3):268-281. doi:10.7150/ijbs.6.268

von Maltzahn J, Chang NC, Bentzinger CF, Rudnicki MA (2012) Wnt signaling in myogenesis. *Trends Cell Biol* 22 (11):602-609. doi:10.1016/j.tcb.2012.07.008

Webb BD, Metikala S, Wheeler PG, Sherpa MD, Houten SM, Horb ME, Schadt EE (2017) Heterozygous Pathogenic Variant in DACT1 Causes an Autosomal-Dominant Syndrome with Features Overlapping Townes-Brocks Syndrome. *Hum Mutat* 38 (4):373-377. doi:10.1002/humu.23171

Wen J, Chiang YJ, Gao C, Xue H, Xu J, Ning Y, Hodes RJ, Gao X, Chen YG (2010) Loss of Dact1 disrupts planar cell polarity signaling by altering dishevelled activity and leads to posterior malformation in mice. *J Biol Chem* 285 (14):11023-11030. doi:10.1074/jbc.M109.085381

Wolf E, Kim PS, Berger B (1997) MultiCoil: a program for predicting two- and three-stranded coiled coils. *Protein Sci* 6 (6):1179-1189. doi:10.1002/pro.5560060606

Xue B, Dunbrack RL, Williams RW, Dunker AK, Uversky VN (2010) PONDR-FIT: a meta-predictor of intrinsically disordered amino acids. *Biochim Biophys Acta* 1804 (4):996-1010. doi:10.1016/j.bbapap.2010.01.011

Xue B, Dunker AK, Uversky VN (2012) The roles of intrinsic disorder in orchestrating the Wnt-pathway. *J Biomol Struct Dyn* 29 (5):843-861. doi:10.1080/073911012010525024

Yablonka-Reuveni Z, Anderson SK, Bowen-Pope DF, Nameroff M (1988) Biochemical and morphological differences between fibroblasts and myoblasts from embryonic chicken skeletal muscle. *Cell Tissue Res* 252 (2):339-348. doi:10.1007/bf00214376

Yaffe D, Saxel O (1977) Serial passaging and differentiation of myogenic cells isolated from dystrophic mouse muscle. *Nature* 270 (5639):725-727

Yang X, Cheyette BN (2013) SEC14 and spectrin domains 1 (Sestd1) and Dapper antagonist of catenin 1 (Dact1) scaffold proteins cooperatively regulate the Van Gogh-like 2 (Vangl2) four-pass transmembrane protein and planar cell polarity (PCP) pathway during embryonic development in mice. *J Biol Chem* 288 (28):20111-20120. doi:10.1074/jbc.M113.465427

Yang X, Fisher DA, Cheyette BN (2013) SEC14 and Spectrin Domains 1 (Sestd1), Dishevelled 2 (Dvl2) and Dapper Antagonist of Catenin-1 (Dact1) co-regulate the Wnt/Planar Cell Polarity (PCP) pathway during mammalian development. *Commun Integr Biol* 6 (6):e26834. doi:10.4161/cib.26834

Yin X, Xiang T, Li L, Su X, Shu X, Luo X, Huang J, Yuan Y, Peng W, Oberst M, Kelly K, Ren G, Tao Q (2013) DACT1, an antagonist to Wnt/beta-catenin signaling, suppresses tumor cell growth and is frequently silenced in breast cancer. *Breast Cancer Res* 15 (2):R23. doi:10.1186/bcr3399

Yu Y, Qi L, Wu J, Wang Y, Fang W, Zhang H (2013) Kindlin 2 regulates myogenic related factor myogenin via a canonical Wnt signaling in myogenic differentiation. *PLoS One* 8 (5):e63490. doi:10.1371/journal.pone.0063490

Yuan G, Wang C, Ma C, Chen N, Tian Q, Zhang T, Fu W (2012) Oncogenic function of DACT1 in colon cancer through the regulation of beta-catenin. *PLoS One* 7 (3):e34004. doi:10.1371/journal.pone.0034004

Zammit PS (2017) Function of the myogenic regulatory factors Myf5, MyoD, Myogenin and MRF4 in skeletal muscle, satellite cells and regenerative myogenesis. *Semin Cell Dev Biol* 72:19-32. doi:10.1016/j.semcdb.2017.11.011

Zammit PS, Cohen A, Buckingham ME, Kelly RG (2008) Integration of embryonic and fetal skeletal myogenic programs at the myosin light chain 1f/3f locus. *Dev Biol* 313 (1):420-433. doi:10.1016/j.ydbio.2007.10.044

Zhang K, Zhang Y, Gu L, Lan M, Liu C, Wang M, Su Y, Ge M, Wang T, Yu Y, Li L, Li Q, Zhao Y, Yu Z, Wang F, Li N, Meng Q (2018) Islr regulates canonical Wnt signaling-mediated skeletal muscle regeneration by stabilizing Dishevelled-2 and preventing autophagy. *Nat Commun* 9 (1):5129. doi:10.1038/s41467-018-07638-4

TABLES

Table 1. Evaluation of *Dact1* expression in normal muscle biopsy specimens (Control; N=16) performed by shinyGEO compared to diseased samples. Fold change (FC) and P-value were calculated by a two-sample t-test. Significance P <0.01 (**) and P<0.001 (***) are indicated by comparison to control samples.

Disease	N	Fold Change
Acute quadriplegic myopathy**	5	0.77 (P=0.018)
Amyotrophic lateral sclerosis	9	0.97 (P=0.826)
Becker muscular dystrophy	5	1.05 (P=0.729)
Duchenne muscular dystrophy***	10	1.73 (P=0.001)
Emery Dreifuss muscular dystrophy (autosomal dominant Lamin A/C form)	4	1.58 (P=0.218)
Emery Dreifuss muscular dystrophy (X linked recessive Emerin form)	4	1.35 (P=0.111)
Facioscapulohumeral muscular dystrophy	14	1.03 (P=0.752)
Hereditary spastic paraplegia	4	0.79 (P=0.387)
Juvenile dermatomyositis	21	0.99 (P=0.904)
Limb-girdle muscular dystrophy type 2A**	10	1.45 (P=0.017)
Limb-girdle muscular dystrophy type 2B	10	1.52 (P=0.130)
Limb-girdle muscular dystrophy type 2I	7	0.99 (P=0.942)

Table 2. Gene-specific primers used in qRT-PCR assays. *Gg*=*Gallus gallus*, *Mm*=*Mus musculus*.

Species	Genbank ID	Symbol	Forward primer	Reverse primer
Gg	723789	<i>Dact1</i>	5'-GCCCTTTGGAGGCAACACT-3'	5'-GGTGGACATCTGCAACGACA-3'
	100861561	<i>RN18s</i>	5'-CGAAAGCATTTGCCAAGAAT-3'	5'-GGCATCGTTTATGGTCGC-3'
	374004	<i>MYOG</i>	5'-CGGAGGCTGAAGAAGGTGAA-3'	5'-CGGTCCTCTGCCTGGTCAT-3'
	777472	<i>TMEM8C</i> (<i>Myomaker</i>)	5'-TGGGTGTCCCTGATGGC-3'	5'-CCCGATGGGTCCTGAGTAG-3'
Mm	14433	<i>Gapdh</i>	5'-TGCACCACCAACTGCTTAGC-3'	5'-GGCATGGACTGTGGTCATGAG-3'
	59036	<i>Dact1</i>	5'-TGTTTCGTCCGAGTGTCTGG-3'	5'-ATCAAGGAGCGAGTTGACGG-3'
	17928	<i>MYOG</i>	5'-GGCAATGCACTGGAGTTCG-3'	5'-AGCCGCGAGCAAATGATC-3'
	66139	<i>TMEM8C</i> (<i>Myomaker</i>)	5'-ATCGCTACCAAGAGGCGTT-3'	5'-CACAGCACAGACAAACCAGG-3'

Figures & Legends

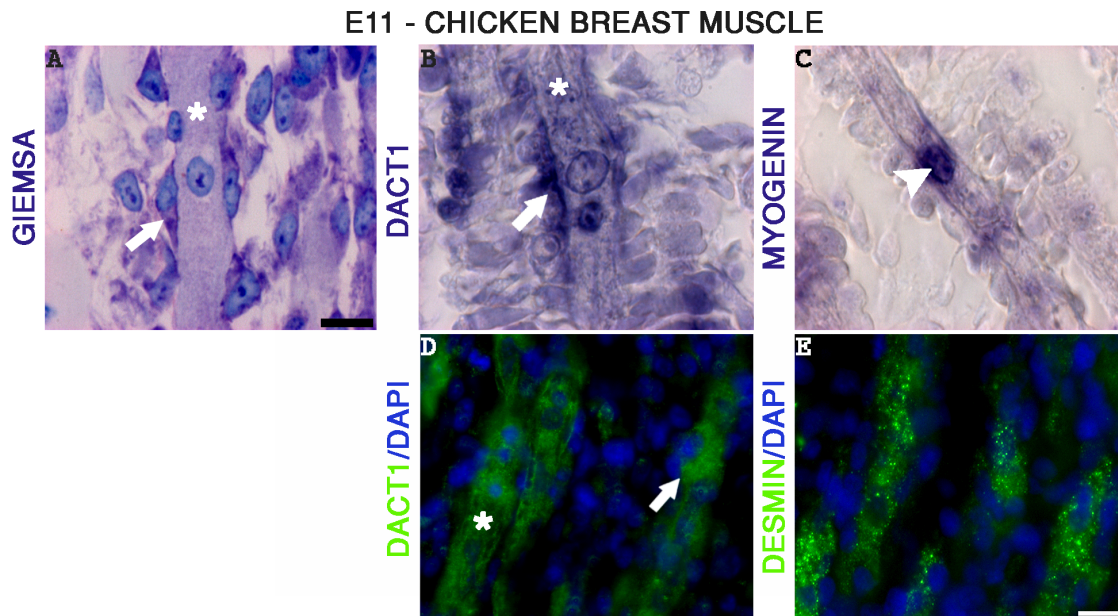


Fig. 1. Gene expression and protein distribution of *Dact1* in chicken fetal muscle. (A) Giemsa-stained breast muscle of E11 chicken embryo. (B, C) *Dact1* and *Myogenin* transcripts distribution as established by ISH. (D, E) *Dact1* and Desmin protein distribution as determined by immunofluorescence. Myotubes were indicated by asterisks; cells adjacent to myotubes (mainly myoblasts and fibroblasts) were indicated by arrows; the *Myogenin*⁺ nucleus of a myotube was indicated by an arrowhead. Nuclei were stained with DAPI (blue). Scale bars: (A-C) 10 μ m; (D-E) 30 μ m.

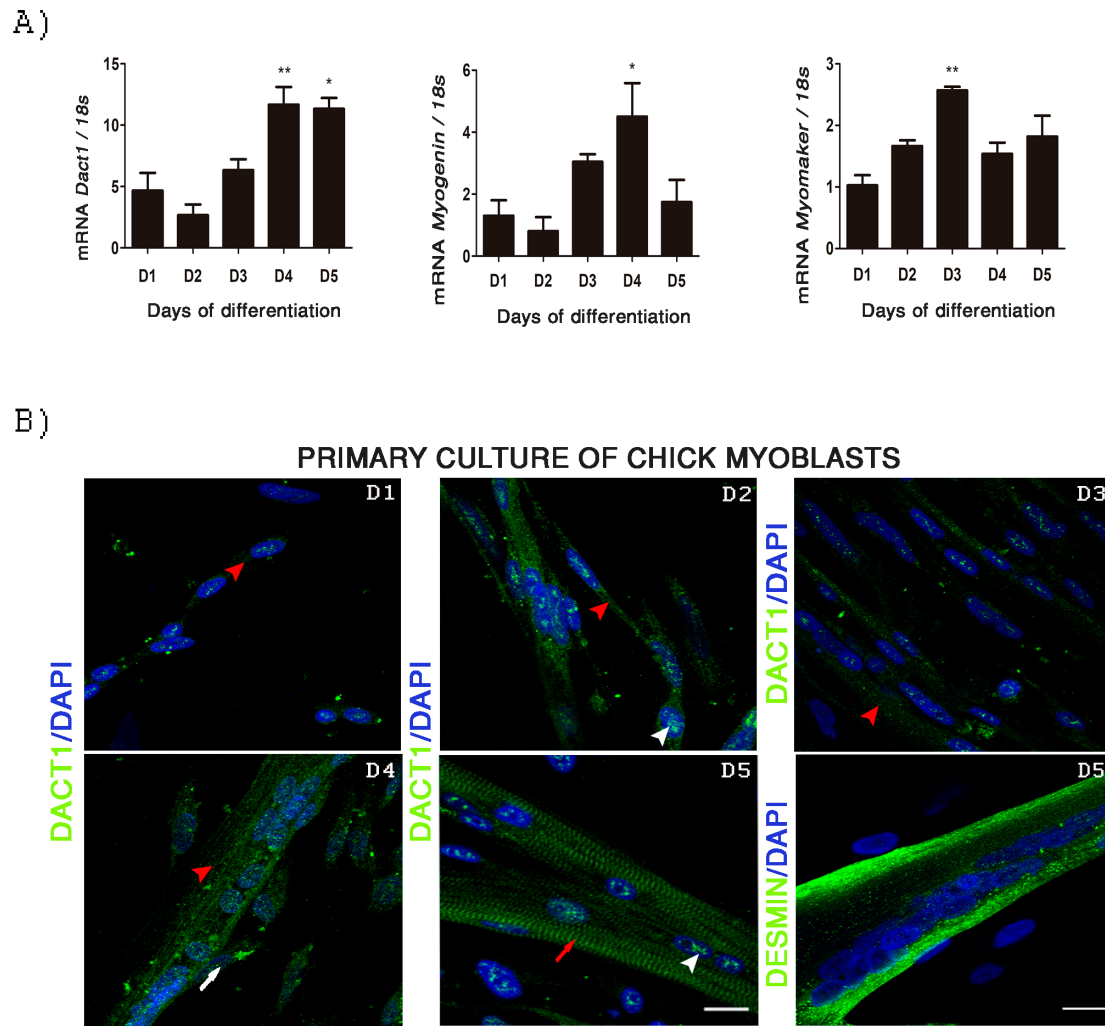


Fig. 2. Gene expression and protein distribution of Dact1 during chicken primary myoblasts differentiation *in vitro*. (A) The relative gene expression profiles of *Dact1*, *Myogenin*, and *Myomaker* were determined by qRT-PCR. 18S rRNA was used as the endogenous control gene. Significance * $P < 0.05$ and ** $P < 0.01$ are indicated in comparison to culture Day 1 (D1). Data represent mean $\pm$ SEM from three independent experiments. (B) Immunolocalization of Dact1 (green) at culture Days 1 to 5 (D1-D5) and Desmin at D5. Cytoplasmic (red arrowheads), nuclear (white arrowhead) and sarcomeric distribution (red arrow) of Dact1 protein were indicated. A Dact1+ myoblast juxtaposed to a myofiber was indicated (D4, white arrow). Desmin was used as a marker of muscle differentiation. Nuclei were stained with DAPI (blue). Scale bars: 20 μ m.

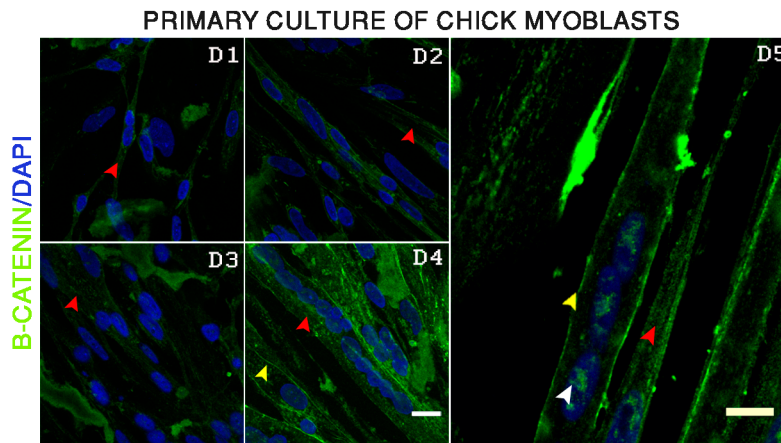


Fig. 3. Distribution of the β -catenin protein during chicken primary myoblasts differentiation *in vitro*. β -catenin immunolocalization (green) at culture Days 1 to 5 (D1-D5). Protein location in the cytoplasm (red arrowheads), plasma membrane (yellow arrowheads) and nucleus (D5, white arrowhead) were indicated. Nuclei were stained with DAPI (blue). Scale bars: 10 μ m.

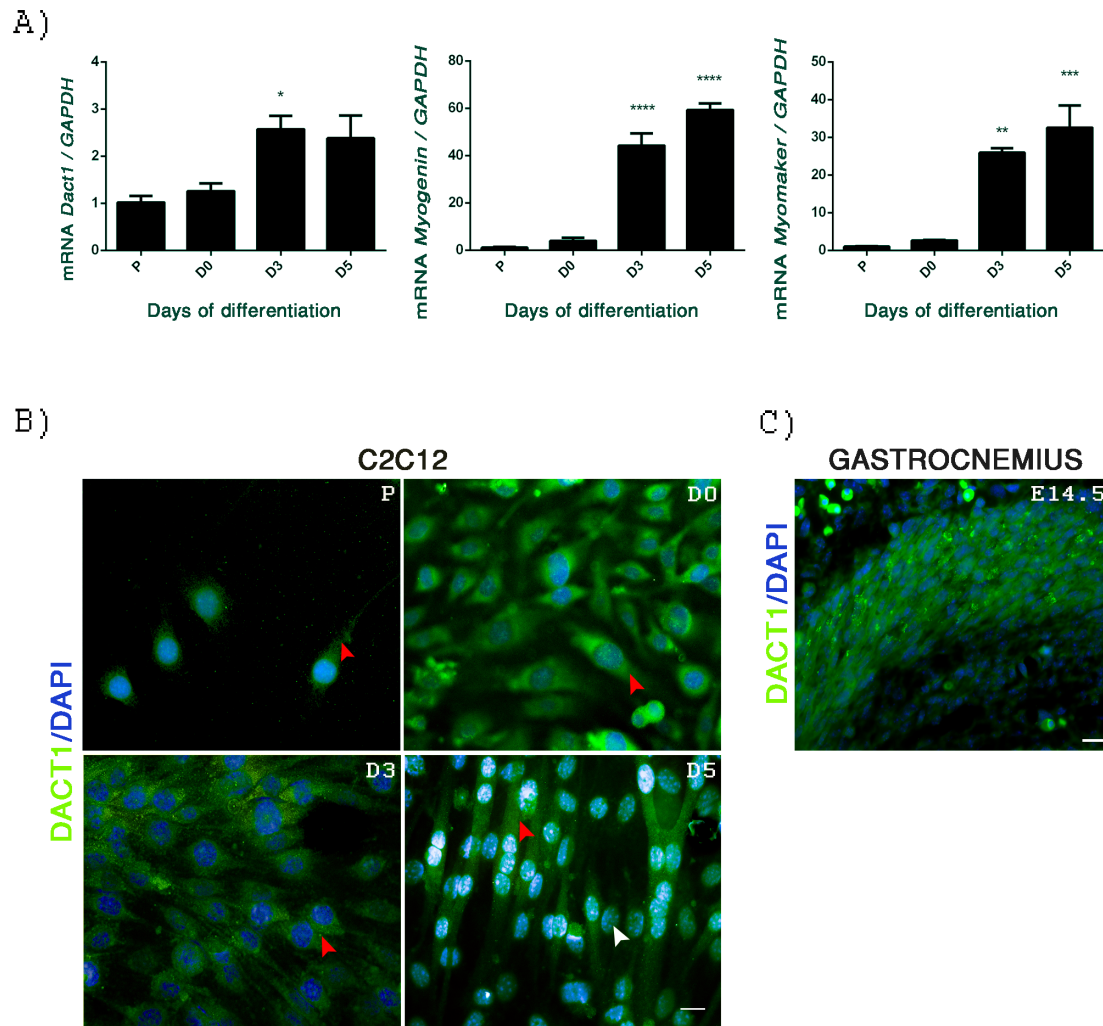


Fig. 4. Gene expression and protein distribution of *Dact1* during C2C12 myogenic differentiation and in mouse fetal muscle. (A) The relative gene expression profiles of *Dact1*, *Myogenin*, and *Myomaker* were determined by qRT-PCR. GAPDH was used as the endogenous control gene. Significance *P < 0.05, **P < 0.01, and ***P < 0.001 are indicated in comparison to C2C12 proliferation (P). Data represent mean $\pm$ SEM from three independent experiments. (B) Immunolocalization of *Dact1* (green) at proliferation (P) and D0, D3 and D5 after induction of C2C12 differentiation. *Dact1* displays cytoplasmic (red arrowheads) and nuclear distribution as discrete dots (D5, white arrowhead). (C) *Dact1* immunolocalization (green) in the developing gastrocnemius muscle of E14.5 mouse embryo. Nuclei were stained with DAPI (blue). Scale bar: 20 μ m.

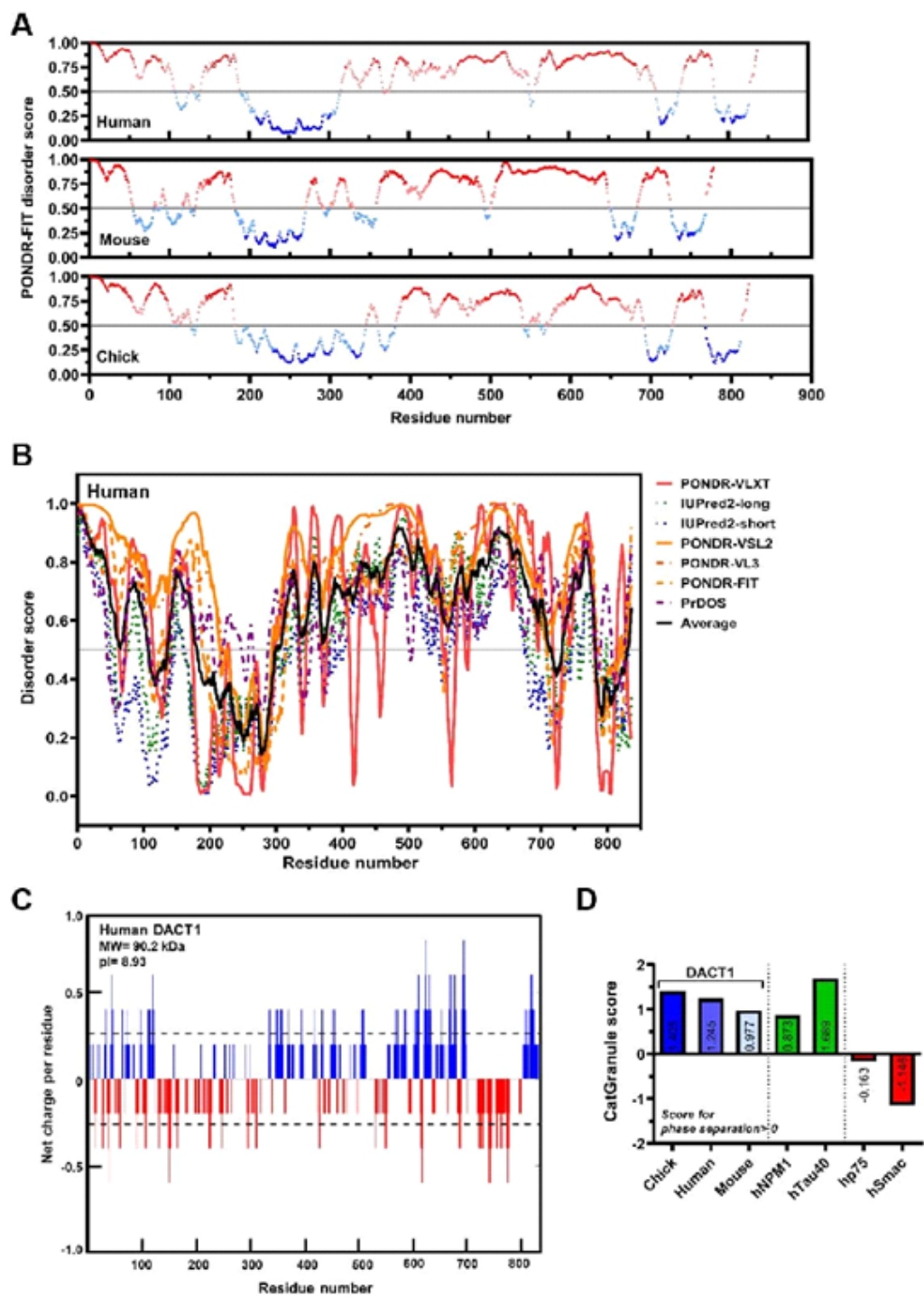
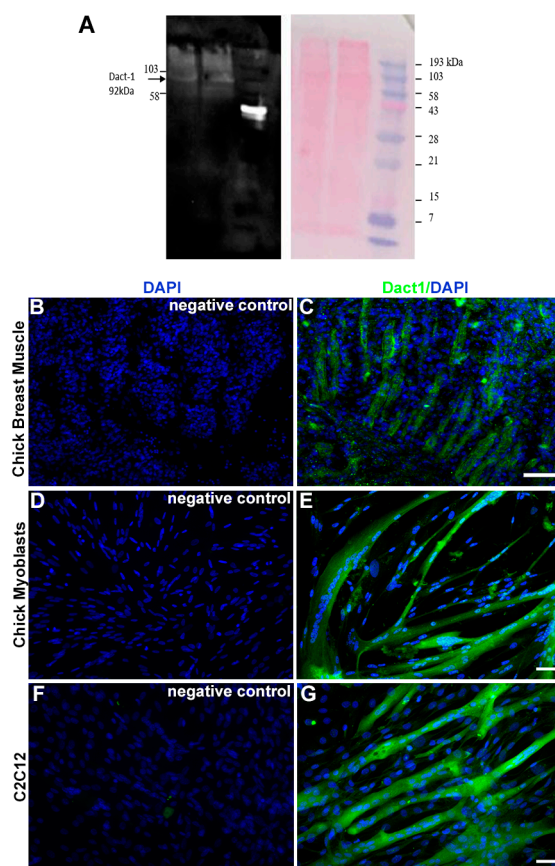


Fig. 5. Evaluation of Dact1 intrinsic disorder and propensity for phase separation. (A) PONDR-FIT analysis of Dact1 disordered protein regions in human, mouse and chicken. (B) Human Dact1 primary structure analysis by seven disorder predictors, displayed as curves in different colors/patterns. Black curve indicates the average disorder calculated among all algorithms. Scores above 0.5 represent disordered amino acids/regions and values between 0.2 to 0.5 indicate

flexible regions. (C) Analysis of net charge per residue distribution for human Dact1 using CIDER platform. (D) Overall scores for liquid-liquid phase separation for chicken, human and mouse Dact1 based on catGRANULE (score > 0 indicates phase separation; score > 1 indicates strong propensity). Scores of Dact1 homologs (blue bars) were compared with those of hNPM1 and hTau40 (green bars, positive controls), and with hAlb and hSmac/Diablo scores (red bars, negative controls).



Supplemental Fig. 1. Validation of the Dact1 antibody for chicken tissues (A) Lysates of primary myoblasts isolated from fetal chicken muscle were denatured and separated by SDS PAGE, transferred to nitrocellulose and blotted with a polyclonal antibody against Dact1 (ab51260 Abcam®). A molecular weight marker was used as control (1610318 Bio-Rad®). The expected 92kDa Dact1 protein was indicated by an arrow, as estimated by the molecular weight marker. (B-C) Immunofluorescence assays with breast muscle sections from E11 chicken embryo comparing the negative control of reaction (no primary antibody) with the reaction containing

the Dact1 antibody (green), respectively. (D-E) Immunofluorescence assays at Day 5 of chicken primary myoblasts differentiation, comparing the negative control of reaction (no primary antibody) with the reaction containing the Dact1 antibody, respectively. (F-G) Immunofluorescence assays with C2C12 myoblasts at Day 5 of differentiation, comparing the negative control of reaction (no primary antibody) with the reaction containing the Dact1 antibody, respectively. Scale bars: 20 μm .

4.3 DISCUSSÃO

Neste estudo realizamos a caracterização da expressão gênica e distribuição da proteína Dact1 ao longo da diferenciação das células miogênicas C2C12 em paralelo à caracterização em cultura primária de mioblastos de galinha. Em nossos ensaios observamos níveis de expressão aumentados de Dact1 concomitantemente à superexpressão de *MyoG* e *Myomaker* na miogênese tanto de camundongo, quanto de galinha. Isto aponta para um possível papel de Dact1 durante a diferenciação do músculo esquelético e/ou fusão de mioblastos. Importaneamente, nossos achados mostram uma conservação dos padrões de expressão de *Dact1* em mamíferos e aves, sugerindo que este gene é necessário para o desenvolvimento da musculatura esquelética dos vertebrados amniotos. No entanto, a ampla distribuição da proteína Dact1 tanto no citoplasma quanto no núcleo das células ao longo da diferenciação miogênica não permite diagnosticar com precisão a função desta proteína em relação a sinalização Wnt, que participa na miogênese esquelética e geralmente exibe uma expressão espaço-temporal específica. A nossa hipótese é que Dact1 seja capaz de se converter de um repressor em um ativador da via Wnt canônica, a depender do contexto. Isto pois, uma transição de um estado de baixos níveis para outro de altos níveis de sinalização Wnt/ β -catenina ocorre durante a miogênese fetal e reparativa, como reproduzido na cultura primária de mioblastos de galinha e nas células C2C12, respectivamente. Por exemplo, as etapas iniciais da miogênese fetal ocorrem sob baixos níveis de sinalização canônica Wnt e uma ativação prematura e sustentada da via Wnt/ β -catenina reduz drasticamente a massa dos músculos esqueléticos em embriões de camundongo (Kuroda et al., 2013). Já durante a miogênese reparativa, a ativação precoce e a proliferação de células satélite são coordenadas pela sinalização Notch, mas a transição para um estado de níveis aumentados de sinalização Wnt/ β -catenina é necessária para promover a formação de mioblastos competentes para a fusão (Tsivitse, 2010).

Assim, é possível que Dact1 se converta de um inibidor (miogênese inicial) em um ativador da via Wnt/ β -catenina (miogênese tardia). Em seu estado inibitório de Wnt, Dact1 possivelmente atua promovendo a ubiquitinação de Dvl levando a sua degradação por autofagia (Ma et al., 2015). Ou então, Dact1 também poderia inibir a sinalização Wnt interrompendo a

interação entre β -catenina e LEF1 no núcleo, agindo para reduzir ainda mais a atividade de genes-alvo de Wnt (Gao et al., 2008). A conversão deste estado inibitório para um ativador possivelmente ocorre uma vez que Dact1 é fosforilado pela caseína quinase 1 δ/ϵ (CK1 δ/ϵ), que promove sua dissociação de Dvl, causando um aumento na sinalização Wnt/ β -catenina (Teran et al., 2009). Em adição, Dact1 pode aumentar a sinalização Wnt canônica ao estabilizar a p120-catenina, permitindo que esta proteína sequestre o inibidor da transcrição (KAISO) de promotores de genes-alvo desta via, aumentando a transcrição mediada por Wnt (Park et al., 2006).

As análises *in silico* da proteína Dact1 de humanos, camundongos e galinhas demonstrou que esta proteína exibe IDRs longos. IDRs são regiões de proteína que não possuem estruturas secundária ou terciária, uma característica conhecida por desempenhar um papel crucial nas interações proteína-proteína (Harnoř et al., 2018), e são encontrados na maioria dos componentes da via Wnt canônica, incluindo Axin, CK1- α , GSK-3 β , APC e β -catenina. Além disso, proteínas envolvidas na autofagia, como é o caso de Dact1 que é capaz de induzir a autofagia de Dvl, costumam apresentar IDRs (Ma et al., 2015; Popelka, 2020). Portanto, é possível que os IDRs em Dact1 contribuam com sua função na modulação da via Wnt e na promoção da autofagia de Dvl, possivelmente facilitando sua interação com moléculas específicas dependendo do contexto celular. Em adição, nossas análises de bioinformática revelaram que Dact1 tem uma forte propensão à separação de fase líquido-líquido, um processo envolvido na organização de moléculas e na formação de organelas sem membranas (Boeynaems et al., 2018; Brangwynne et al., 2009). Essa tendência pode estar associada à formação dos grânulos nucleares que foram observados nas células ao longo da miogênese. Como esta organização de moléculas e consequente formação dos grânulos são muito dinâmicos, estas estruturas são rapidamente arranjadas e desarranjadas mediante estímulos celulares (Boeynaems et al., 2018). Esta característica de Dact1 sugere que ele pode ser um modulador Wnt que alterna os estados "liga-desliga" desta sinalização devido a sua propriedade de separação de fase.

Finalmente, outro achado importante de nosso estudo é a modulação da expressão de *Dact1* nas distrofias musculares DMD e LGMD-2A e na AQM. A distrofia muscular de

Duchenne (DMD) é um distúrbio degenerativo causado por mutações no gene da distrofina ligada ao cromossomo X, afetando 1 em cada 3.500-5.000 recém-nascidos do sexo masculino (Lim et al., 2017). Os pacientes com DMD sofrem de perda muscular progressiva com sintomas clínicos detectados ainda na primeira infância. No músculo esquelético, o complexo distrofina-glicoproteína está localizado dentro da membrana ao longo das fibras musculares e nas junções miotendíneas e neuromusculares. A deficiência de distrofina resulta na ruptura da ligação desse complexo que conecta membrana e matriz extracelular, levando ao rompimento do sarcolema (Suzuki et al., 1994). Isto causa um influxo de cálcio nas células musculares, a ativação de proteases dependentes deste cálcio e a degeneração muscular durante a contração (Gillis, 1996). A carência característica de distrofina observada em pacientes com DMD, bem como no camundongo *mdx*, é compensada pela regulação positiva da utrofina (Deconinck et al., 1997). Dado que a proteína *Dact1* tem uma região semelhante à distrofina e utrofina (Gloy et al., 2002), é possível que a regulação positiva de *Dact1* em DMD seja parte de um mecanismo compensatório em células musculares para suprir a ausência de distrofina e que nesta situação ele desempenhe um papel estrutural nas células musculares.

Em pacientes portadores da LGMD-2A, a ausência de calpaína 3 é responsável pela distrofia muscular por alterar a clivagem de substratos (Richard et al., 1995), que afetam o equilíbrio de proteínas estruturais. No entanto, como a calpaína 3 tem múltiplas funções no músculo esquelético e os mecanismos que levam à distrofia muscular em indivíduos mutados não são totalmente compreendidos (Huang et al., 2008), ainda não é possível determinar uma correlação precisa entre aumento da expressão de *Dact1* expressão nesses pacientes. Uma outra patologia em que *Dact1* tem sua expressão modulada é a AQM, uma miopatia que comumente afeta pacientes acamados e é caracterizada pela fraqueza muscular severa devido à perda da proteína motora miosina e de proteínas de filamentos espessos associadas à miosina e também a alterações na excitabilidade da membrana muscular (Larsson et al., 2000). A perda de massa muscular na AQM tem sido associada à ativação de diferentes vias proteolíticas (Di Giovanni et al., 2004; Helliwell et al., 1998; Showalter and Engel, 1997). Considerando que *Dact1* é expresso em níveis basais nas células musculares, é possível que a ativação da proteólise característica da AQM também afete *Dact1*, resultando em sua regulação negativa e repercutindo no seu possível

papel estrutural, como levantado no caso da DMD. No entanto, o significado biológico desses achados *in silico* precisa ser testado por ensaios funcionais e também em modelos animais para distrofias musculares.

4.4 CONCLUSÕES

Dact1 é expresso no músculo esquelético fetal de camundongo e de galinha, bem como durante a diferenciação de mioblastos primários destes dois modelos experimentais. A proteína *Dact1* está localizada no citoplasma e nos núcleos dos mioblastos, miócitos, miotubos e miofibras, como também nos sarcômeros das miofibras. Esta ampla expressão em células da linhagem miogênica indica que *Dact1* pode exercer diferentes funções em etapas específicas da miogênese, possivelmente atuando como um modulador bivalente (inibidor e ativador) da via Wnt/ β -catenina. Importaneamente, a transcrição de *Dact1* é modificada nas doenças musculares DMD, LGMD-2A e AQM, o que prevê um possível papel fisiopatológico de *Dact1* em doenças musculares, sugerindo que este possa ser um potencial alvo terapêutico. Apesar das inúmeras hipóteses levantadas neste estudo sobre os mecanismos pelos quais *Dact1* modula a sinalização Wnt durante a miogênese, estudos funcionais são necessários para determinar como essa molécula funciona durante a formação, crescimento e regeneração do músculo esquelético, bem como para entender a participação de *Dact1* em doenças musculares em humanos.

5. CAPÍTULO 2 – IDENTIFICAÇÃO E CARACTERIZAÇÃO DE ELEMENTOS CIS-REGULADORES

5.1 INTRODUÇÃO

O desenvolvimento dos seres vivos, da fertilização ao nascimento, e a manutenção da homeostase na fase pós-natal dependem de um refinado controle da expressão espaço-temporal dos genes. Os genes são modulados por elementos reguladores da transcrição gênica, denominados elementos cis-reguladores, os quais podem atuar como promotores, *enhancers*, silenciadores ou isoladores gênicos. Via de regra, os genes que participam do desenvolvimento são regulados por vários elementos cis-reguladores distintos, organizados de forma modular, de tal maneira que diferentes elementos controlam a atividade gênica em contextos específicos. Assim, um mesmo gene pode contribuir em diferentes processos do desenvolvimento, sob a coordenação de elementos cis-reguladores específicos. Esta estrutura modular assegura, ainda, que havendo uma mutação em um elemento cis-regulador específico, a expressão do gene será afetada apenas em um contexto particular, enquanto todos os demais aspectos de sua expressão são preservados.

Dada a importância do estudo da expressão e da regulação gênica de genes multifuncionais, tal como *Dact1*, neste trabalho, inicialmente, buscamos caracterizar a expressão gênica e proteica de *Dact1* na miogênese esquelética em células C2C12, um papel até então não caracterizado, buscando também correlacionar a modulação deste gene com miopatias em humanos. Neste estudo identificamos a participação de *Dact1*/Dact1 no processo miogênico e também alterações na sua expressão gênica em distrofias musculares, conforme apresentado no capítulo 1. À luz destes achados, nosso objetivo foi identificar sequências reguladoras da expressão de *Dact1* no desenvolvimento embrionário e na fase pós-natal que pudessem também atuar de maneira contexto-específica na miogênese e/ou no tecido muscular.

Para isto, realizamos uma análise sistemática do locus de *Dact1*, inicialmente comparando este locus no humano com o correspondente em camundongo e galinha, por meio do navegador genômico *ECR Browser*. Por esta análise identificamos vários elementos não

codificadores conservados (CNEs). Os CNEs identificados foram, então, sobrepostos com o dados fornecidos pelo *Ensembl Regulatory Build* (versão GRCh38), que contém as localizações e informações sobre regiões promotoras, regiões flangeadoras de promotores e *enhancers*, inferidas a partir de ensaios de cromatina (DNase-seq), ensaios de modificação de histonas e ensaios de ligação de fator de transcrição (sequenciamento de imunoprecipitação da cromatina) em vários tipos celulares (Zerbino et al., 2015). Para todos os elementos reguladores que se sobrepõem ou que são adjacentes aos CNEs, foi compilada uma lista de tipos celulares nos quais estes elementos são ativos ou reprimidos. Embora nem todos os CNEs se sobrepusessem a reguladores previamente caracterizados no *Ensembl*, todos estes foram analisados quanto aos seus potenciais sítios de ligação para fatores de transcrição, dado que estes podem ainda, potencialmente, corresponder a elementos cis-reguladores não caracterizados experimentalmente. O compilado de sítios de ligação para fatores de transcrição (TFBSs) preditos pelo *MultiTF* para cada CNE foi utilizado para montagem de redes de interação proteína-proteína, visando mapear associações físicas ou funcionais conhecidas entre estas proteínas, relacionando os contextos de interação aos processos, aos tipos celulares e às doenças humanas em que *Dact1* possa estar implicado.

Após identificarmos um CNE com potencial de regulação em contextos relacionados ao desenvolvimento e patologias do músculo esquelético, bem como na regulação metabólica relacionada ao aporte nutricional, um papel já descrito na literatura para *Dact1*, realizamos uma análise de dados de *Hi-C* em músculo humano disponíveis no *Gene Expression Omnibus* (GEO) para validarmos o potencial de interação deste elemento selecionado com o gene *Dact1*. Este CNE (CNE4) foi então validado experimentalmente por ensaios de transfecção com construção repórter, por ter características de elemento regulador no contexto miogênico (mioblastos murinos C2C12) e/ou adipogênico (pré-adipócitos murinos imortalizados 9W). Em seguida, selecionamos organismos-modelo (camundongo e rato), animais domésticos (cão e gato) e também aqueles de interesse econômico (boi, ovelha e cabra) para, por meio do alinhamento múltiplo de sequências, localizarmos e compararmos as sequências ortólogas deste CNE em outros grupos de vertebrados amniotos. Subsequentemente, foi realizada uma análise grupo-específica de TFBSs e de interação proteína-proteína para dois grandes grupos de animais

(*Primates* e *Bovidae*), buscando contextualizar a possível atividade do CNE4 para cada um dos grupos. Todos os achados deste estudo serão apresentados no capítulo 2.

5.2 MATERIAIS E MÉTODOS

5.2.1 Identificação de CNEs no locus de *Dact1*

A busca por elementos conservados evolutivamente (CNEs) foi realizada no intervalo genômico *KIAA0586-Dact1-Daam1*, que compreende toda a extensão entre *Dact1* e os seus genes flanqueadores *KIAA0586* e *Daam1*. Esta busca foi feita utilizando o navegador genômico *ECR Browser* (<https://ecrbrowser.dcode.org>) (Ovcharenko et al., 2004), com os seguintes parâmetros para identificação de CNEs: mínimo de 200 pb e 70% de identidade de sequência, tendo como base o genoma humano (hg19). Após a identificação dos CNEs, as sequências individuais de nucleotídeos para cada CNE e espécie foi coletada e catalogada no formato FASTA em arquivos texto para uso nas análises subsequentes realizadas nesta tese.

5.2.2 Comparação dos CNEs presentes no locus de *Dact1* com elementos reguladores validados pelo *Ensembl*

A fim de verificar a correspondência entre os CNEs previamente identificados em 5.2.1 e os elementos regulatórios funcionais anotados no genoma humano (Zerbino et al., 2015), a sequência de nucleotídeos humana de cada CNE, no formato FASTA, foi utilizada como *bait* para realizar uma busca com o algoritmo de alinhamento BLAT (*BLAST-like alignment tool*) do *browser* genômico *Ensembl* (<https://www.ensembl.org/>). Para cada *hit* encontrado na busca, foi analisado o *location-based display* do *Ensembl* para confirmar a posição da sequência identificada no intervalo *KIAA0586-Dact1-Daam1*, bem como para verificar sua sobreposição com elementos regulatórios do *Ensembl*, utilizando a legenda de elementos regulatórios do *browser* genômico. No caso de haver correspondência CNE/elemento regulador *Ensembl*, foi realizado o *download* da sequência do elemento regulador identificado. Como passo final desta análise, foi feito um alinhamento entre as sequências do CNE humano e do elemento regulador *Ensembl*, utilizando-se a ferramenta BLAST do NCBI (<https://blast.ncbi.nlm.nih.gov/Blast.cgi>).

5.2.3 *Comparação dos CNEs presentes no locus de Dact1 com elementos reguladores validados do projeto GeneHancer*

A correspondência entre os CNEs identificados no genoma humano e os elementos reguladores de *Dact1* catalogados pelo projeto *GeneHancer* (Fishilevich et al., 2017) também foi analisada. Para este levantamento, foi realizada uma busca textual no banco de dados de genes humanos *GeneCards* (<https://www.genecards.org/cgi-bin/carddisp.pl?gene=DACT1>), uma interface que exibe os potenciais *enhancers* catalogados pelo *GeneHancer* e suas respectivas anotações. Os dados de sequência e localização de cada *enhancer* de *Dact1* catalogado no *GeneCards* foram anotados e comparados sistematicamente com todas as CNEs previamente identificadas no genoma humano.

5.2.4 *Identificação dos contextos celulares e órgãos ou tecidos onde os CNEs no locus de Dact1 estão ativos e reprimidos*

A partir das coordenadas dos elementos reguladores que se relacionam aos CNEs no genoma humano, identificados em 5.2.2, foi realizado um BLAT no *Ensembl* e analisado o *Regulatory Build Track* (Zerbino et al., 2015) para coletar dados sobre os contextos celulares e órgãos ou tecidos onde essas regiões estão ativas ou reprimidas. Esta ferramenta contém um conjunto de regiões do genoma que, provavelmente, estão envolvidas na regulação do gene, podendo ser classificadas em: promotores, regiões flaqueadoras de promotor, *enhancers*, sítios de ligação CTCF, TFBSs e regiões de cromatina aberta. Um algoritmo de segmentação do genoma é então aplicado em todos os ensaios epigenômicos para os tipos de células para os quais um conjunto mínimo de ensaios está disponível. Este conjunto inclui sete ensaios de modificação de histona: H3K4me1, H3K4me3, H3K9ac ou H3K9me3, H3K27ac, H3K27me3 e H3K36me3. A etapa de segmentação primeiro define um número fixo de padrões de sinais recorrentes em todos os conjuntos de dados, chamados de estados, e então atribui um desses

estados a cada par de base (pb) no genoma para cada tipo de célula, determinando sua característica regulatória.

Em seguida, estabelecemos uma correlação com os dados de expressão de *Dact1* em adultos humanos saudáveis, disponíveis na plataforma de dados públicos de expressão e regulação gênica tecido-específicas – *GTEX Portal Release V7 (dbGap Accession phs000424.v7.p2)* (<https://www.gtexportal.org/>) (Consortium, 2013).

5.2.5 *Predição dos sítios de ligação para fatores de transcrição (TFBSs) conservados evolutivamente*

Para identificar TFBSs conservados entre as sequências de elementos reguladores do *Ensembl* que se sobrepõem aos CNEs, inicialmente foi realizado o *download* das sequências humanas que foram empregadas como *bait* para localizar sequências correspondentes no genoma de outros vertebrados. Isto foi feito utilizando-se o *location-based display* do *Ensembl* para delimitar um fragmento de 1 kb, a partir do centro de cada elemento regulador, realizando-se em seguida a exportação da sequência FASTA. Na etapa seguinte, as sequências humanas foram empregadas como *baits* para obter as sequências correspondentes de galinha e camundongo no *browser* genômico *Ensembl* por meio de uma busca utilizando-se o algoritmo BLAT. Após checar a posição de cada sequência no intervalo genômico *KIAA0586-Dact1-Daam1*, foi feito o *download* das sequências FASTA de camundongo e galinha.

Na etapa seguinte, alinhamentos múltiplos de sequência (humano x camundongo x galinha) foram realizados com o programa *Mulan* (<https://mulan.dcode.org/>), que faz interface com a ferramenta *MultiTF*, a qual identifica sítios de ligação para fatores de transcrição evolutivamente conservados (Loots and Ovcharenko, 2007; Ovcharenko et al., 2004). O *MultiTF* utiliza como base para análise a biblioteca do TRANSFAC, que possui em seu banco de dados mais de 400 famílias de fatores de transcrição de vertebrados catalogadas. A similaridade de matriz foi pré-definida como otimizada para função (*optimized for function*), de forma a utilizar diferentes parâmetros para diversos sítios de ligação, minimizando e balanceando a abundância de *hits* falso-negativos de diferentes matrizes (Loots and Ovcharenko, 2007). Os TFBSs

conservados localizados nos CNEs/elementos *Ensembl* foram catalogados em arquivo texto, visando a elaboração das redes de associação funcional. Além disto, para aprofundarmos a análise dos TFBSs, buscamos informações sobre as principais funções dos fatores de transcrição com sítios de ligação preditos para os CNEs na literatura científica.

5.2.6 *Redes de associação de proteínas funcionais (interações proteína-proteína) para os CNEs no locus de Dact1*

O banco de dados de interações proteína-proteína *STRING v.11.0* (<https://string-db.org/>) (Szklarczyk et al., 2019) foi utilizado para determinar as interações conhecidas e previstas entre os fatores de transcrição, identificados em 5.2.5. Essa análise é importante porque as proteínas que interagem umas com as outras provavelmente estão envolvidas em processos ou funções biológicas comuns. Informações sobre interações funcionais e a identificação de outros parceiros de interação foram obtidas a partir de dados de fusão, coexpressão, vizinhança, co-ocorrência, banco de dados experimentais, referências em publicações científicas e homologia de genes. Como critério para a análise de redes, foi adotado o mínimo de cinco TFBSs presentes em cada CNE a ser analisada.

5.2.7 *Deteção de interações entre as regiões que compreendem os CNEs no locus de Dact1*

Com o navegador genômico *3D Genome Browser* (<http://promoter.bx.psu.edu/hic/virtual4c.php>) (Wang et al., 2018) foram analisados os dados de *Hi-C/Virtual 4C* utilizando como *bait* a coordenada inicial do *core* dos CNEs para os quais há correspondência com elementos reguladores do *Ensembl*, identificado em 5.2.2. Para cada CNE foi escolhido um contexto celular onde esta região está ativa, de acordo com o obtido em 5.2.4. Além disso, foram analisados os dados de *Capture Hi-C* tendo como alvo o promotor de *Dact1*. Em adição, para melhor examinar as ligações promotor-regulador, também foram explorados os dados de ligação por locais hipersensíveis à DNase I (DHS), que são marcadores de DNA com papel regulador, nesta região no mesmo contexto celular selecionado. Este método baseia-se no cálculo dos

coeficientes de correlação de Pearson (PCC) entre um gene proximal e os DHSs distais e apenas os pares (gene-DHS) com $PCC > 0,7$ e dentro de 500 kb foram considerados ligados (Thurman et al., 2012).

5.2.8 Geração da construção repórter *pGL2-CNE4-TK-Luc*

As análises *in silico* da região intergênica dos genes *KIAA0586*, *Dact1* e *Daam1* sugeriram a presença de um CNE com alto potencial de atividade em tipos celulares como adipócitos e mioblastos, denominado CNE4. Para testar a capacidade desse fragmento de DNA de modular a atividade transcricional de um gene, o CNE4 foi clonado no vetor pGL2-TK-Luc (Figura 1), que carrega o gene repórter da *Luciferase Firefly* conduzido por um promotor TK mínimo. O fragmento correspondente ao CNE4 em murinos foi amplificado por reação em cadeia da polimerase (PCR) do DNA genômico de *Mus musculus*, extraído com o *kit Wizard Genomic DNA Purification* (Promega), utilizando os seguintes *primers*: CNE4-f: 5' - TCC ACC CTA CAC CCT CCT TT - 3' e CNE4-r: 5' - TTG CCA AAA CAG AAC ACA AGC - 3' produzindo um fragmento de PCR de 665 pb. Resumidamente, as condições para PCR compreenderam: uma mistura de reação contendo: 300 ng de DNA molde, 1 μ M de cada iniciador, 0,5 unidades de DNA polimerase (Fermentas), tampão Taq 10X (Fermentas), dNTP 0,2 mM, $MgCl_2$ 2 mM e dH_2O para um volume final de 50 μ L por reação. A PCR foi realizada por 30 ciclos, consistindo em uma etapa de desnaturação a 95°C (30 seg), uma etapa de anelamento a 55°C (30 seg) e uma etapa de extensão de 72 °C (1 min) em termociclador *Mastercycler Pro* (Eppendorf). O fragmento gerado por PCR foi clonado no vetor pGL2-TK-Luc por digestão enzimática com *SacI* e *XhoI* (Invitrogen). O fragmento foi verificado por sequenciamento.

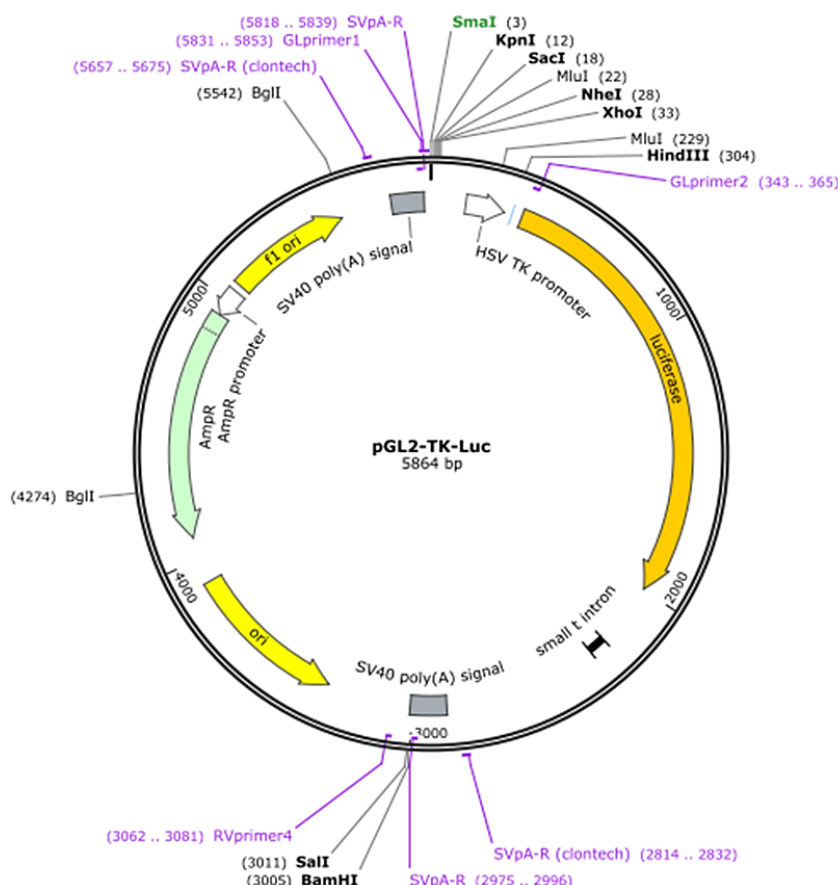


Figura 1. Vetor pGL2-TK-Luc utilizado para clonagem. A seta em branco destaca o promotor mínimo TK, upstream ao gene repórter da Luciferase Firefly (laranja). Os sítios utilizados para abertura do vetor e clonagem do fragmento foram SacI e XhoI. Nesta mesma região foi inserida a sequência correspondente ao CNE4 murino.

5.2.9 Cultura de células HEK293T, C2C12 e pré-adipócitos de camundongo

As linhagens HEK293T, C2C12 e pré-adipócitos imortalizados de tecido adiposo branco de camundongo (9W), gentilmente doados pelo Dr. Marcelo A. Mori, foram cultivados com *Dulbecco's Modified Eagle High Glucose Medium* (DMEM) suplementado com 10% de soro fetal bovino (FBS) contendo os antibióticos penicilina (100 U/mL) e estreptomicina (100 µg/mL). As células foram mantidas em placas de 24 poços a 37°C em atmosfera umidificada contendo 5% de CO₂.

5.2.10 Ensaios de transfecção celular

A construção repórter pGL2-CNE4-TK-Luc (CNE4) foi transfectada em células HEK293T, C2C12 e em pré-adipócitos murinos imortalizados para medir a diferença na

atividade da *Luciferase* em relação à atividade da *Luciferase* do vetor de controle pGL2-TK-Luc. As células foram transfectadas usando *Lipofectamine 2000* (ThermoFisher Scientific). Um total de 1×10^5 células foi colocado em solução de transfecção com 0,5 µg do construto de expressão pGL2-CNE4-TK-Luc. Em todos os ensaios, 10 ng do vetor de normalização pRL-SV40 (Promega) foi co-transfectado para normalizar a eficiência de transfecção (PROMEGA CELL NOTES ISSUE 17 (2007)³³. As células foram coletadas 24 horas após a transfecção para quantificação da atividade do gene repórter da *Luciferase* pelo *Dual-Glo Luciferase Assay System* (Promega), de acordo com as instruções do fabricante, e as leituras foram realizadas no luminômetro *GloMax-Multi Detection System* (Promega). Os valores médios da atividade da *Luciferase* foram expressos como uma razão para a expressão do gene repórter dividido pelos valores de pGL2-TK-Luc não modificado (*Fold expression*).

5.2.11 Análise estatística

A significância estatística foi analisada pela comparação de pares com o teste t de Student bicaudal ($p \leq 0,05$). Os dados foram apresentados como média $\pm$ erro padrão da média (SEM) com base em 3 experimentos independentes (poços em triplicata).

5.2.12 Alinhamentos múltiplos de sequência do CNE4

As sequências ortólogas ao CNE4 de diferentes animais, incluindo organismos-modelo e várias espécies de interesse econômico, foram coletadas no *browser* genômico *Ensembl* utilizando-se a ferramenta BLAT (<https://www.ensembl.org/>). Nestas buscas, a sequência humana foi empregada como *bait* para identificação das sequências de mamíferos, enquanto a sequência de galinha foi utilizada para rastrear a sequência do peru. Para cada espécie analisada, foi baixado um segmento de DNA de 1 kb no formato FASTA, definido a partir do centro da

33

<https://www.promega.com/~media/Files/Resources/Cell%20Notes/cn017/Normalizing%20Genetic%20Reporter%20Assays.ashx>

sequência identificada pelo BLAT. O alinhamento das sequências ortólogas ao CNE4 das espécies comparadas foi realizado no *Mulan* (<https://mulan.dcode.org/>).

A Figura 2 a seguir resume as estratégias de pesquisa utilizadas no estudo apresentado neste capítulo.

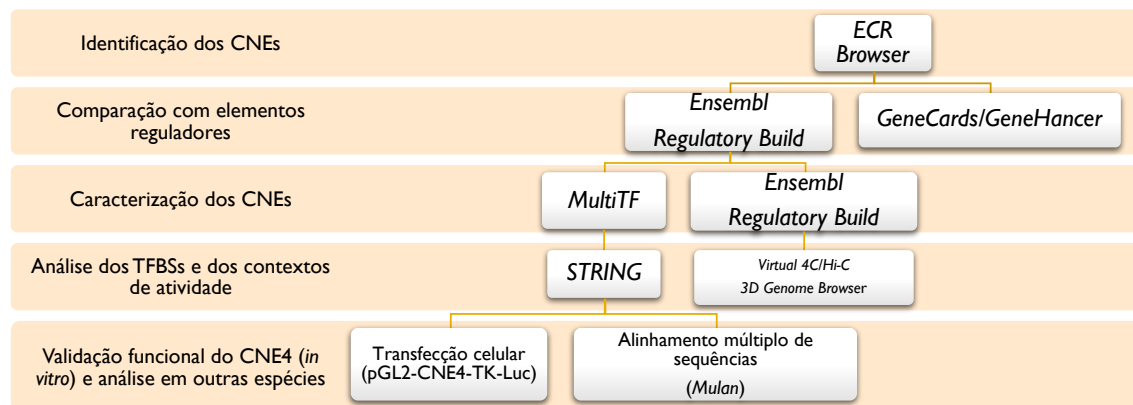


Figura 2. Fluxo de trabalho da pesquisa realizada no capítulo 2 desta tese.

5.3 RESULTADOS

5.3.1 Oito CNEs foram identificados nas regiões intergênicas entre os genes *KIAA0586*, *Dact1* e *Daam1*

A pesquisa por CNEs no intervalo genômico *KIAA0586-Dact1-Daam1* incluiu os genomas do humano (*Homo sapiens*), de camundongo (*Mus musculus*) e da galinha (*Gallus gallus*). O genoma do peixe-zebra também foi inicialmente incluído nas análises, mas nenhum CNE foi identificado nesta espécie. Por meio desta análise, oito CNEs (extensão > 200 pb e mínimo de 70% de identidade de sequência) foram identificados (Figura 3; Apêndice 1). Como estes CNEs estão presentes no genoma de mamíferos e da galinha, é provável que estas sequências tenham surgido a pelo menos 310 milhões de anos, considerando o tempo de divergência entre o último ancestral comum entre aves e mamíferos (Kumar and Hedges, 1998).

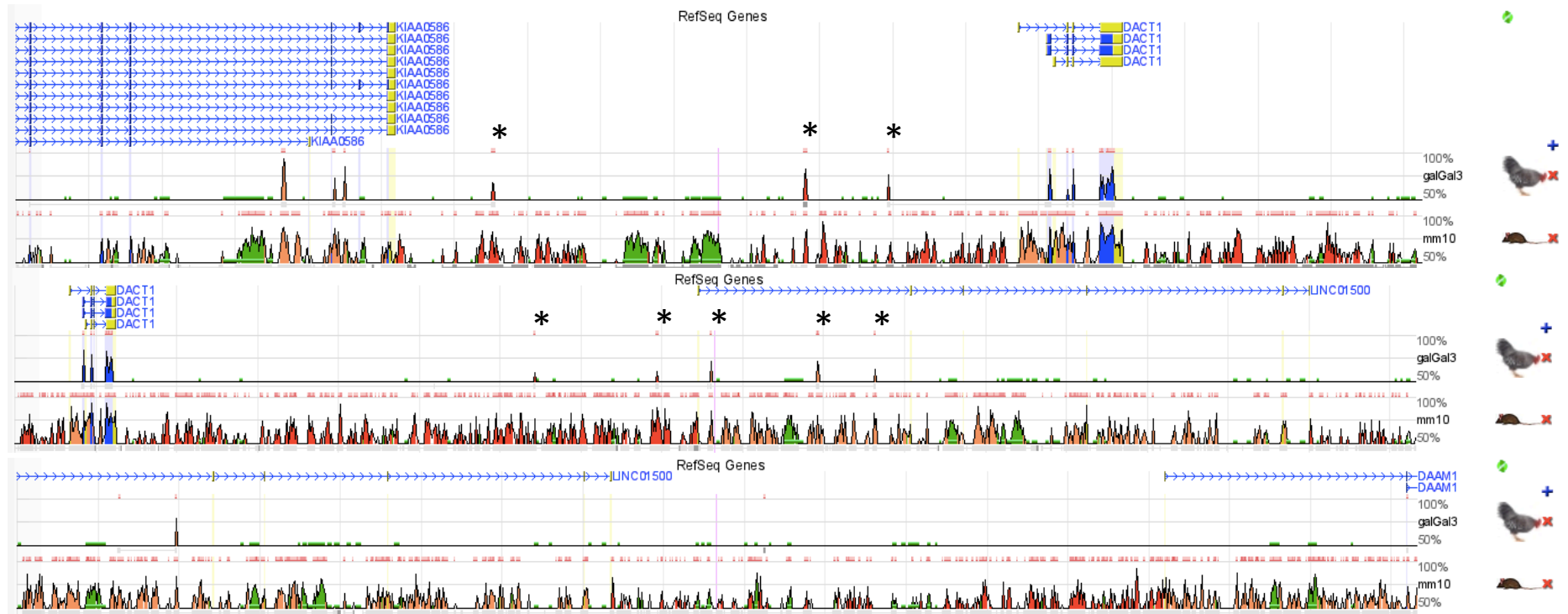


Figura 3. Elementos não codificadores conservados (CNEs) em torno do locus de *Dact1* humano (CRCh37: chr14:59,100,685-59,115,039). A comparação do genoma humano com as sequências de duas outras espécies permitiu a identificação de três CNEs entre *KIAA0586* e *Dact1* e cinco outros entre *Dact1* e *Daam1*. Em azul estão representados os exons, em salmão os íntrons, em amarelo as UTRs 5' e 3'. O verde indica sequências repetitivas e o vermelho indica as regiões intergênicas. A linha superior refere-se ao gene humano, utilizado como referência. Há indicação de orientação do gene, pelas cabeças de seta na linha superior azul mais fina. Os asteriscos representam picos de conservação evolutiva.

5.3.2 Os CNEs se localizam em regiões de elementos reguladores previamente identificados

A fim de verificar a correspondência entre a sequência e a localização dos CNEs identificados por genômica comparativa com elementos reguladores da transcrição estabelecidos por dados experimentais, cada coordenada dos CNEs foi analisada no genoma humano por meio do *Regulatory Build Track* do *Ensembl* (Zerbino et al., 2015). Este *track* regulatório utiliza dados experimentais para prever características que indicam potencial para regulação da expressão gênica. A anotação dessas características regulatórias é baseada em uma ampla variedade de dados produzidos a partir dos projetos de epigenômica *Blueprint* e *Roadmap* e do projeto ENCODE. Este último, abriga uma ampla quantidade de dados para diversos tipos de células em cultura, incluindo ensaios de hipersensibilidade de DNA, ensaios de metilação de DNA e *ChIP-Seq* de proteínas que interagem com o DNA, ou seja, histonas modificadas, fatores de transcrição e reguladores da cromatina (Consortium, 2012). A comparação de dados mostrou que o CNE1 é adjacente a uma região caracterizada como flanqueadora de promotor e que os CNEs 5, 7 e 8 se sobrepõem às regiões flanqueadoras de promotor. Os CNEs 4 e 6 se sobrepõem às regiões caracterizadas como *enhancers* no genoma humano (Tabela 1).

Em adição, os CNEs identificados previamente foram comparados com os elementos reguladores de *Dact1* catalogados no banco de elementos reguladores *GeneHancer* (Fishilevich et al., 2017), usando sua posição e sequência no genoma. O catálogo *GeneHancer* lista milhares de *enhancers* humanos, que foram identificados empregando um algoritmo de integração para remover a redundância de dados e integrar informações. Neste catálogo, cada *enhancer* recebe uma pontuação referente ao nível de confiança da análise, derivada da anotação. Em adição, o *GeneHancer* associa *enhancers* a genes específicos usando vários dados, tais como: correlação de coexpressão de genes em tecidos específicos, expressão de RNAs em *enhancers*, bem como dados sobre fatores de transcrição e análises de *Hi-C*, a partir de dados integrados da Enciclopédia de Elementos de DNA (ENCODE), do *track* regulatório do *Ensembl*, da anotação funcional do projeto genoma de mamíferos (FANTOM) e do *browser VISTA Enhancer*. Contudo, não foi observada sobreposição entre os CNEs aqui identificados e os elementos reguladores do *GeneHancer* (Tabela 2). Esse resultado pode ser devido ao fato de que os elementos reguladores

catalogados pelo *GeneHancer* são restritos ao genoma humano e, possivelmente, as análises tenham sido realizadas em poucos contextos celulares, insuficientes para determinar a relação com *Dact1* mais amplamente.

Tabela 1. Localizações genômicas dos CNEs de acordo com o banco de dados do Ensembl e elementos reguladores humanos correspondentes que estão sobrepostos a cada CNE.

	CNE1	CNE2	CNE3	CNE4	CNE5	CNE6	CNE7	CNE8
Localização (GRCh37/hg19)	chr14:59,028,673 - 59,029,142	chr14:59,071,439 - 59,071,876	chr14:59,082,951 -59,083,157	chr14:59,244,279 -59,244,528	chr14:59,282,228 - 59,282,497	chr14:59,298,964 -59,299,212	chr14:59,331,774 - 59,331,996	chr14:59,349,427 -59,349,727
Localização (GRCh38/hg38)	chr14:58,561,931 -58,562,447	chr14:58,604,699 -58,605,179	chr14:58,616,222 -58,616,449	chr14:58,777,548 -58,777,822	chr14:58,815,496 -58,815,792	chr14:58,832,233 -58,832,506	chr14:58,865,044 -58,865,289	chr14:58,882,694 -58,883,024
Distância relativa ao <i>Dact1</i>	72 kb	29,2 kb	17,7 kb	143,5 kb	181,5 kb	198,2 kb	231 kb	248,7 kb
ID do Elemento Regulador	ENSR00000498555	-	-	ENSR00000985362	ENSR00000498642	ENSR00000985376	ENSR00000498653	ENSR00000985382
Tipo	Região flanqueadora de promotor chr14:	-	-	Região <i>Enhancer</i> chr14:	Região flanqueadora de promotor chr14:	Região <i>Enhancer</i> chr14:	Região flanqueadora de promotor chr14:	Região flanqueadora de promotor chr14:
<i>Core</i>	58,560,174- 58,561,199	-	-	58,777,401- 58,778,001	58,815,001- 58,817,599	58,832,000- 58,832,800	58,864,575- 58,866,988	58,882,401- 58,883,200

Tabela 2. Elementos reguladores do gene *Dact1*, segundo o banco de sequências reguladoras humanas GeneHancer. É importante notar que há evidências de correção de *Dact1* com outros genes vizinhos, como *KIAA0586*, mas não com *Daam1*.

GeneHancer	Tipo	Localização (GRCh38/hg38)	Ensembl ID (localização)	Genes-alvo
GH14J058636	Promotor/ <i>Enhancer</i>	chr14:58,636,083-58640640	ENSR00000069061 (chr14:58,636,600-58,640,601)	DACT1; piR-42116; lnc-TIMM9-4
GH14J058633	Promotor/ <i>Enhancer</i>	chr14:58,633,388-58,635,991	ENSR00000498577 (chr14:58,635,000-58,635,401)	DACT1; piR-42116; lnc-TIMM9-4
GH14J058238	Promotor/ <i>Enhancer</i>	chr14:58,238,020-58,240,716	ENCODE chr14:58,238,020-58,240,716	TRK-CTT1-1; ARID4A; DACT1; KIAA0586; ENSG00000258682; PSMA3-AS1; ARMH4; NONHSAG015102.2-002
GH14J058532	<i>Enhancer</i>	chr14:58,532,883-58,535,639	ENSR00000498549 (chr14:58533002-58535600)	DACT1; PSMA3-AS1; HNRNPCP1; piR-33605-155; piR-43105-167; KIAA0586
GH14J058730	<i>Enhancer</i>	chr14:58730002-58733999	ENSR00000069067 (chr14:58730002-58733999)	DACT1; RF01045-040; piR-53063-064; LINC01500

5.3.3 *As regiões regulatórias identificadas são ativas em tipos celulares, tecidos e órgãos onde o gene *Dact1* é expresso durante o desenvolvimento e/ou no pós-natal*

A partir das coordenadas dos elementos reguladores que se relacionam aos CNEs no genoma humano, identificados em 5.3.2, foi realizado um *BLAT* no *Ensembl* e analisado o *Regulatory Build Track* para coletar dados sobre os contextos celulares e órgãos ou tecidos onde essas regiões estão ativas ou reprimidas. Em seguida, estabelecemos uma correlação com os dados de expressão de *Dact1* em adultos humanos saudáveis, disponíveis na plataforma de dados públicos de expressão e regulação gênica tecido-específicas – *GTEx Portal Release V7* (Consortium, 2013).

O elemento regulador que está na região do CNE1 se destaca por sua atividade em tipos celulares relacionados ao sistema imunológico, como em células CD4 + ab T, CD8 + ab T, células CMP CD4 +, mielócitos e neutrófilos, mas também em células e órgãos do sistema digestório, como intestino, estômago, hepatócitos e em HepG2, uma linhagem humana de células provenientes de câncer de fígado, bem como na linhagem A673, proveniente de rhabdomyosarcoma. É importante ressaltar que, em amostras desses órgãos advindas de humanos adultos saudáveis, *Dact1* é pouco expresso (Figura 4). Por outro lado, marcadores de atividade foram encontrados em amostras de aorta, um tecido onde *Dact1* é altamente expresso em humanos. Em relação aos marcadores de repressão encontrados ao longo dessa região flangeadora de promotor, foram reportados marcadores de repressão em células do sistema imunológico (células B e CD4+ CD25+ ab Treg), células mamárias epiteliais e em linhagens celulares de câncer de mama e de leucemia de células T.

O elemento regulador que se sobrepõe CNE4 tem sua atividade descrita em diferentes linhagens de células-tronco (iPS-15b, HUES6, HUES64, H1-hESC), glândula adrenal, neurônios e células progenitoras neurais (NPC), pulmões, intestinos grosso e delgado, placenta e células do músculo esquelético de diferentes compartimentos, como músculos do tronco e das pernas (Tabela 3). É importante notar que alguns desses órgãos apresentam expressão marcante de *Dact1* em adultos, incluindo a glândula adrenal, pulmões e em diversas estruturas do cérebro (Figura 4). Quanto aos marcadores de repressão de atividade desta região, estes foram

reportados nos tipos celulares do sistema imunológico (células B e CD4+ CD25+ ab Treg), em fibroblastos e melanócitos do prepúcio, em tecidos provenientes do ventrículo esquerdo e dos pulmões, e também aparece reprimida nas células HCT116 e nas células PC-9, que são linhagens celulares de câncer de cólon e de adenocarcinoma de pulmão, respectivamente (Tabela 4).

Em relação ao elemento que se sobrepõe ao CNE5, a atividade desta região é encontrada em vários tipos de células dérmicas, osteoblastos, ovários, mas também em células musculares, como miotubos, células mioblásticas do músculo esquelético (HSMM), células musculares esqueléticas do tronco e na linhagem de rhabdomyosarcoma (A673) (Tabela 3). Já em células B e CD4+ CD25+ ab Treg há marcadores de repressão identificados na região que compreende o CNE5, em adição aos tecidos provenientes do ventrículo esquerdo, pulmões, átrio direito, cólon sigmoide e do trofoblasto, um tecido embrionário que forma a parede externa do embrião em fase de blástula. Nas células progenitoras neurais (NPC) esta região também aparece reprimida. Além disso, esta região genômica tem sua atividade reprimida em importantes contextos que remetem a vários tipos de câncer, como dos pulmões (PC-9 e A549) e de leucemia mieloide crônica (K562) (Tabela 4).

Na região que compreende o CNE6, destacam-se os marcadores de atividade encontrados em órgãos do aparelho digestório, como intestino e estômago, e hepatócitos, mas também é importante notar que nesta região também são encontrados marcadores de atividade em células musculares esqueléticas, como nos miotubos e músculos da perna e no músculo Psoas (Tabela 3). Amostras de cérebro e em progenitores neurais (neuroesferas) também foram consideradas positivas na expressão de atividade neste segmento genômico (Tabela 3). A região que sobrepõe ao CNE6 encontra-se reprimida nas células B, CD4+ CD25+ ab Treg e na linhagem de células linfoblastóides (GM12878), em tecidos provenientes da aorta, do cólon sigmoide, no tecido endodérmico, queratinócitos e nas células endoteliais da veia umbilical (HUVEC) e também na linhagem de células de adenocarcinoma de pulmão (PC-9) (Tabela 4).

Na região do CNE7, marcadores de atividade foram detectados em HSMM, miotubos, músculos da perna e do tronco, rhabdomyosarcoma (A673), vários órgãos digestórios e urinários, como fígado, intestino, estômago, rins, mas também em células dérmicas e tecidos do coração, nos astrócitos e macrófagos (M1 e M2) (Tabela 3). O CNE7 aparece reprimido em células do

sistema imunológico: células B, CD4+ CD25+ ab Treg, CD8+ ab T e na linhagem de células linfoblastóides (GM12878), bem como nas células endoteliais da veia umbilical (HUVEC) (Tabela 4).

Por último, no segmento que se sobrepõe ao CNE8, a atividade da região genômica foi relatada em células musculares, rabdomiossarcoma (A673), células-tronco mesenquimais (MSC), neurônios, osteoblastos e ovário – um órgão com expressão marcadamente importante de *Dact1* (Tabela 3; Figura 4). Quanto aos marcadores de repressão de atividade, foram reportados em células B, CD4+ CD25+ ab Treg, CD4+ ab T, na linhagem de células linfoblastóides (GM12878), nas células endoteliais da veia umbilical (HUVEC) e, por fim, nas células progenitoras neurais (NPC) (Tabela 4).

Todas as informações sobre os tipos celulares e tipo de atividade predito para os elementos reguladores em cada contexto encontram-se sumarizados nas Tabelas 3 e 4.

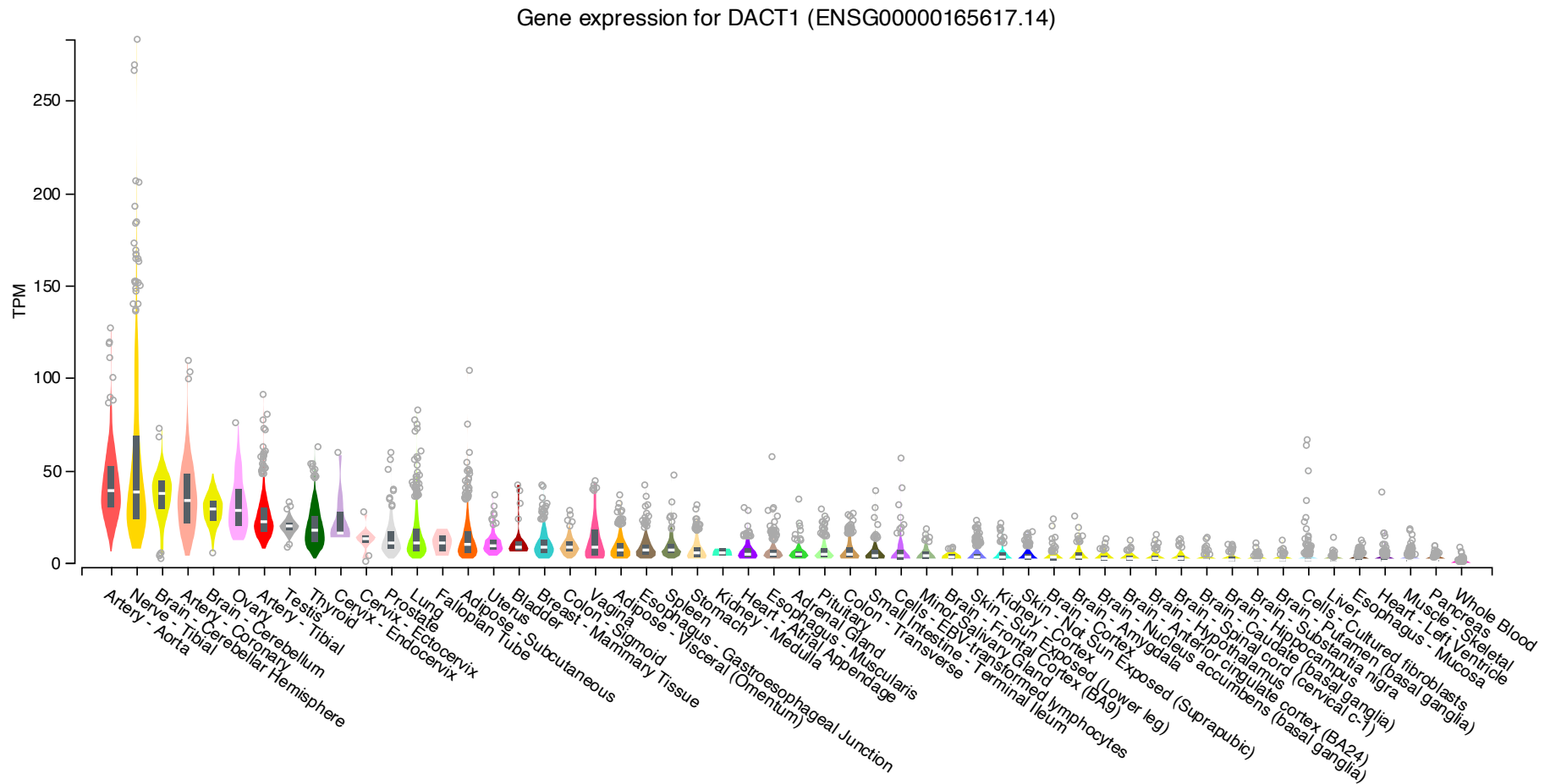


Figura 4. Gráfico de expressão do gene *Dact1* em amostras provenientes de órgãos e tecidos de adultos humanos saudáveis (GTEx Portal), na escala transcritos por milhão (do inglês, Transcripts Per Million, TPM).

Tabela 3. Linhagens celulares, tecidos e órgãos com marcadores de atividade identificados no banco de dados Ensembl Regulatory Build, considerando os CNEs relacionados aos elementos reguladores humanos identificados em 5.3.2.

Tipo celular/Tecido/Órgão	CNE1	CNE2	CNE3	CNE4	CNE5	CNE6	CNE7	CNE8
A673	+				+		+	+
Glândula adrenal				+				
Aorta	+							
Astrócito							+	
Neurônio bipolar				+				+
Cérebro_1						+		
CD4+ ab T (VB)	+							
CD8+ ab T (CB)	+							
CMP CD4+_1	+			+				
Derme (fibroblasto)					+		+	
Endodérmico					+			
Prepúcio (fibroblasto)_2					+		+	
Prepúcio (queratinócito)_2					+		+	
Prepúcio (melanócito)_2					+			
H1-hESC_2				+				
H1-hESC_3				+				
He-La-S3							+	
Hepatócito	+					+	+	
HepG2	+							
HSMM					+		+	+
HUES6				+				
HUES64				+				
HUVEC-prol (CB)	+							
iPS-15b				+				
Rim							+	
Intestino grosso	+			+			+	
Músculo da perna				+	+	+	+	
Pulmão_1				+		+	+	
M1 (VB)							+	
M2 (CB)							+	

Tipo celular/Tecido/Órgão	CNE1	CNE2	CNE3	CNE4	CNE5	CNE6	CNE7	CNE8
M2 (VB)							+	
Mioepitélio mamário						+		
MCF-7						+		
MSC					+		+	+
MSC (VB)							+	
Miotubo					+	+	+	+
Neurônio				+				
Neuroesfera (C)						+		
Neutro mielócito	+							
Neutrófilo (CB)	+							
Neutrófilo (VB)	+							
NHLF					+		+	
NPC_1				+				
NPC_3				+				
Osteoblasto					+		+	+
Ovário							+	+
Pâncreas							+	
PC-3						+		
Placenta				+		+	+	
Músculo Psoas						+		
Átrio direito							+	
Cólon sigmoide							+	
Intestino delgado	+							
Intestino delgado_1				+		+	+	
Intestino delgado_2						+		
Estômago	+							
Estômago_1					+	+	+	
Trofoblasto							+	
Músculo do tronco				+	+		+	

A673: *primary rhabdomyosarcoma*; CMP CD4+: *common myeloid progenitor CD4+*; H1-hESC: *H1 human embryonic stem cell line*; He-La-S3: *human cervix adenocarcinoma*; HepG2: *human liver cancer cell line*; HSMM: *skeletal muscle myoblasts cells*; HUES6: *human embryonic stem cell line with a normal 46XX karyotype*; HUES64: *human embryonic stem cell line with a normal 46XY karyotype*; HUVEC: *human umbilical vein endothelial cells*; iPS-15b: *human induced stem cell line derived from dermal fibroblasts*; MCF-7: *human breast cancer cell line*; MSC: *mesenchymal stem cell*; M1: *macrophages*; M2: *macrophages*; NHLF: *normal human lung fibroblasts*; NPC: *neural progenitor cells*; PC-3: *caucasian prostate adenocarcinoma*.

Tabela 4. Linhagens celulares, tecidos e órgãos com marcadores de repressão identificados no banco de dados Ensembl Regulatory Build, considerando os CNEs relacionados aos elementos reguladores humanos identificados em 5.3.2.

Tipo celular/Tecido/Órgão	CNE1	CNE2	CNE3	CNE4	CNE5	CNE6	CNE7	CNE8
A549					+			
Aorta						+		
B	+			+	+	+	+	+
CD4+ ab T								+
CD4+ CD25+ ab Treg (PB)	+			+	+	+	+	+
CD8+ ab T (PB)							+	
DND-41	+							
Endodérmico						+		
Prepúcio (fibroblasto)_2				+				
Prepúcio (melanócito)_1				+				
GM12878						+	+	+
HCT116				+				
HUVEC						+	+	+
K562					+			
Queratinócito						+		
Ventrículo esquerdo				+	+			
Pulmão_2				+	+			
Pulmão_2 NHLF								
Epitélio mamário_1	+							
MCF-7	+							
NPC_3					+			+
PC-9				+	+	+		
Átrio direito					+			
Cólon sigmoide					+	+		
Trofoblasto					+			

DND-41: Human T cell leukemia; MCF-7: human breast cancer cell line; HCT116: human colon cancer cell line; PC-9: human adenocarcinoma from lung tissue; A549: adenocarcinomic human alveolar basal epithelial cells; K562: human myelogenous leukemia cell line; NPC: neural progenitor cells; HUVEC: human umbilical vein endothelial cells; GM12878: human lymphoblastoid cell line; NHLF: normal human lung fibroblasts.

5.3.4 Vários sítios de ligação para fatores de transcrição (TFBSs) foram identificados nas regiões correspondentes aos CNEs, reforçando o potencial papel regulador destes elementos

Para identificar TFBSs evolutivamente conservados entre as sequências de CNEs, utilizamos como *bait* um segmento de 1 kb a partir do centro de cada elemento regulador do *Ensembl*, identificado em humanos, para realizar um BLAT no banco de dados genômicos do *Ensembl* para, assim, obter as sequências correspondentes em camundongo e galinha (Apêndice 2). Isto foi feito visando ampliar o tamanho da região passível de comparação entre as diferentes espécies, permitindo identificar eventuais sítios conservados localizados além dos limites inicialmente estabelecidos para os CNEs, segundo a análise feita no *ECR browser*. Em seguida, as três sequências de espécies foram alinhadas utilizando a plataforma *Mulan* e os TFBSs conservados foram identificados com o *software MultiTF*, que armazena informações do banco de dados *TRANSFAC* (Figura 5). Esta análise é importante pois o enriquecimento de TFBSs em CNEs são indicadores significativos da presença de sinais regulatórios, aumentando a chance desses *sites* serem funcionais (Levy et al., 2001).

Nossos resultados mostraram que todos os CNEs contêm TFBSs conservados, exceto o CNE6. Os TFBSs localizados nos CNEs foram listados na Tabela 5. Da mesma forma, também foram analisadas as outras duas sequências de CNEs identificados em 5.3.1 – os CNEs 2 e 3. Com isto, foi possível identificar que os CNEs 2 e 3 também possuem TFBSs conservados (Tabela 5). Isto sugere que, embora não haja dados experimentais validados, estas duas regiões possam ser funcionais, mas dada a limitação dos tipos celulares utilizados nos experimentos depositados nos bancos de dados, estes elementos não tenham sido descritos até o momento.

Para detalhar a análise dos TFBSs, buscamos informações sobre as principais funções dos fatores de transcrição (TFs) com sítios de ligação previstos nos CNEs, utilizando o banco de dados de interações proteína-proteína *STRING* e dados da literatura (Apêndice 3). Este levantamento de informações é importante para levantar hipóteses sobre o possível papel regulador dos CNEs identificados.

O CNE1 apresenta TFBSs principalmente relacionados à regulação geral da transcrição como MAX, um componente da rede MYC-MAX-MXD1 de fatores de transcrição *basic helix-*

loop-helix leucine zipper (bHLHZ) que regulam a proliferação, diferenciação e apoptose celular (Nair and Burley, 2003), o ativador transcricional FOXJ2 (Gómez-Ferrería and Rey-Campos, 2003), o elemento promotor transcricional E-box (Church et al., 1985) e USF, um fator de transcrição celular expresso de forma ubíqua (Gregor et al., 1990). Já PBX1 regula vários processos embrionários, incluindo padronização, organogênese e, especialmente, a hematopoese (Aspland et al., 2001; Zhou et al., 2020). É importante ressaltar que este CNE também apresenta sítios de ligação para PAX2 e HFH3 (FOXI1), fatores de transcrição importantes para a formação do rim e que também são expressos nos olhos e ouvidos internos do embrião humano durante o desenvolvimento (Dressler, 2009; Enerbäck et al., 2018; Torres et al., 1996). O fator de transcrição HFH4 (FOXJ1) é expresso no desenvolvimento do epitélio pulmonar e renal (Pelletier et al., 1998) e também é necessário para a ciliogênese (Stauber et al., 2017), enquanto o HFH8 (FOXF1) é expresso nas células endoteliais e musculares lisas no pulmão do embrião e do adulto (Kalinichenko et al., 2001).

O CNE2 também abriga TFBSs gerais e outros relacionados à organogênese, como HLF, expresso no fígado fetal e adulto e em uma linhagem celular de carcinoma hepático (Hunger et al., 1992), LEF1 (diferenciação de células capilares e morfogênese folicular) (Zhou et al., 1995) e PAX5, que é importante para a diferenciação de linfócitos B, bem como para a espermatogênese (Adams et al., 1992). PAX5 também está relacionado ao desenvolvimento neural, bem como, o VBP (Hemberger et al., 1999).

O CNE3 apresenta um menor número de TFBSs conservados entre humano, camundongo e galinha, no entanto, ele possui sítios de ligação para HOXD13, um regulador chave da morfogênese esquelética do autópode (que no ser humano corresponde à mão ou ao pé) e TFE, uma família de fatores de transcrição implicados nas translocações associadas ao câncer. Por exemplo, as translocações de TFE3 ocorrem em carcinomas renais distintos da infância e de adultos jovens e também no sarcoma alveolar (Kuss et al., 2014; Ladanyi et al., 2001; Sidhar et al., 1996).

O CNE4 destaca-se pela presença de múltiplos e diversos TFBSs. Vários TFBSs estão relacionados à organogênese, como FOXI1, FOXJ1 e HFH8, todos compartilhados com o CNE1, mas também possui sítios para outros membros da família *Forkhead*, como FOXQ1, que

desempenha um papel na diferenciação do folículo, e a família de fatores de transcrição SMAD, que é envolvida em um grande número de processos celulares por meio da sinalização de TGF- β e tem sido associada ao controle do metabolismo, fisiologia muscular e adipogênese (Goldstein et al., 2014; Tan et al., 2012). Ainda neste último contexto, este CNE também se destaca pelos TFBSs relacionados à adipogênese como a família PPAR, em especial PPARG, e também os membros da família *Forkhead* FOXO1 e FOXO4, alvos da sinalização da insulina e reguladores da homeostase metabólica em resposta ao estresse oxidativo (Webb and Brunet, 2014). É importante ressaltar que o CNE4 também apresenta TFBSs para importantes ativadores transcricionais da miogênese, como MYOD e Miogenina (MYOG), reguladores-chave deste processo de diferenciação (Sabourin and Rudnicki, 2000). Os TFBS para SRY compreendem uma matriz que também poderia interagir com membros da SOX, dado o fato de pertencerem à mesma família de TFs e a similaridade das sequências. Isso sugere que os fatores SOX, que desempenham papéis na determinação do sexo, desenvolvimento dos membros, condrogênese e também na neurogênese (Bell et al., 1997; Collignon et al., 1996; Cox et al., 2011; Raspopovic et al., 2014), poderiam interagir com o CNE4 para modular a expressão de genes do desenvolvimento. Isto tudo levantou a hipótese de que o CNE4 poderia apresentar alto potencial como regulador da expressão gênica em contextos específicos como a miogênese e a adipogênese.

O CNE5 também apresenta TFBSs relacionados ao metabolismo, como o PAX4, que desempenha um papel importante na diferenciação e desenvolvimento de células β pancreáticas (Mansouri et al., 1999). Neste CNE, também é importante notar que o fator de transcrição ALX4 é importante para muitos processos de desenvolvimento, mas é especialmente necessário para o correto posicionamento da zona de atividade de polarização (ZPA) e subsequente desenvolvimento do membro (Takahashi et al., 1998), bem como no desenvolvimento craniofacial, desenvolvimento da pele e do folículo capilar (Kayserili et al., 2009). TEF é um membro da família de fatores de transcrição PAR bZIP, que inclui TEF, DBP e HLF – este último presente no CNE2 –, e eles apresentam alta expressão no núcleo supraquiasmático, o marca-passo circadiano em mamíferos. Em contrapartida, eles são expressos em níveis quase invariáveis na maioria das regiões do cérebro, nas quais a expressão do gene *Clock* só aparece em

níveis mais baixos. Além disso, esta família de fatores de transcrição, incluindo TEF, tem alta expressão, conforme o ciclo circadiano, no fígado (Gachon et al., 2004). É importante notar que no CNE1 também encontramos TFBS para o fator de transcrição CLOCK que é essencial para o ritmo circadiano (King et al., 1997). Neste contexto, CRX, também presente no CNE1, atua com outros fatores de transcrição para regular a transcrição do gene específico da célula fotorreceptora na retina em desenvolvimento (Furukawa et al., 1997). No CNE5 também encontramos um TFBS para proteínas CLOX, que são expressas predominantemente em embriões em estágios iniciais e são reguladas positivamente em mioblastos e condrócitos comprometidos, mas reguladas negativamente na diferenciação terminal (Andres et al., 1992). Em contraste, espécies maiores (em peso molecular) de proteínas CLOX são detectadas principalmente em tecidos adultos e em células terminalmente diferenciadas. CLOX, como seus homólogos na *Drosophila*, são reguladores candidatos da especificação do destino celular em vários programas de diferenciação (Andres et al., 1992). E47 (TCF3) é um fator de transcrição importante para a regulação da expressão do gene da imunoglobulina e para a diferenciação das células B (Bain et al., 1994), e também é essencial para a diferenciação muscular. O reconhecimento de E-box em promotores musculares requer heterodimerização de MYOD com E47, culminando na formação deste dímero funcional, que é essencial para a miogênese esquelética (Lassar et al., 1991).

Na análise da região do CNE7 foi encontrada a matriz de ligação para o fator de transcrição IK1; este fator é um regulador da diferenciação das células hematopoiéticas e desempenha papel essencial no desenvolvimento de linfócitos, células B e células T (Georgopoulos et al., 1994). Por fim, para o CNE8 vale destacar a presença de matriz de ligação para os fatores CEBPs. Os fatores de transcrição CEBP α e β (CEBPA e CEBPB) são amplamente relacionados à coordenação da proliferação e diferenciação de mielóides, adipócitos, hepatócitos e células do pulmão e placenta, e também à regeneração hepática e hematopoese (Greenbaum et al., 1998; Lin et al., 1993; Swart et al., 1997; Zhang et al., 2004b). CEBPB também reprime o crescimento e proliferação de cardiomiócitos no coração de mamíferos adultos, desempenhando um papel central na hipertrofia fisiológica e proliferação neste órgão (Boström et al., 2010). Neste contexto, os adipócitos deficientes em CEBPA acumulam menos lipídios e não induzem PPARG

[illegible]

Figura 5. Diagrama dos picos de conservação e dos TFBSs, gerados pelo Mulan, presentes nas regiões correspondentes aos CNEs que estão contidos nos elementos reguladores identificados pelo Ensembl Regulatory Track.

Tabela 5. TFBSs conservados entre humano, camundongo e galinha, identificados pelo MultiTF.

Região correspondente	TFBSs conservados
CNE1	FREAC4; ARNT; MYCMAX; MAX; STRA13; USF; E47; E-BOX_Q6; NMYC; USF_Q6; SREBP1; CLOCKBMAL_Q6; USF_C; USF_Q6; PBX1; POU6F1; CDXA; FOXJ2; HFH3; HFH4; HFH8; XFD1; FOXO1; PAX2; CRX_Q4
CNE2	CBF; CEBP; CEBPGAMMA; E4BP4; FREAC2; HLF; ISRE; LEF1; NFAT; PAX5; SOX5; VBP; VMYB; ZEC
CNE3	CDPCR3; HOX13; NRF2; TFE
CNE4	CDXA; AP4; MYOD_Q6; AP4_Q6; MYOGENIN_Q6; LBP1_Q6; XVENT1; COMP1; SMAD_Q6; PPARG; HNF4_01_B; PPAR_DR1_Q2; FOXO4; FREAC2; XFD2; FOXO3; FOXO4; FOXD3; FOX_Q2; HFH1; HFH3; HFH4; HFH8; FOXO1; SRY
CNE5	ZTA_Q2; VMYB; AP4; E47; AP4_Q6; LBP1_Q6; CDP; CLOX; TEF_Q6; GR_Q6; ALX4; PAX4; KAISO
CNE6	Não foram encontrados TFBSs conservados entre as espécies
CNE7	IK1
CNE8	CEBP_Q3; P53; P53_DECAMER_Q2; T3R_Q6

5.3.5 Os fatores de transcrição para os quais os CNEs 1, 2 e 4 possuem sítio de ligação estabelecem relações diretas entre si

A análise dos grupos de TFBSs contidos em cada CNE realizada pela ferramenta *STRING* v.11.0 (Szklarczyk et al., 2019) mostrou que, dentro de cada CNE, essas moléculas estabelecem redes em diversos processos biológicos. O banco de dados *STRING* tem como objetivo coletar, pontuar e integrar todas as fontes publicamente disponíveis de informações sobre interações proteína-proteína e complementá-las com previsões computacionais. Seu objetivo é alcançar uma rede global abrangente e objetiva, incluindo interações diretas (físicas) e indiretas (funcionais). Além disso, esta ferramenta integra o banco de dados KEGG (*Kyoto Encyclopedia of Genes and Genomes*), utilizado para a compreensão de funções biológicas a partir de informações de nível molecular, especialmente utilizando *datasets* moleculares de larga escala, provenientes de sequenciamento genômico e de tecnologias experimentais *high-throughput* (Kanehisa et al., 2017). Outro teste de enriquecimento do *STRING* consiste no mapeamento de todas as publicações que contêm os genes de entrada listados pelo usuário. Este recurso rastreia os genes pesquisados em todos os resumos do PubMed e em alguns textos adicionais, identificando publicações que são enriquecidas na classificação de entrada do usuário (Szklarczyk et al., 2019). Com essas ferramentas pudemos construir as redes de interação

proteína-proteína presentes nos CNEs. Como critério para inclusão nesta análise, foi considerada a ocorrência de, no mínimo, cinco TFBSs para cada CNE.

De acordo com estas análises, no CNE1, a rede de fatores N-MYC, MYC, MAX, FOXO1, TCF3 e CX1 destaca-se na via KEGG hsa05202 (Apêndice 4) pelo envolvimento em vias relacionadas a diversos tipos de leucemia, sarcomas e cânceres epiteliais, por exemplo. Em adição, os fatores USF1, FOXO1, SREBF1 e MYC, presentes no CNE1, aparecem envolvidos em um estudo que trata de câncer hepático (Zhang and Wang, 2011). A rede de fatores MYC, MAX, N-MYC, USF1 e ARNT foi demonstrada estar presente em uma região regulatória do gene *mir-142*, que codifica o microRNA-142, sendo, este último, capaz de modular a proliferação e, possivelmente, a migração e a capacidade de formação de colônias das células mesenquimais (Skårn et al., 2013). O aumento do microRNA-142 também já foi associado ao câncer colorretal (Toyota et al., 2008). Já a rede de fatores ARNT, USF1, TCF3, N-MYC, MYC e MAX está relacionada à resposta mediada por dieta, mostrando enriquecimento de termos em relação à pesquisa de Karweina de colaboradores (2015) (Karweina et al., 2015), enquanto os fatores CRX, FOXD1, MYC, N-MYC, SREBF1, FOXO1, TCF3 e POU6F1 aparecem ligados à adipogênese conforme associação com o artigo de Ambele & Pepper (2017) (Ambele and Pepper, 2017). É importante ressaltar que este é um contexto – adipogênese e resposta a estímulo nutricional – onde *Dact1* conhecidamente atua (Lagathu et al., 2009). Por fim, o artigo de Altman e colaboradores (2015) relaciona os fatores USF1, MYC e CLOCK aos distúrbios acometidos ao ciclo circadiano e ao metabolismo das células cancerosas (Altman et al., 2015) (Figura 6).

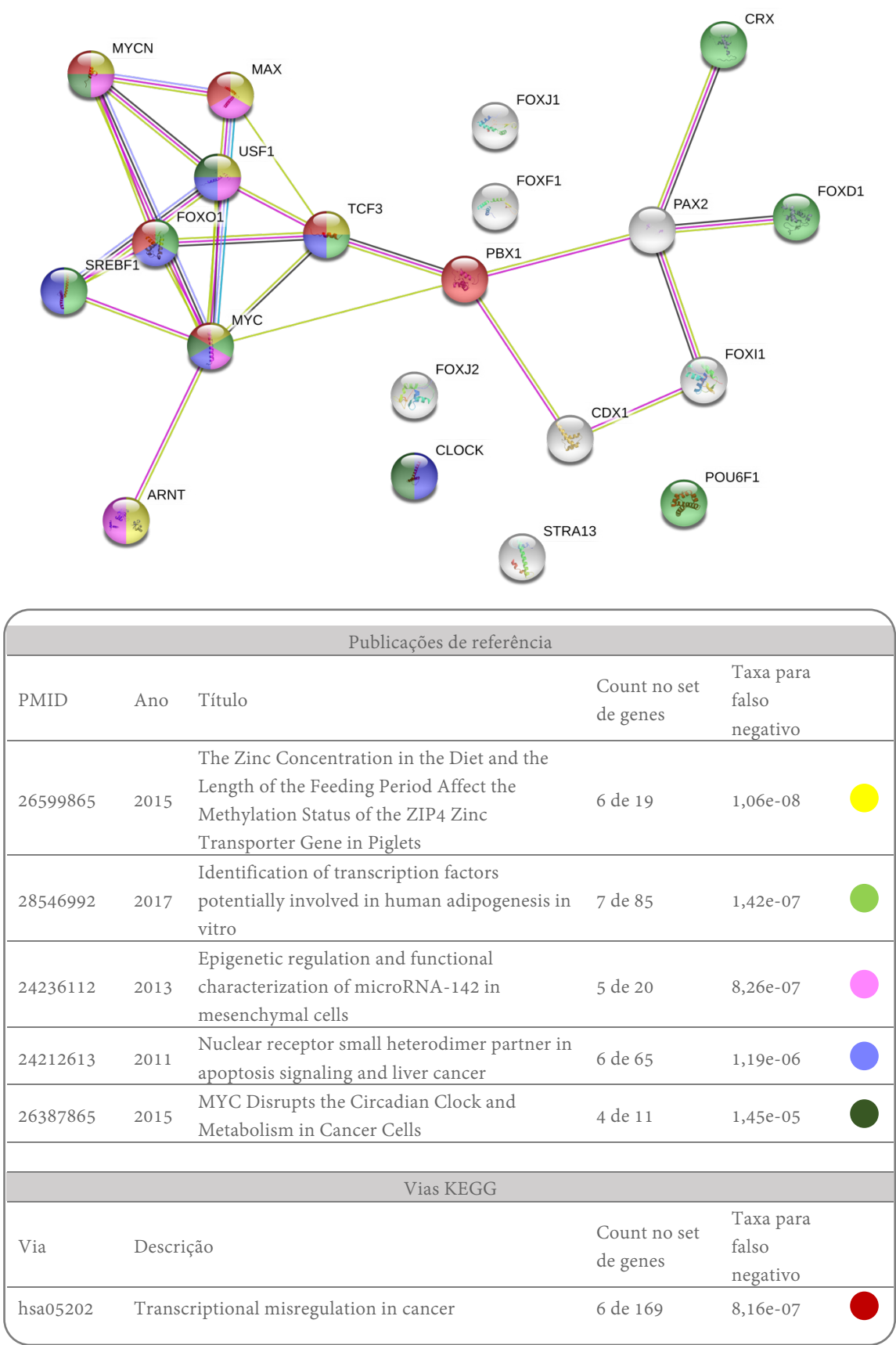
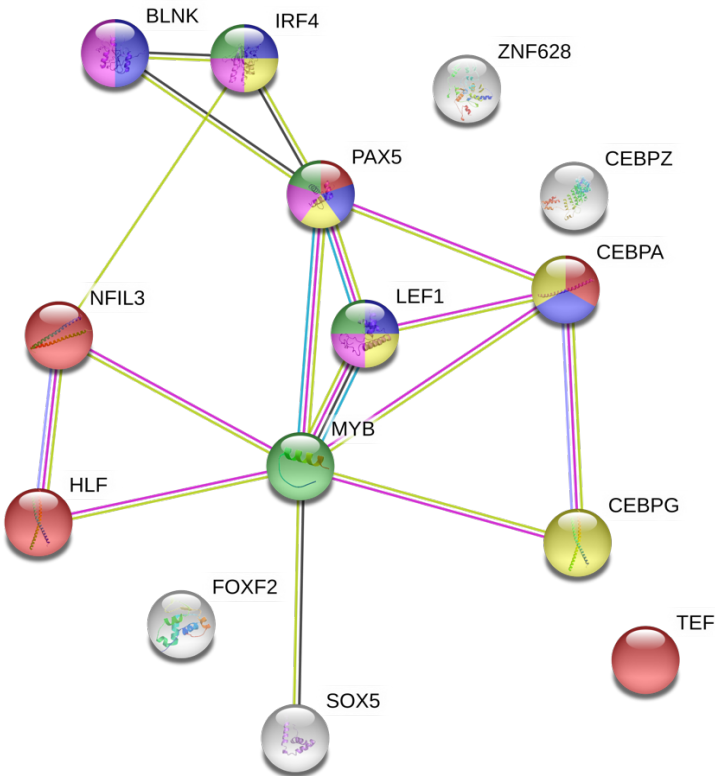


Figura 6. Rede de interações proteína-proteína gerada pelo STRING para o CNE1. Em amarelo estão destacadas as interações referentes ao estudo sobre os fatores em resposta à dieta; em azul está a rede relacionada ao câncer hepático; em verde destaca-se a rede de interações na adipogênese; em lilás estão as

ligações entre os fatores envolvidos na regulação da apoptose em câncer hepático e em verde escuro estão as ligações entre fatores da regulação circadiana e do metabolismo em células cancerosas. Por fim, em vermelho, estão as interações da via KEGG hsa05202 que remetem à desregulação transcricional em diversos tipos de câncer. Linhas em cor rosa indicam interações determinadas experimentalmente e, em azul turquesa, interações conhecidas em banco de dados curados; linhas em cor verde claro indicam correlações presentes na literatura científica; em lilás, homologia entre proteínas e, em verde escuro, co-expressão. Count no set de genes indica o total de genes listados em cada estudo e quantos destes estão presentes no input desta análise. PPI enrichment p-value: 5,17e-12.

A análise de interações proteína-proteína para os sítios do CNE2 mostrou que esta rede de moléculas aparece fortemente associada a processos de diferenciação de células do sistema linfático, estando, inclusive, o fator HLF intimamente ligado a situações patológicas como o câncer linfático (Figura 7). Por exemplo, foi observado um envolvimento dos fatores de transcrição presentes no CNE2 nos artigos listados na Figura 7, os quais relacionam majoritariamente esses fatores à diversos processos do sistema linfático humano, inclusive a doenças associadas, como as leucemias.

No artigo de Wahlestedt e colaboradores (2017) é mostrado um envolvimento dos fatores E4BP4 (NFIL3), CEBPA, PAX5 e TEF na modulação do destino da linhagem hematopoiética através de HLF (Wahlestedt et al., 2017). Neste mesmo contexto, aparecem outros estudos que relacionam, além destes fatores citados, NFAT (BLNK), ISRE (IRF4), LEF, CEBPG e MYB à diversos processos da diferenciação das linhagens linfopoiéticas (Guo et al., 2018; Rumfelt et al., 2006), bem como com doenças típicas relacionadas, como as leucemias (Gutiérrez et al., 2007). É importante ressaltar que os fatores de transcrição LEF-1/TCF medeiam uma resposta nuclear aos sinais Wnt ao interagir com β -catenina para ativar a transcrição de genes-alvo específicos (Eastman and Grosschedl, 1999). Importaneamente, *Dact1* é muito bem caracterizado como um modulador da via Wnt/ β -catenina. A presença de sítios de ligação para LEF-1/TCF no CNE2 indica que a transcrição de *Dact1* é modulada pela via Wnt/ β -catenina, o que sugere um mútuo controle entre a proteína Dact1 e a via Wnt/ β -catenina. Por fim, tendo MYB como fator central, esses fatores, juntamente com IRF4 e PAX4, também aparecem relacionados em um estudo sobre a modulação da via Wnt/ β -catenina no desenvolvimento das células B (Yu et al., 2008).

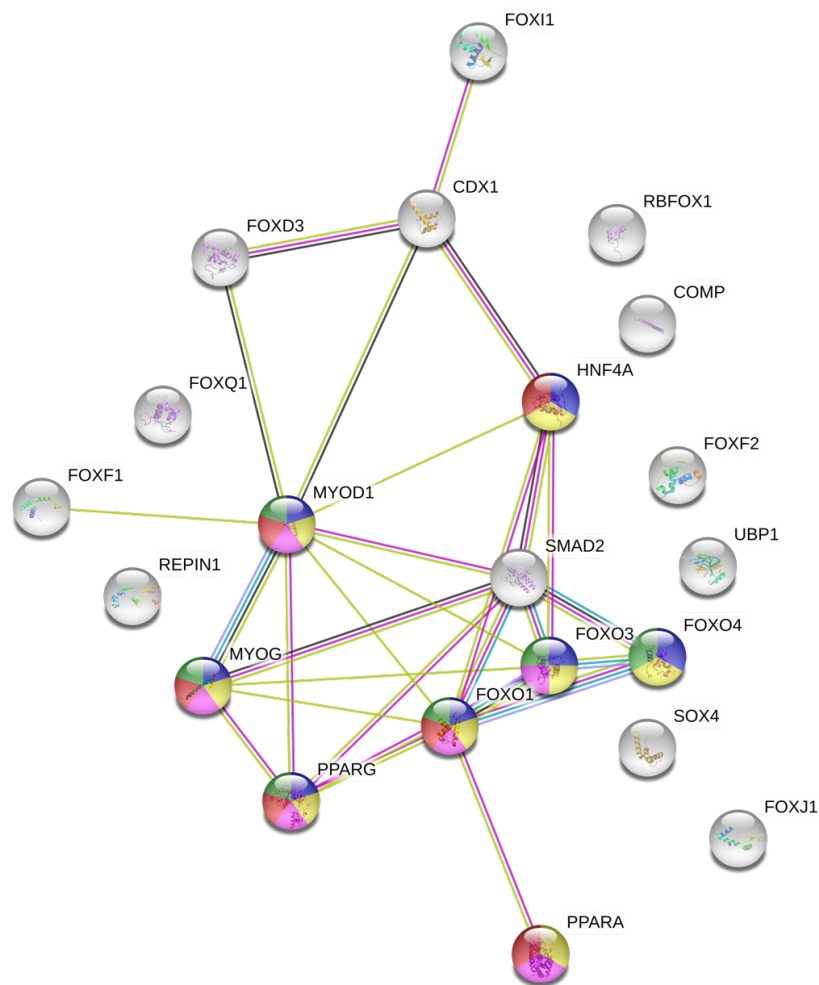


Publicações de referência				
PMID	Ano	Título	Count no set de genes	Taxa para falso negativo
29166614	2017	Critical Modulation of Hematopoietic Lineage Fate by Hepatic Leukemia Factor	5 de 37	2,19e-06
16505143	2006	Lineage specification and plasticity in CD19- early B cell precursor	5 de 41	2,19e-06
30061199	2018	Progression from the Common Lymphoid Progenitor to B/Myeloid PreproB and ProB Precursors during B Lymphopoiesis Requires C/EBPα	5 de 54	3,58e-06
17252022	2007	Gene expression profiling of B lymphocytes and plasma cells from Waldenström's macroglobulinemia: comparison with expression patterns of the same cell counterparts from chronic lymphocytic leukemia, multiple myeloma and normal individuals	4 de 13	3,58e-06
18768830	2008	Role of beta-catenin in B cell development and function	4 de 34	3,39e-05

Figura 7. Rede de interações proteína-proteína gerada pelo STRING para o CNE2. Os TFBSs desta região destacam-se pela associação com a modulação transcricional em células do sistema linfático. Em vermelho estão as moléculas que modulam o destino da linhagem hematopoiética através de HLF; em lilás, destacam-se as proteínas envolvidas na diferenciação das células B; em amarelo destaca-se a rede de fatores envolvidos

na diferenciação dos progenitores linfocitários na linfopoiese B através de CEBP α ; em rosa estão os fatores ligados à leucemia linfocítica crônica e, por fim, em verde está destacada a rede relacionada à β -catenina e o desenvolvimento das células B. Linhas em cor rosa indicam interações determinadas experimentalmente e, em azul turquesa, interações conhecidas em banco de dados curados; linhas em cor verde claro indicam correlações presentes na literatura científica; em lilás, homologia entre proteínas e, em verde escuro, co-expressão. Count no set de genes indica o total de genes listados em cada estudo e quantos destes estão presentes no input desta análise. PPI enrichment p-value: 4,95e-14.

As interações reveladas para o CNE4 remetem, em sua grande maioria, a processos do contexto muscular: regulação em células miogênicas, distrofia muscular de Duchenne (DMD), rabdomiossarcomas e também aparecem relacionadas aos efeitos anti-envelhecimento em resposta à restrição calórica em células musculares cardíacas e esqueléticas (Figura 8). No artigo de Chung e colaboradores (2013) é discutido o envolvimento de fatores da família FOXO (FOXO1, FOXO3 E FOXO4) na regulação transcricional para definir o nível de metabolismo oxidativo no músculo esquelético, junto com outros fatores presentes no CNE4, como MYOG, MYOD1, PPARG e HNF4A (Chung et al., 2013). Este mesmo conjunto de fatores, somados a PPARG, também aparecem relacionados aos efeitos anti-envelhecimento em resposta à restrição calórica em células musculares cardíacas e esqueléticas, através da mediação por sirtuínas (Zullo et al., 2018). Foi observado também um envolvimento de fatores MYOD1, MYOG, PPARG, PPARG, FOXO1 e FOXO3 no desenvolvimento da atrofia do músculo esquelético (Manickam et al., 2018). Na mesma forma, estes fatores aparecem relacionados à distrofia muscular de Duchenne (González-Sánchez et al., 2018). Finalmente, a rede HNF4A, MYOD1, MYOG, PPARG, PPARG e FOXO1 aparece implicada em um artigo de Ragab e colaboradores (2018) que relaciona a via Wnt/ β -catenina canônica com o desenvolvimento de rabdomiossarcomas (Ragab et al., 2018).



Publicações de referência				
PMID	Ano	Título	Count no set de genes	Taxa para falso negativo
23639108	2013	FoxO6 and PGC-1α form a regulatory loop in myogenic cells	7 de 20	1,78e-11
29561771	2018	Sirtuins as Mediator of the Anti-Ageing Effects of Calorie Restriction in Skeletal and Cardiac Muscle	8 de 55	3,95e-11
30115857	2018	Metronidazole Causes Skeletal Muscle Atrophy and Modulates Muscle Chronometabolism	6 de 28	5,00e-09
30568936	2018	Canonical WNT/β-Catenin Signaling Plays a Subordinate Role in Rhabdomyosarcomas	6 de 33	1,02e-08
30216693	2018	Improvement of Duchenne muscular dystrophy phenotype following obestatin treatment	6 de 54	7,26e-08

Figura 8. Rede de interações proteína-proteína gerada pelo STRING para o CNE4. Em lilás está destacada a rede de regulação em células miogênicas por meio de fatores da família FOXO; em amarelo, aparecem as moléculas associadas aos efeitos anti-envelhecimento em resposta à restrição calórica em células musculares cardíacas e esqueléticas mediadas por sirtuínas; em rosa destaca-se a rede de fatores envolvida na atrofia muscular e o cronometabolismo deste tecido; em vermelho estão representados os fatores

regulados em rabdomiossarcomas pela via Wnt/ β -catenina, e, finalmente, em verde as moléculas relacionadas à distrofia muscular de Duchenne. Linhas em cor rosa indicam interações determinadas experimentalmente e, em azul turquesa, interações conhecidas em banco de dados curados; linhas em cor verde claro indicam correlações presentes na literatura científica; em lilás, homologia entre proteínas e, em verde escuro, co-expressão. Count no set de genes indica o total de genes listados em cada estudo e quantos destes estão presentes no input desta análise. PPI enrichment p-value: 1,11e-16.

A análise do CNE5 mostrou que a rede de fatores de transcrição não tem significativamente mais interações do que o esperado. Isso significa que o conjunto de proteínas pode não ser grande o suficiente para inferir interações ou que se trata essencialmente de uma coleção aleatória de proteínas que não estão muito bem conectadas. Isso não significa, necessariamente, que não seja uma seleção de proteínas biologicamente significativa – pode simplesmente ser que essas proteínas não foram muito estudadas e que suas interações podem ainda não ser conhecidas pelo *STRING* (Szklarczyk et al., 2019).

5.3.6 O potencial regulatório do CNE4 foi validado por Hi-C/Virtual 4C

Com o navegador genômico *3D Genome Browser* (Wang et al., 2018) foram analisados os dados de *Hi-C/Virtual 4C* utilizando como *bait* a coordenada inicial do *core* de cada CNE para os quais há correspondência com os elementos reguladores do *Ensembl*, identificados em 5.3.2 e com potencial funcional determinado pelas análises de interação entre as redes de TFBSs, realizados em 5.3.5. Os contextos celulares onde as regiões dos CNEs 1, 2 e 4 têm potencial atividade foram determinados de acordo com o obtido em 5.3.3. Os resultados mostraram que apenas a região que compreende o CNE4 possui *reads* gerados pelos sequenciamentos *high-throughput*, em amostra de músculo Psoas, depositados no banco de dados da plataforma, corroborando a hipótese de que este segmento genômico se liga à outras regiões da cromatina no contexto muscular, representadas pelos picos da Figura 9A. Para os CNEs 1 e 2 não foi possível determinar a interação com *Dact1*, dentre o limitado número de tipos celulares catalogados na plataforma. Esta observação pode se dever ao fato de que, embora os dados de *Hi-C* possam demonstrar ligações entre pequenas regiões do cromossomo, a maioria destes mapas publicados atualmente estão em resolução de 10-40 kb e, portanto, não são ideais para

descobrir as interações dos elementos reguladores com promotores gênicos a curta distâncias (Wang et al., 2018), além da própria limitação quanto ao número de tipos celulares testados e depositados. Como o observado nesta análise, embora haja indicadores de interações entre as regiões e o promotor de *Dact1*, o banco de dados conta com resoluções de 40 kb o que limita a identificação exata de coordenadas que são muito curtas, como ocorre com os CNEs identificados. Assim, este tipo de análise *in silico* serve como um importante indicativo, porém não substitui outros métodos experimentais mais refinados. Métodos baseados em captura de sequência, como o *Capture Hi-C* ou *ChIA-PET*, geralmente, têm maior resolução e, portanto, são mais eficazes na identificação de interações da cromatina entre o gene e seus elementos cis-reguladores.

Em adição, para melhor examinar as ligações promotor-regulador, também exploramos os dados de ligação por locais hipersensíveis à DNase I (DHS) nesta região no mesmo contexto celular adotado acima. Os DHSs são marcadores de DNA com papel regulador e são utilizados para caracterizar as mais diversas classes de elementos cis-reguladores, incluindo *enhancers*, promotores, isoladores e silenciadores. Este método, aqui utilizado, baseia-se no cálculo dos coeficientes de correlação de Pearson entre um gene proximal e os DHSs distais e apenas os pares (gene-DHS) com $PCC > 0,7$ e dentro de 500 kb são considerados ligados (Thurman et al., 2012). Os resultados mostraram que, tomando *Dact1* como *bait*, ele se liga ao segmento distal *downstream* correspondente à região que abriga o CNE4, que é hipersensível à DNase I (Figura 9B).

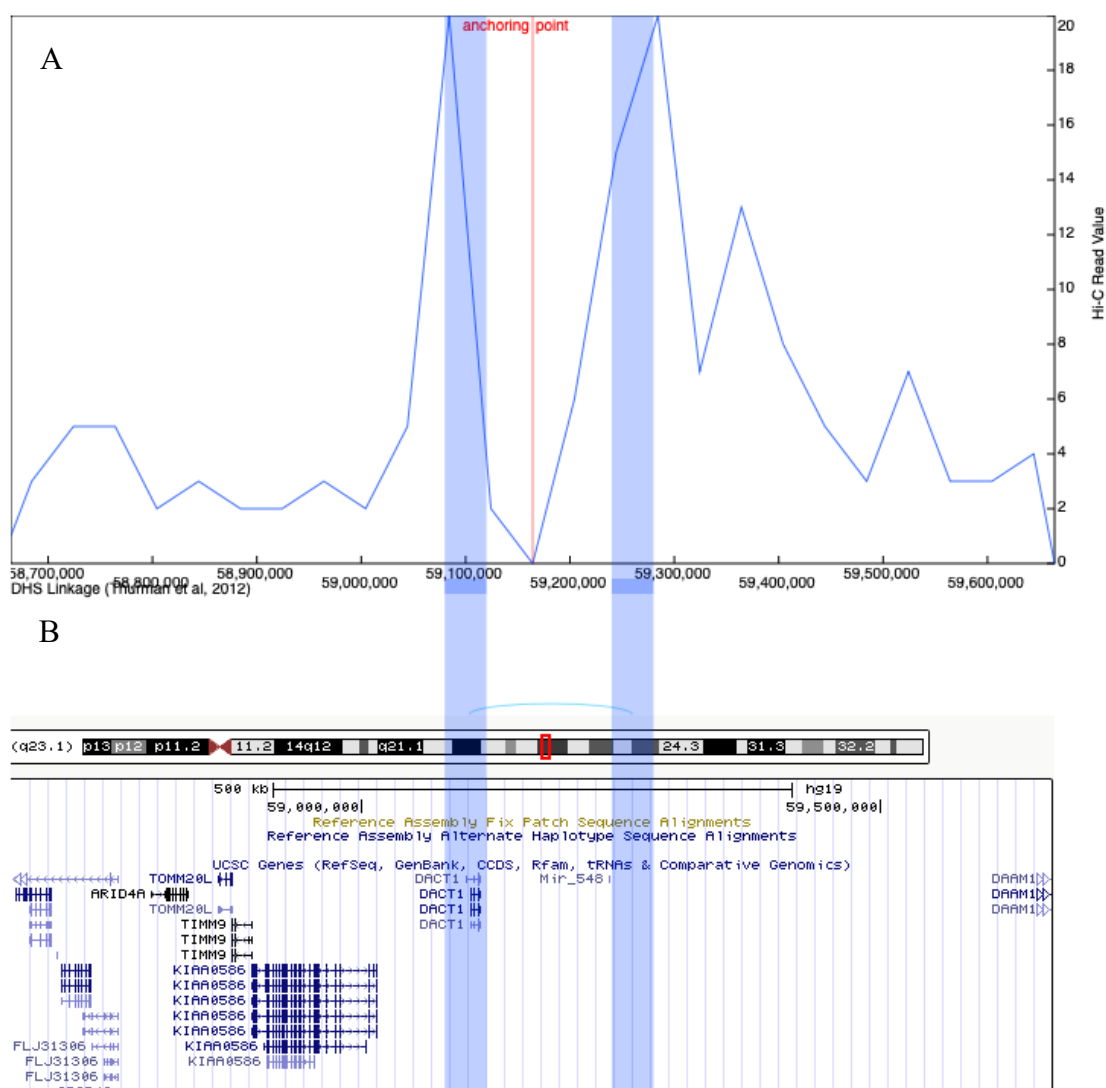


Figura 9. CNE4 Virtual 4C gerado pelo dado de Hi-C proveniente do músculo Psoas. A) Picos de reads identificados nas regiões que compreendem a coordenada inicial de Dact1 e a coordenada do CNE4, ambos delimitados pelas colunas em azul. Os picos indicam a frequência de eventos de interação da cromatina nos segmentos. B) Arco em azul representa a potencial interação entre a região de Dact1 e a região que abriga o CNE4, hipersensível à DNaseI. Montagem: hg19; resolução: 40 kb (Schmitt et al 2016).

5.3.7 A análise funcional do CNE4 de camundongo demonstrou sua capacidade de modificar a expressão diferencial da Luciferase em células HEK293T e C2C12

Depois de examinado o potencial regulatório dos CNEs, focamos no CNE4 para os ensaios funcionais, uma vez que a análise dos TFBSs deste elemento apresentou resultados promissores para ser uma potencial região regulatória de contextos como o da miogênese e da adipogênese e, além disso, se sobrepor à uma região caracterizada como *enhancer* pelo *Ensembl*. A atividade de transcrição do CNE4 de camundongo foi validada usando o ensaio de gene

repórter clássico com *Luciferase* (Figura 10). O vetor contendo o CNE4 e o promotor mínimo TK conduzindo a expressão da *Luciferase* como gene repórter foi capaz de atuar como um intensificador da expressão da *Luciferase* em cultura de células HEK293T e C2C12. Nestas células, a atividade do repórter foi significativamente maior do que a observada no plasmídeo de controle pGL2-TK-Luc não modificado (teste t de Student, pGL2-TK-Luc vs. CNE4, $p < 0,05$, indicado por * na Figura 10A-B). No entanto, os pré-adipócitos de camundongo imortalizados não exibiram alterações significativas na atividade da *Luciferase*.

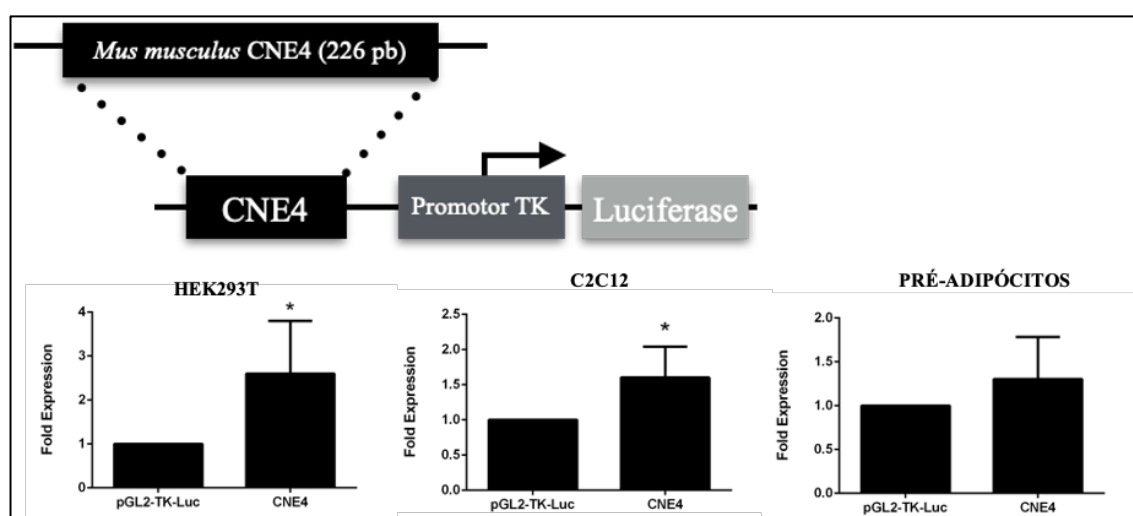


Figura 10. Gráficos da atividade do gene repórter da *Luciferase* dirigido pelo CNE4 murino. A construção de expressão contendo o CNE4 (0,5 μ g) foi transfectada em HEK293T, mioblastos (C2C12) e pré-adipócitos em estado proliferativo. O CNE4 de camundongo (*Mus musculus*) exibe atividade transcricional em cultura de células HEK293T e C2C12. Este efeito foi considerado significativo pelo teste t de Student ($p < 0,05$, *). Na cultura de células de pré-adipócitos imortalizados a diferença não foi considerada significativa. Valores obtidos em três experimentos independentes por triplicata.

5.3.8 O alinhamento múltiplo de sequências do CNE4 sugere haver potencial regulatório grupo-específico em contextos distintos.

Após a descoberta de que o CNE4 pertence a um segmento genômico no humano caracterizado como um *enhancer* pelo *Ensembl* (item 5.3.2; Tabela 1), objetivamos saber se outros animais de interesse (organismos-modelo) ou então aqueles economicamente importantes, como os bovinos, também possuíam esta sequência presente nos seus genomas. A par dessa identificação, analisamos se o CNE4 apresenta características grupo-específicas, por

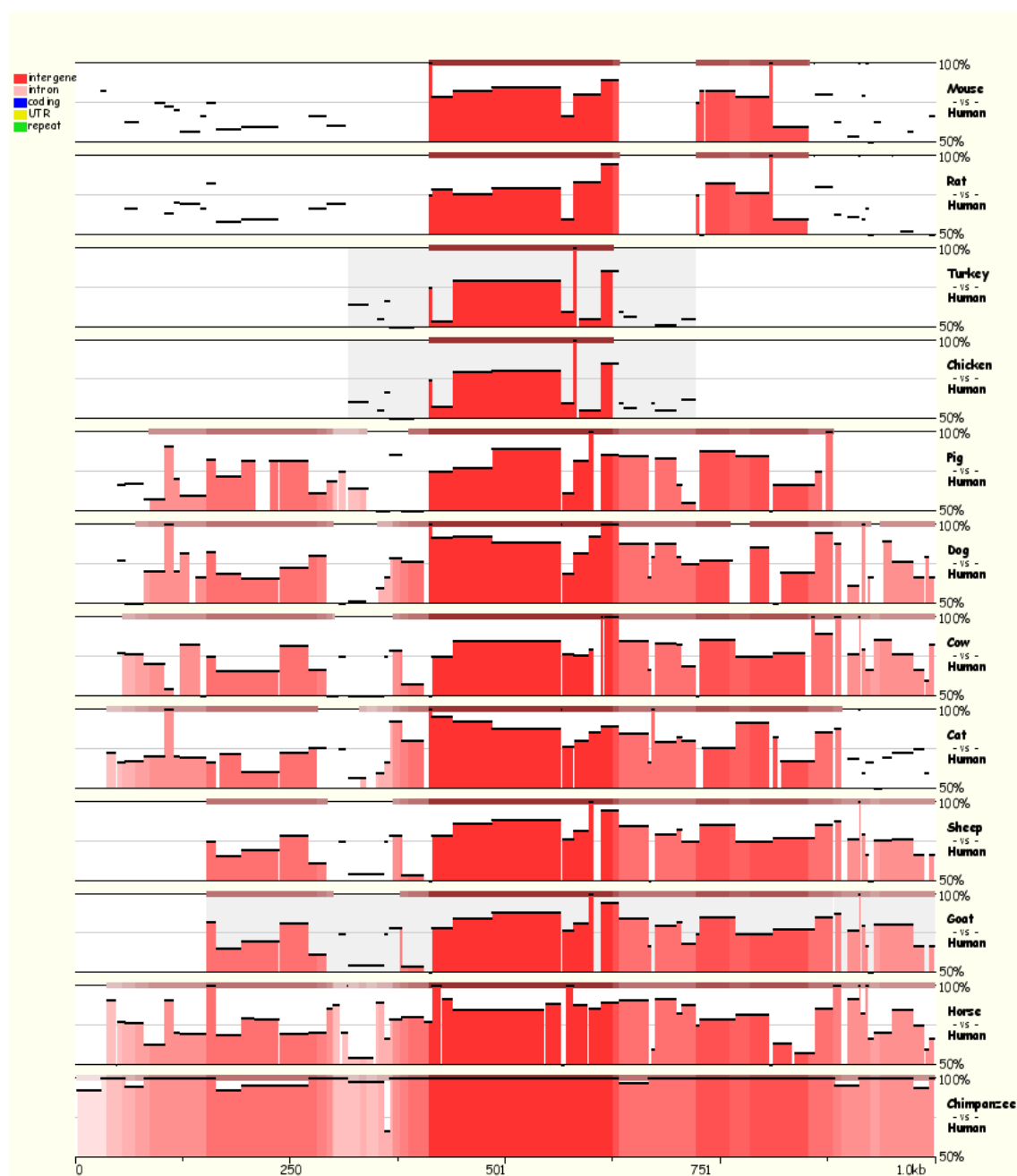
exemplo entre *Primates* e *Bovidae*, em relação a sua conservação, como um *fingerprint*, e também em relação aos TFBSs presentes.

Os alinhamentos múltiplos das sequências correspondentes ao CNE4 de diferentes animais (Apêndice 5), incluindo organismos-modelo e várias espécies de interesse econômico, mostraram que sequências ortólogas ao CNE4 estão em todas as espécies analisadas, as quais localizam-se no locus de *Dact1* (Tabela 6). Além disso, há um *core* comum aos mamíferos e aves (Figura 11A; Apêndice 6). Quando analisadas espécies mais próximas, por exemplo as pertencentes à família *Bovidae*, observa-se um maior nível de conservação evolutiva (Figura 11B).

Tabela 6. Localização genômica das sequências correspondentes ao CNE4 identificadas em 12 animais distintos, além do humano (*Homo sapiens*).

Organismo	Localização no genoma
Humano (<i>Homo sapiens</i>)	GRCh38: chr14:58777185-58778186
Chimpanzé (<i>Pan troglodytes</i>)	Pan_tro_3.0: chr14:43034200-43035302
Cão (<i>Canis lupus familiaris</i>)	CanFam3.1: chr8:34074699-34075700
Vaca (<i>Bos taurus</i>)	ARS-UCD1.2: chr10:71048740-71049741
Ovelha (<i>Ovis aries</i>)	Oar_rambouillet_v1.0: chr7:73579129-73580130
Cabra (<i>Capra aegagrus hircus</i>)	ARS1: chr10:32171403-32172404
Porco (<i>Sus scrofa</i>)	Sscrofa11.1: chr1:187854618-187855619
Cavalo (<i>Equus ferus caballus</i>)	EquCab3.0: chr24:6261570-6262571
Gato (<i>Felis catus</i>)	Felis_catus_9.0: chrB3:107913493-107914494
Rato (<i>Rattus norvegicus</i>)	Rnor_6.0: chr6:94149174-94150175
Camundongo (<i>Mus musculus</i>)	GRCm38: chr12:71437819-71438820
Peru (<i>Meleagris gallopavo</i>)	Turkey_2.01: chr5:57458957-57459958
Galinha (<i>Gallus gallus</i>)	GRCg6a: chr5:55230575-55231576

A)



B)

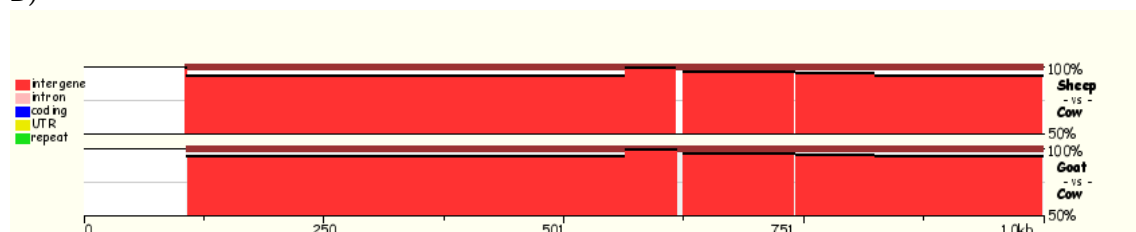
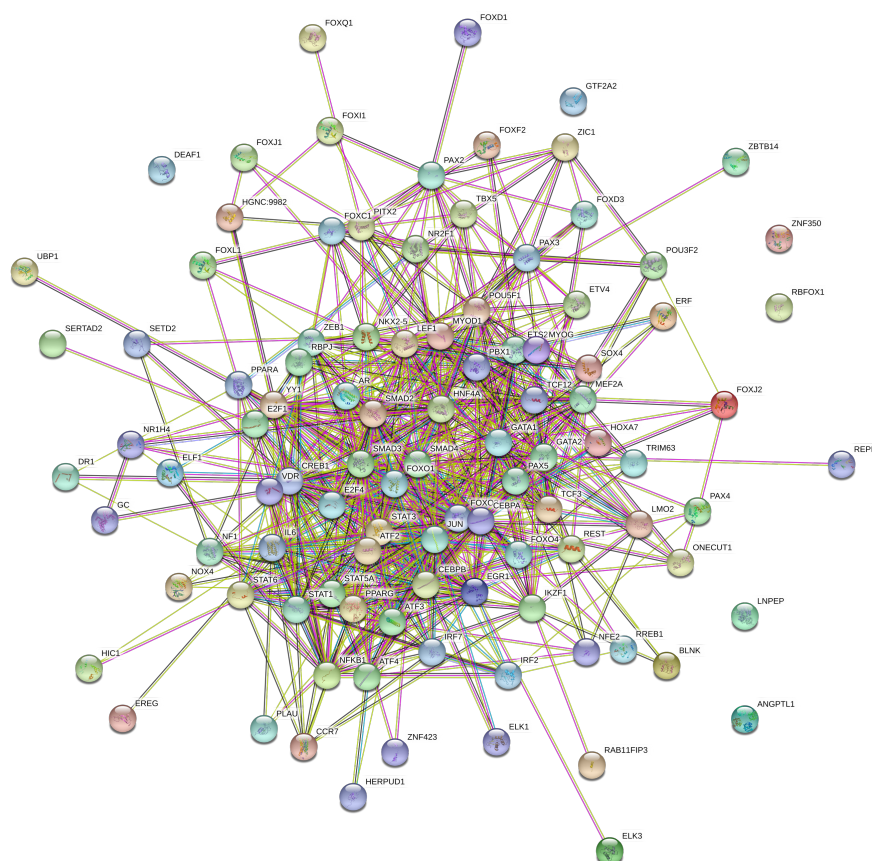


Figura 11. Comparação das sequências ortólogas do CNE4 entre as 12 espécies de animais estudadas e a sequência humana. A) Há um core comum a todas as espécies, quando galinha e peru são considerados nas análises. Dentre os mamíferos, a região conservada com o humano estende-se por toda a extensão do CNE4. B) Comparação do CNE4 da vaca com a sequência ortóloga de ovelha e cabra, mostrando uma maior conservação evolutiva entre espécies próximas.

As análises de conservação dos TFBSs presentes no CNE4 mostraram que há três TFBSs conservados entre todas as espécies de animais quando comparadas com o humano: XVENT1, SMAD e HFH3 (Figura 12A); no entanto, centenas de TFBSs são comuns entre espécies mais próximas, como vaca, ovelha e cabra, que pertencem ao grupo dos *Bovidae* (Figura 12B) e entre os *Primates*, humano e chimpanzé (Figura 12C).

A partir dos TFBSs conservados entre *Primates* (humano e chimpanzé) e daqueles conservados entre *Bovidae* (vaca, ovelha e cabra) foram estabelecidas as redes de interações proteína-proteína para estes dois grandes grupos utilizando o *STRING*. Como esperado, estes grupos apresentam redes bastante distintas, sugerindo que, embora o CNE4 seja uma região evolutivamente conservada, ele possivelmente divergiu a fim de suprir diferentes demandas regulatórias em diferentes espécies. Assim, vários sítios de ligação comuns foram perdidos, porém novos sítios apareceram e, então, devem ter sido seletivamente mantidos para desempenhar papéis importantes na regulação gênica destas espécies.

A Figura 13 mostra as redes entre os fatores de transcrição para os quais o CNE4 dos primatas possui sítios de ligação. A extensa rede remete principalmente às vias KEGG de diversas doenças, como vários tipos de câncer, diabetes, e também aos processos de envelhecimento celular. Quando observados os *clusters* de proteínas formados (*Local STRING Network Cluster*), que são agrupamentos de proteínas pré-computados hierarquicamente a partir de toda a rede *STRING* por um algoritmo de ligação média (*average linkage*), nota-se a evidente ligação desses fatores à miogênese, ao comprometimento celular e também aos processos diversos regulados pela família FOXO. Estas características recapitulam fortemente aquelas encontradas quando analisadas as redes do CNE4 conservado entre humano, camundongo e galinha. Na análise das redes estabelecidas entre os fatores de transcrição para os quais o CNE4 dos bovídeos possui sítios de ligação observa-se pouca similaridade com as dos primatas (Figura 14). No entanto, a análise de enriquecimento funcional dos TFBSs deste CNE para vias KEGG em bovinos revelou uma forte correlação com vias relacionadas ao câncer. A via bta05166 diz respeito à infecção por vírus HTLV-1, que pertence à família Oncovírus, responsável pela leucemia bovina (BVL) (Hajj et al., 2012).



Vias KEGG			
<i>via</i>	<i>descrição</i>	<i>count no set de genes</i>	<i>taxa de falso negativo</i>
hsa05161	Hepatitis B	15 de 142	4,07e-13
hsa05202	Transcriptional misregulation in cancer	14 de 169	2,85e-11
hsa05166	HTLV-I infection	16 de 250	2,85e-11
hsa05200	Pathways in cancer	17 de 515	4,03e-08
hsa04933	AGE-RAGE signaling pathway in diabetic complications	12 de 98	2,85e-11
hsa04659	Th17 cell differentiation	10 de 102	8,50e-09
hsa05321	Inflammatory bowel disease (IBD)	8 de 62	5,41e-08
hsa05215	Prostate cancer	9 de 97	6,72e-08
hsa05203	Viral carcinogenesis	10 de 183	9,23e-07
hsa04550	Signaling pathways regulating pluripotency of stem cells	8 de 138	9,45e-06
hsa05212	Pancreatic cancer	7 de 74	2,78e-06
hsa04218	Cellular senescence	8 de 156	2,11e-05
hsa05210	Colorectal cancer	6 de 85	5,88e-05
hsa04211	Longevity regulating pathway	7 de 88	7,27e-06
hsa04068	FoxO signaling pathway	8 de 130	7,27e-06
hsa05167	Kaposi's sarcoma-associated herpesvirus infection	8 de 183	5,71e-05
hsa04668	TNF signaling pathway	7 de 108	2,11e-05

Redes de <i>clusters</i> locais do STRING (<i>Local STRING Network Clusters</i>)			
cluster	descrição	count no set de genes	taxa de falso negativo
CL:5182	Cell fate commitment	22 de 187	1,66e-20
CL:5414	CDO in myogenesis, and regulation of stem cell differentiation	12 de 71	1,83e-12
CL:5415	CDO in myogenesis, and cardiac conduction system development	10 of 53	8,19e-11
CL:5416	CDO in myogenesis and positive regulation of cardioblast differentiation	8 de 30	1,17e-09
CL:5183	Homeobox conserved site, and cell fate commitment	10 de 116	5,68e-08
CL:5184	Homeobox, conserved site, and Maturity onset diabetes of young	9 de 95	1,58e-07
CL:5419	Myogenic Basic domain, and Holliday junction regulator protein family C-terminal	5 de 12	8,11e-07
CL:2376	Transactivation domain of FOXO protein family	4 de 6	4,45e-06
CL:2190	regulation of TOR signaling, and MARK targets/Nuclear events mediated by MAP kinases	10 de 196	3,99e-06
CL: 2482	AP-1 transcription factor, and transcription factor Jun	4 de 12	3,39e-05
CL:5186	Neuron fate commitment, and Homeobox KN domain	6 de 62	3,62e-05

Figura 13. Análise das redes de fatores do CNE4 humano vs. chimpanzé (Primates). Vias KEGG e redes de clusters associadas aos TFBSs conservados entre os primatas analisados. Linhas em cor rosa indicam interações determinadas experimentalmente e, em azul turquesa, interações conhecidas em banco de dados curados; linhas em cor verde claro indicam correlações presentes na literatura científica; em lilás, homologia entre proteínas e, em verde escuro, co-expressão. Count no set de genes indica o total de genes listados em cada estudo e quantos destes estão presentes no input desta análise. PPI enrichment p-value: < 1.0e-16.

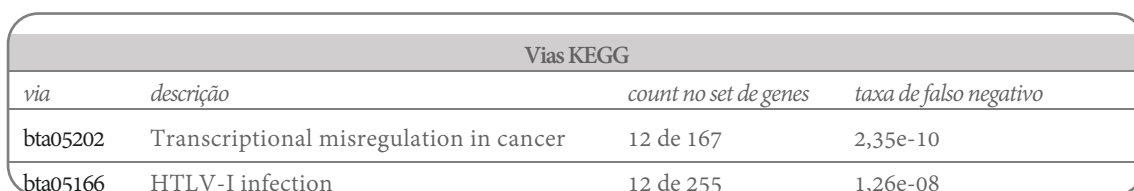


Figura 14. Análise das redes de fatores do CNE4 vaca vs. ovelha vs. cabra (Bovidae). Vias KEGG associadas aos TFBSs conservados entre os bovinos analisados. Linhas em cor rosa indicam interações determinadas experimentalmente e, em azul turquesa, interações conhecidas em banco de dados curados; linhas em cor verde claro indicam correlações presentes na literatura científica; em lilás, homologia entre proteínas e, em verde escuro, co-expressão. Count no set de genes indica o total de genes listados em cada estudo e quantos destes estão presentes no input desta análise. PPI enrichment p-value: $< 1.0e-16$.

5.4 DISCUSSÃO

O estudo de elementos cis-reguladores contribuiu para revelar mutações associadas à inúmeras doenças humanas, bem como para elucidar os mecanismos envolvidos na diversificação fenotípica entre os animais (Mathelier et al., 2015; Stenson et al., 2009; Wray, 2007). Devido à grande relevância deste tipo de estudo, nesta tese tivemos como objetivo identificar e caracterizar possíveis elementos reguladores da transcrição de *Dact1*, um gene regulador do desenvolvimento embrionário, que também vem sendo amplamente associado a patologias humanas. Utilizando ferramentas de genômica comparativa, um importante indicador da atividade desses elementos, pudemos identificar regiões com potencial regulatório em torno deste gene. Tendo *Dact1* como personagem central, buscamos então correlacionar as características dos CNEs identificados pelas predições *in silico* com as funções descritas e os contextos celulares onde este gene age, por meio de banco de dados que contêm informações sobre interações proteína-proteína, outra importante ferramenta para predizermos as possíveis relações funcionais entre *Dact1* e os sítios de ligação para fatores de transcrição identificados em diferentes segmentos deste locus gênico.

5.4.1 O que se sabe sobre a regulação transcricional do gene *Dact1*?

Os estudos voltados para a identificação e a caracterização de elementos cis-reguladores do gene *Dact1* são escassos, justificando a necessidade de pesquisas nesta área, como a que foi realizada nesta tese. Portanto, esta discussão se inicia pela apresentação dos principais achados dos dois principais artigos que abordam a regulação transcricional de *Dact1*, para situar o leitor em relação ao que se sabe sobre este tema até o momento e permitir uma comparação com os achados desta tese.

Suzuki e colaboradores (2014) criaram camundongos transgênicos contendo um fragmento de 6 kb *upstream* ao gene *Dact1*, que incluiu o promotor gênico, dirigindo a expressão do gene repórter *LacZ*. Este estudo demonstrou que este fragmento contém elementos cis-reguladores da transcrição de *Dact1*, que atuam em múltiplos tecidos derivados do mesoderma e do neuroectoderma durante o desenvolvimento embrionário. Além disto, os autores observaram que um subconjunto de tecidos adultos, incluindo o cérebro, olhos, coração e alguns órgãos reprodutores, também têm sua expressão regulada por esta região reguladora (Suzuki et al., 2014). Nestes ensaios o repórter *LacZ* foi altamente

expresso em vários tipos de neurônios e estruturas do cérebro dos camundongos transgênicos, como o hipocampo, córtex cerebral, prosencéfalo ventral, tálamo e os lobos temporais mediais, revelando que o fragmento de 6 kb avaliado neste artigo contém elementos cis-reguladores que controlam a expressão de *Dact1* nestas estruturas do sistema nervoso.

Em um outro estudo, realizado por Le Guen e colaboradores (2020), foi utilizado o banco de dados *UK Biobank*, que inclui genótipos de milhares de indivíduos e suas respectivas imagens de ressonância magnética. Neste artigo, os autores demonstraram que variações genéticas em um *enhancer* de *Dact1* modula características anatômicas e funcionais do sulco temporal superior em seres humanos. Este *enhancer* localiza-se a 26,5 kb *upstream* ao gene *Dact1* e modula a profundidade de uma parte do cérebro chamada de sulco temporal superior (STS), responsável pelo desenvolvimento da linguagem. É digno de nota que esta característica anatômica difere entre os cérebros de humanos e dos chimpanzés (Le Guen et al., 2020), revelando que variações em um elemento cis-regulador de *Dact1* estão associadas ao desenvolvimento de um fenótipo específico (profundidade do STS) em seres humanos e que devem ter ocorrido ao longo da evolução dos primatas, destacando a importância de um *enhancer* específico de *Dact1* para diversificação fenotípica e funcional de um atributo do sistema nervoso.

Nesta tese encontramos vários CNEs no locus gênico de *Dact1*, os quais possivelmente são elementos cis-reguladores da transcrição deste gene que ainda não foram descritos, e que podem controlar a atividade deste gene em múltiplos contextos, por exemplo a miogênese, a adipogênese e inúmeros tipos de câncer, como aqueles do sistema digestório e imunológico, além de miopatias. Estes são alguns exemplos sobre quais serão discutidos a seguir.

5.4.2 Múltiplos CNEs com potencial regulador estão presentes na região intergênica entre os genes *KIAA0586*, *Dact1* e *Daam1*

A comparação do segmento genômico que contém o locus de *Dact1* revelou oito regiões evolutivamente conservadas entre humano, camundongo e galinha. Dentre estes segmentos de DNA, denominados elementos não codificadores conservados (CNEs), três se localizam na região intergênica entre *KIAA0586* e *Dact1*, enquanto cinco CNEs se localizam na região intergênica entre *Dact1* e

Daam1. Por meio da caracterização destes elementos, demonstramos haver fortes evidências de que *Dact1* é regulado por um conjunto de elementos cis-reguladores localizados em diferentes regiões deste intervalo genômico. Portanto, os elementos reguladores proximais, identificados no estudo com camundongos transgênicos *LacZ*, não devem ser os únicos reguladores de *Dact1*. Isto porque, neste estudo de Suzuki e colaboradores (2014) foi analisada uma região restrita a 6 kb *upstream* a *Dact1*. Os achados desta tese indicam haver elementos que estão em regiões mais distantes de *Dact1*, mas que também apresentam potencial de regulador transcricional. Também é importante ressaltar que a regulação da expressão de genes que atuam no desenvolvimento, como *KIAA0586*, *Dact1* e *Daam1*, em via de regra é modular. Isto é, elementos cis-reguladores diferentes controlam a expressão espacial e temporal dos genes em diferentes etapas, de tal forma que a perda de elementos específicos não inviabiliza a atividade global do gene, preservando ainda os outros meios onde são expressos.

A seguir, serão discutidos os contextos celulares putativos dos CNEs que apresentaram número suficiente de sítios de ligação para fatores de transcrição conservados entre humano, camundongo e galinha, capazes de gerar redes de interação proteína-proteína e então possibilitar o mapeamento da associação física e/ou funcional conhecida entre estas proteínas identificadas, relacionando os contextos em que *Dact1* possa ser modulado.

5.4.3 *A rede de interações entre os TFBSs do CNE1 apresenta características regulatórias em diversos tipos de câncer e de modulação em resposta às variações nutricionais*

O CNE1 é adjacente a um elemento regulador previamente identificado no *Ensembl* como uma região flanqueadora de promotor e se localiza a 72 kb *upstream* a *Dact1*. Este elemento regulador caracterizado no *Ensembl* tem atividade detectada em linhagens celulares de câncer de fígado e apresenta marcadores de repressão em alguns tipos de câncer, como o de mama e na leucemia de células T. Estes contextos regulatórios aparecem de acordo com aqueles observados nas análises das redes de fatores de transcrição identificadas no CNE1, pois estas remetem, majoritariamente, às vias ligadas ao desenvolvimento de diversos tipos de carcinomas, dentre eles o câncer hepático epitelial e leucemias. O carcinoma hepatocelular (HCC) é um dos tipos de câncer mais comuns no mundo (Eric C, 2008) e foi demonstrado que *Dact1* está modulado negativamente em 43% das amostras de HCC

humanas (Clarke et al., 2014; Yau et al., 2005). Assim sendo, é possível que o CNE1 esteja envolvido em aspectos da regulação transcricional de *Dact1* relacionados ao seu papel como um supressor tumoral. Por exemplo, os fatores SREBP1, USF1 e FOXO1 – encontrados no CNE1 –, são conhecidos por interagir com o receptor nuclear SHP (do inglês, *Small Heterodimer Partner*) (Zhang and Wang, 2011). SHP é um repressor transcricional regulado epigeneticamente que tem importante papel na supressão do câncer de fígado em humanos e em camundongos (He et al., 2008; Zhang et al., 2008), impedindo defeitos da via apoptótica e o aumento da proliferação celular, comumente ligados ao aparecimento de tumores (Zhang and Wang, 2011). Nas células do fígado, foi demonstrado que SHP, ao se ligar ao coativador de Nur77, leva à repressão da função transcricional desta proteína (Zhang and Wang, 2011). Paralelamente, em um outro tipo de câncer, o melanoma, Nur77 reprime *Dact1*, permitindo a ativação da via Wnt/ β -catenina nesta patologia (Farmer, 2016). Isto tudo sugere que *Dact1* pode ter sua expressão modulada negativamente por Nur77, por intermédio dos TFBSs do CNE1 que são capazes de se ligar ao SHP, por um mecanismo indireto.

Outros TFBSs presentes no CNE1 que têm um potencial papel regulador da transcrição de *Dact1* no contexto da carcinogênese são USF1, MYC, MYB, N-MYC e ARNT. Estes fatores foram identificados como reguladores da expressão do gene *mir-142*, que está envolvido na migração celular, na proliferação e na apoptose em diferentes cânceres humanos (Pahlavan et al., 2020). Dentre os tipos de câncer onde há uma desregulação do *mir-142* estão o câncer de pulmão, de mama e de cólon, nos quais há um envolvimento significativo de *Dact1* (Kaduthanam et al., 2013; Troschel et al., 2018; Zhiying et al., 2013). Importaneamente, o estudo publicado por Skårn e colaboradores (2013) mostrou que a transcrição do microRNA *mir-142* é silenciada por metilação nas células mesenquimais (Skårn et al., 2013). Neste trabalho foi observado a presença dos mesmos sítios de ligação conservados para USF1, MYC, MYB, N-MYC e ARNT, em uma região *upstream* ao gene *mir-142*. Estes fatores são capazes de se ligar a um E-box, que é importante para a ativação transcricional mediada por estes fatores (Griswold and Kim, 2001; Prendergast and Ziff, 1991). A metilação do E-box, conforme observada no estudo, mostrou diminuir significativamente sua afinidade de ligação para FTs como o USF1 e o MYC. Assim, a inibição da ligação do TF por metilação do DNA pode ser um mecanismo importante na regulação epigenética de *mir-142* e na sua participação no desenvolvimento de cânceres. Da mesma forma, este mesmo conjunto de TFBSs está presente no CNE1, *upstream* a *Dact1*, sugerindo

que esta região possa funcionar como um elemento regulador da mesma maneira observada na pesquisa de Skårn e colaboradores (2013). Finalmente, os fatores MYC e N-MYC, conservados no CNE1, também são fortemente associados a doenças humanas. Estima-se que MYC e N-MYC têm expressão alterada em pelo menos 50% dos cânceres (Dang, 2012). Em um estudo Altman e colaboradores (2015) demonstraram que MYC interrompe o relógio molecular e altera o metabolismo circadiano da glicose favorecendo o crescimento em células cancerosas. Estes fatores ativam diretamente reguladores negativos de CLOCK-BMAL1, TFBS também encontrado no CNE1. Esta supressão prediz resultados clínicos ruins em neuroblastomas humanos, por exemplo (Altman et al., 2015).

Um outro conjunto de TFBSs presentes no CNE1 aponta para um possível papel deste elemento cis-regulador no controle da transcrição de *Dact1* mediante estímulo nutricional. Neste sentido, é importante mencionar que *Dact1* é um gene modulado nutricionalmente, mobilizando o tecido adiposo em resposta à ingestão alimentar (Lagathu et al., 2009). Neste conjunto de fatores encontram-se ARNT, N-MYC, USF, TCF3 e MYC, os quais estão presentes no CNE1, e foram identificados como reguladores da transcrição do gene *ZIP4* (do inglês, *Zinc transporter 4*) no intestino delgado de porcos de maneira dependente da concentração de zinco na dieta (Karweina et al., 2015). O gene *ZIP4* codifica uma proteína transmembranar que é responsável pelo transporte de zinco em diferentes órgãos do sistema digestório e dos rins (Karweina et al., 2015). Isto sugere que este mesmo conjunto de fatores pode atuar respondendo às variações da dieta também em outros organismos, inclusive no humano, onde *Dact1* e a sinalização Wnt/ β -catenina exibem perfis de expressão coordenados em resposta ao estado alimentar e em que *Dact1* atua facilitando a adaptação do tecido adiposo às necessidades de armazenamento de gordura com base no estado nutricional e metabólico (Lagathu et al., 2009).

Neste mesmo contexto adipogênico, no qual *Dact1* controla a diferenciação dos adipócitos por meio da regulação negativa da sinalização Wnt/ β -catenina, os fatores CRX, FOXD1, MYC, N-MYC, SREBF1, FOXO1, TCF3 e POU6F1, presentes no CNE1, têm papel fundamental na adipogênese. Isto foi demonstrado no artigo de Ambele e Pepper (2017), que utilizaram células estromais/tronco derivadas do tecido adiposo humano (ASCs) para estudar a diferenciação adipogênica em diferentes momentos. Além dos já conhecidos fatores adipogênicos FOXO1A E FOXD1, também foi

demonstrado que *c-MYC*, sítio presente no CNE1, atua negativamente na diferenciação dos pré-adipócitos e que o fator de transcrição Ikaros (IK1) promove a diferenciação de células 3T3-L1 em adipócitos por meio da supressão da expressão de *c-Myc* (Park and Pyo, 2013), alvo da via Wnt/ β -catenina. Curiosamente, o sítio de ligação para IK1 foi aqui identificado como o único TFBS conservado no CNE7 que, por este motivo, não foi analisado com o *STRING*. Isto posto, estes novos achados sugerem que o CNE1 regula a transcrição de *Dact1* no contexto da adipogênese, por uma gama de fatores de transcrição aqui relacionados a este elemento cis-regulador, levando à modulação da via Wnt/ β -catenina.

5.4.4 *A rede de interações entre os TFBSs do CNE2 apresenta características regulatórias da expressão gênica em células do sistema linfático humano*

O CNE2 localiza-se a 29,2 kb *upstream* a *Dact1*. Ainda que este CNE não apresentasse sobreposição com nenhum elemento previamente identificado no *Ensembl*, foram realizadas as análises de interação entre os fatores de transcrição para os quais este segmento genômico apresenta sítios de ligação. Isto porque este elemento exibe um número significativo de TFBSs conservados entre o humano, o camundongo e a galinha, sugerindo que este possa ser um elemento cis-regulador até então não caracterizado nos bancos de dados disponíveis.

As análises realizadas para o CNE2 apontam fortemente para uma correlação com a regulação da expressão gênica de *Dact1* no sistema imunológico humano. Como abordado no artigo de Wahlestedt e colaboradores (2017), há um envolvimento dos fatores E4BP4 (NFIL3), CEBPA, PAX5 e TEF na modulação do destino da linhagem hematopoiética através de HLF (Wahlestedt et al., 2017). Importaneamente, todos os sítios para esses fatores de transcrição estão presentes no CNE2. A superexpressão de NFIL3 leva à diferenciação quase exclusiva em células NK (Kamizono et al., 2009), enquanto HLF/TEF favorece a mielopoiese, ou seja, a formação das células granulocíticas (eosinófilos, basófilos e neutrófilos) (Wahlestedt et al., 2017).

Durante a hematopoese, as linhagens linfoides (células B, T e NK) são originadas a partir de um progenitor linfóide comum (CLP) (Kondo et al., 1997). CLPs dão origem a progenitores de células B comprometidos, um processo guiado por uma rede de fatores de transcrição que inclui FOXO1, EBF1

e PAX5 (Mercer et al., 2011), sendo o primeiro TFBS encontrado no CNE1 e o último no CNE2. Neste contexto, é digno de nota que a via Wnt canônica preserva o fenótipo imaturo das células-tronco hematopoiéticas, sendo essencial para a autorrenovação desses progenitores (Staal and Clevers, 2005). Isto sugere que *Dact1*, como tipicamente um modulador negativo desta via, possa ser alvo de supressão nestes contextos das doenças linfocitárias, contribuindo para a superexpressão da via Wnt. Corroborando esta hipótese, Zhu e colaboradores (2017) demonstraram que a superexpressão de *Dact1* inibiu a sinalização de Wnt/ β -catenina ao reduzir os níveis de β -catenina nuclear em células KG-1 α , uma linhagem celular derivada de leucemia mieloide aguda. Os achados deste artigo revelaram um importante papel inibitório de *Dact1* no desenvolvimento e progressão desta patologia (Zhu et al., 2017). Em adição, foi demonstrado que as células B de pacientes com leucemia linfocítica crônica expressam de forma elevada vários genes-alvo da via Wnt em comparação com células B normais (Lu et al., 2004). Este achado mostra que a sinalização Wnt desregulada, ou seja, superexpressa, contribui para a leucemogênese (Staal and Clevers, 2005). Isto posto, este conjunto de observações sugere que o CNE2 pode atuar como um regulador de *Dact1* nas linhagens celulares linfocitárias, abrindo novas perspectivas para o aprofundamento sobre a expressão e a modulação deste gene neste contexto, fortemente implicado em doenças, e que até então é pouco explorado.

5.4.5 *A rede de interações entre os TFBSs do CNE4 apresenta características regulatórias em diversos contextos do músculo esquelético – do desenvolvimento até patologias*

O CNE4 localiza-se a 143,5 kb *downstream* a *Dact1*. Importaneamente, esta região se sobrepõe a um elemento regulador caracterizado no banco de dados do *Ensembl* como *enhancer*. Este *enhancer* apresenta marcadores de atividade em diversos tipo celulares e dentre estes destacam-se os neurônios e células progenitoras neurais (NPC), pulmões, intestinos grosso e delgado, placenta e o músculo esquelético. Além disso, este elemento regulador apresenta marcadores de repressão nas linhagens celulares de câncer de cólon e de adenocarcinoma de pulmão, duas patologias nas quais *Dact1* tem sua expressão alterada. Estes achados corroboram a hipótese de que o CNE4 atua como um elemento regulador de *Dact1* nestes contextos, uma vez que este gene está implicado em todos eles.

Nas análises das redes de interação proteína-proteína para os fatores correspondentes aos TFBSs presentes no CNE4 encontramos um conjunto de fatores-chave relacionados ao desenvolvimento da musculatura esquelética. Tendo como fatores centrais MYOD e MYOG, são diversos os estudos que estabelecem relações entre estes e outros fatores presentes no CNE4 com a miogênese e com miopatias, tais como a distrofia muscular de Duchenne (DMD) e rabdomiossarcomas. Os fatores regulatórios miogênicos (MRFs) compõem uma pequena família de fatores de transcrição *basic helix-loop-helix*, com papel central no desenvolvimento do músculo esquelético. Enquanto MYOD controla a determinação miogênica e as fases iniciais da miogênese, MYOG é necessária para a diferenciação dos mioblastos para que os progenitores miogênicos possam formar miofibras contráteis (Sabourin and Rudnicki, 2000). Neste contexto, já foi demonstrado que MYOD é capaz de se ligar a um elemento cis-regulador de *Dact1*, o qual dirige a sua expressão em mioblastos C2C12 na fase de proliferação celular (50% de confluência) (Santolini et al., 2016), corroborando a nossa hipótese de que *Dact1* atua em resposta aos fatores miogênicos nestas células. Em adição aos sítios para MRFs, o CNE4 apresenta sítios de ligação para diferentes fatores da família FOXO, conhecidos por sua função inibidora da miogênese e facilitadora da atrofia muscular (Sandri et al., 2004). Por isso, é também importante destacar que o CNE4 apresenta sítios para fatores que regulam a miogênese tanto positivamente, como MYOD e MYOG, quanto negativamente, como os fatores da família FOXO, sugerindo que a modulação de *Dact1* deve ocorrer de uma forma dependente do estágio de diferenciação das células musculares esqueléticas.

Os TFBSs presentes no CNE4 também remetem à ação das sirtuínas, que são proteínas com atividade enzimática, principalmente de deacetilases, as quais são capazes de regular a expressão dos fatores MYOD, MYOG, PPARG e HNF4 (Zullo et al., 2018), presentes neste CNE. É importante notar que a expressão das sirtuínas é também regulada pela dieta, tal como *Dact1*. De fato, Zhou e colaboradores (2016) demonstraram que a sirtuína 1 (SIRT1) atua como um supressor da adipogênese, através da desacetilação das histonas dos promotores de *sFRP1*, *sFRP2* e *Dact1*, inibindo a expressão gênica (Zhou et al., 2016). Como consequência, há uma ativação da via de sinalização Wnt canônica, o que leva à inibição do processo adipogênico, como discutido anteriormente. No entanto, as sirtuínas não estão implicadas apenas na adipogênese. Por exemplo, a restrição calórica foi descrita como um importante mecanismo nutricional que leva à ativação destas enzimas, conferindo às sirtuínas um

papel mediador das respostas metabólicas às mudanças na disponibilidade de nutrientes (Yang et al., 2011) e, neste contexto, sabe-se que o músculo esquelético é muito suscetível às variações no aporte de nutrientes (Aspnès et al., 1997). Estudos *in vitro* demonstraram que a expressão de SIRT1 estimula uma resposta antioxidante em células do músculo esquelético após estresse mecânico (Pardo and Boriek, 2011). Em miotubos cultivados em baixa concentração de glicose, a SIRT1 é necessária para a mudança metabólica da utilização da glicose para a oxidação dos ácidos graxos; esta resposta também acontece durante a restrição calórica e é reduzida nas doenças metabólicas (Gerhart-Hines et al., 2007), demonstrando o potencial terapêutico desse mecanismo de regulação, que implica na atividade de *Dact1*, uma vez que ele também é modulado por SIRT1, como mostrado por Zhou e colaboradores (2016). SIRT1 produz estes e outros efeitos ao interagir com seus substratos e cofatores transcricionais, por exemplo, PGC-1 α , MEF2, p53, HNF4, FOXOs, PPARG, MYOD e SMADs (Zullo et al., 2018), sendo estes cinco últimos, fatores para os quais o CNE4 possui sítios de ligação, enquanto o sítio para SREBP1 aparece na região do CNE1. Isto sugere que SIRT1 possa desempenhar um papel regulatório de *Dact1* não apenas pela deacetilação de seu promotor, mas também pela modulação da expressão de outros fatores de transcrição presentes no CNE4, que fazem parte da cascata de ativação de diversos processos metabólicos nos quais *Dact1* também participa.

Um outro contexto, agora patológico, no qual os fatores de transcrição do CNE4 estão envolvidos é o rabdomiossarcoma (RMS), um tumor maligno dos tecidos moles que acomete os mioblastos em diferenciação, principalmente crianças (Ragab et al., 2018). Este tipo de tumor apresenta significativa expressão de MYOD e MYOG, fatores para os quais o CNE4 possui sítios, além da expressão aumentada de Desmina (Saab et al 2011). Neste tipo de tumor, as células do músculo esquelético permanecem pouco diferenciadas, com muitas células apresentando um *status* de diferenciação comparável ao de mioblastos. Ragab e colaboradores (2018) demonstraram que LEF1, coativador de β -catenina na ativação de genes-alvo Wnt, deve ser responsável pela supressão da malignidade e pela indução da miodiferenciação das células RMS, enquanto a atividade da β -catenina desempenha apenas um papel subordinado nesses processos, demonstrando que a sinalização Wnt canônica aparentemente não influencia a incidência ou a taxa de crescimento do tumor (Dräger et al., 2017; Ragab et al., 2018). Em um tratamento de células tumorais com inibidor de β -catenina (FH535) foi demonstrado que a supressão tumoral ocorre devido a um efeito secundário do inibidor empregado,

que na verdade inibe os fatores de transcrição PPARG (Handeli and Simon, 2008; Ragab et al., 2018). Embora ainda não haja dados na literatura científica sobre a modulação de *Dact1* nesse tipo específico de sarcoma, há a possibilidade de que ele possa ser modulado neste contexto por meio dos fatores de transcrição da família PPAR, sítios também presentes no CNE4, além de MYOD e MYOG.

Embora os CNEs 5, 6, 7 e 8 não possuam TFBSs conservados suficientes para estabelecermos as redes de interação proteína-proteína, uma hipótese é que isto pode se dever ao fato destes elementos terem divergido evolutivamente e por isso apresentam poucos sítios conservados. Esta hipótese é corroborada pelo fato de que quando analisamos os TFBSs preditos do CNE8 humano, obtivemos uma rede com alto valor de significância. De fato, dentre os 28 TFBSs preditos para este CNE, seis deles (STAT6, JUN, CREB1, TP53, NF1 e CEBPA) aparecem relacionados a um estudo sobre o desenvolvimento de lipossarcomas, que demonstrou que o bloqueio desses fatores por determinados oncogenes leva à progressão desse tipo de tumor (Kim et al., 2019) (Apêndice 7). Além disso, no CNE8 também há um enriquecimento de fatores envolvidos na via KEGG hsa05200, que trata das sinalizações ativas em câncer, sendo a via Wnt canônica uma delas (Apêndice 8). Em conjunto, estes achados sugerem que *Dact1* possa ser mais um componente da extensa cadeia de elementos que sofrem alterações de expressão também em sarcomas, um contexto, até o momento, pouco explorado em relação a este gene.

Com relação ao envolvimento de *Dact1* e seus elementos cis-reguladores em sarcomas, é digno de nota que há uma correlação significativa entre tumores estromais gastrointestinais com a deleção da região chr14:59,284,472–61,286,479, que contém os genes *Daam1*, *RTN1* e *Dact1* (Astolfi et al., 2010). É interessante ressaltar que esta região compreende também os CNEs 6, 7 e 8, identificados neste trabalho. Em conjunto, esses genes têm um papel fundamental no controle da apoptose e na via Wnt/ β -catenina. Particularmente, *Dact1* é modulado negativamente neste tipo de sarcoma e sua superexpressão é, portanto, capaz de promover um efeito supressor no tumor. Uma hipótese é a de que SNPs presentes nestes elementos cis-reguladores ou nos outros identificados nesta tese tenham correlação com o desenvolvimento de tipos específicos de câncer. Se, como observado por Astolfi e colaboradores (2010), a deleção da região que compreende esses CNEs, ou mesmo a ocorrência de mutações nestes elementos, for também presente em outros tipos de sarcoma, será de suma importância investigar a expressão de *Dact1* em um amplo conjunto destas doenças.

5.4.6 A função do CNE4 confirma o papel regulador deste elemento genômico

Dado que *Dact1* é expresso ao longo da miogênese esquelética de camundongo e outros vertebrados, um dos objetivos deste trabalho foi identificar um possível elemento regulador deste gene neste contexto específico. As múltiplas análises *in silico* realizadas indicaram fortemente que o CNE4 seria tal elemento regulador. Por este motivo, ele foi escolhido para ser testado funcionalmente e para as análises evolutivas.

A análise de dados de *Hi-C* gerados em amostras provenientes do músculo Psoas suportam a hipótese de que este segmento genômico seja capaz de interagir com a região que compreende o promotor de *Dact1*. Em adição, verificamos que o CNE4 corresponde à uma região hipersensível à DNase, um típico marcador de atividade de elementos reguladores. Por meio de ensaios de transfecção celular, demonstramos que o CNE4 foi capaz de aumentar a expressão do gene repórter *Luciferase* em células HEK293T e em mioblastos C2C12. Embora a expressão diferencial não tenha sido detectada em pré-adipócitos, não descartamos a possibilidade de que este elemento seja ativo em momentos mais tardios da adipogênese, uma vez que os ensaios foram realizados em células ainda em estado proliferativo. Por isso, será de grande importância avaliar a atividade deste elemento regulador por construções integrativas que possam ser acompanhadas por um período mais longo da diferenciação adipogênica, justamente porque *Dact1* atua neste contexto durante a maturação em adipócitos maduros (Lagathu et al., 2009). Em suma, pudemos concluir que a metodologia empregada nesta tese para identificar potenciais elementos reguladores demonstrou ser apropriada e eficiente, dado que nossas hipóteses sobre o papel de um CNE específico foram confirmadas.

Em adição, para aprofundarmos a caracterização do CNE4, realizamos também alinhamentos múltiplos de sequência entre o humano e 12 espécies distintas, incluindo organismos-modelo e algumas espécies de interesse econômico, provando que os segmentos genômicos ortólogos ao CNE4 estão presentes em todas as espécies analisadas e que todas estas estão abrigadas no locus de *Dact1*. No entanto, quando consideradas as treze espécies na totalidade, há um número bastante reduzido de TFBSs comuns a todas elas. Este fato sugere que, ao longo da evolução, variações de sequência tenham ocorrido, levando a uma divergência dos sítios envolvidos no controle da atividade transcricional deste CNE. Já a análise restrita ao grupo representante dos *Bovidae* mostrou que há um grande número de TFBSs conservados entre os animais deste grupo. O mesmo foi observado entre os primatas analisados,

sendo detectadas quase duas centenas de TFBSs conservados entre humano e chimpanzé. Portanto, estes achados indicam que existem características grupo-específicas para o CNE4, possivelmente visando adequar a função deste CNE às diferentes demandas transcricionais de *Dact1* nestas espécies. A rede de fatores de transcrição presentes no CNE4 dos primatas analisados se assemelha àquelas encontradas quando analisadas as redes do CNE4 conservado entre humano, camundongo e galinha, revelando que alguns contextos regulatórios, como o muscular, nutricional e tumoral, são centrais na atividade do CNE4. Como esperávamos, os *clusters* de fatores formados pelo *STRING* remetem claramente à miogênese, o que corrobora nossa hipótese de que a modulação de *Dact1* em doenças humanas do músculo esquelético, como as distrofias, por exemplo, possa ser devida à atuação deste elemento conservado ou mesmo deste em conjunto com elementos vizinhos. Como visto nos dados de expressão gênica de *Dact1* durante a miogênese em células C2C12 e também nas análises *in silico* dos *datasets* de mRNA proveniente de biópsias de músculos de pacientes acometidos por miopatias (dados apresentados no capítulo 1), *Dact1* é diferencialmente expresso durante a miogênese, seja ela fetal ou pós-natal (regenerativa). Além disso, será de grande valia apurar se, em determinadas patologias, SNPs presentes neste elemento regulador possam também interferir na expressão de *Dact1*, um campo de pesquisa que ainda necessita de maior investigação.

As análises de interações proteína-proteína para os TFBSs conservados entre humano e chimpanzé mostraram que há um enriquecimento destes fatores nas vias KEGG de diversas doenças como cânceres de próstata, de pâncreas e colorretal. Estes fatores também estão relacionados à carcinogênese viral, como ocorre, por exemplo, no sarcoma de Kaposi pelo Herpesvírus (KSHV) e também na leucemia causada pelo vírus linfotrópico de células T (HTLV-I). Há também um envolvimento destes TFs nas redes de sinalização mediadas por FOXO – fatores implicados tanto no músculo esquelético quando na adipogênese –, e também à diabetes, contextos onde *Dact1* notadamente atua. É importante ressaltar que há um *cluster* de fatores no CNE4 de primatas que são envolvidos no comprometimento das células neuronais e que não está presente, por exemplo, nos *Bovidae*. Curiosamente, *Dact1* é altamente expresso em alguns tipos de neurônios durante o desenvolvimento embrionário, incluindo no gânglio da raiz dorsal, células ganglionares da retina e nos neurônios sensoriais olfatórios (Suzuki et al., 2014). No entanto, como observado por Le Guen e colaboradores (2020), há uma importante diferença entre humanos e chimpanzés na formação das

estruturas cerebrais responsáveis pela linguagem e que está relacionada a uma região *upstream* à *Dact1* caracterizada como *enhancer* nos humanos. De acordo com este estudo, o SNP rs160459 (chr14:59,074,136) parece ter forte correlação com uma característica morfológica típica do sulco temporal em humanos que é ligada ao desenvolvimento da linguagem, de acordo com associações feitas entre genótipo e ressonâncias magnéticas em milhares de indivíduos.

A análise das redes de interação entre os TFBSs do CNE4 de bovinos revelou-se distinta da dos primatas. Um fato que se sobressai é a ausência de TFBSs para fatores miogênicos, como MYOD e MYOG, que estão conservados quando comparamos os genomas de humano, camundongo e galinha. Esta característica do CNE4 bovino sugere que houve uma perda de sítios miogênicos nestas espécies, o que pode ter acarretado a perda da função miogênica dependente destes fatores de transcrição neste grupo de animais. Por outro lado, uma característica do CNE4 bovino que se assemelha ao do CNE4 de primatas é o enriquecimento de fatores implicados nas vias KEGG relacionadas ao câncer. É digno de nota o enriquecimento de fatores na via KEGG bta05166, a qual diz respeito à infecção pelo retrovírus HTLV-1, e que também aparece entre os primatas analisados. A família dos vírus *Retroviridae* têm duas marcas distintivas de replicação: a transcrição reversa do RNA genômico em uma cópia linear de DNA de fita dupla e a subsequente integração covalente desse DNA no genoma do hospedeiro. Dentre os retrovírus, a família Oncovírus compreende diversos vírus com significado clínico, econômico e veterinário. Dentre eles estão o vírus linfotrópico de células T humanas (HTLV-1-4), o vírus da leucemia de células T de símio (STLV-1-5) e o vírus da leucemia bovina (BLV) (Hajj et al., 2012). Em bovinos, este vírus causa uma doença viral agressiva semelhante à leucemia e que é encontrada também em aves domésticas. Isto sugere que o CNE4 neste grupo pode servir como um potencial alvo para a regulação gênica nesta patologia, tendo, inclusive, valor econômico implicado. Além disso, dadas as semelhanças entre as doenças causadas em bovinos e humanos, como exemplo deste último temos a leucemia de células T e linfomas, as ovelhas são utilizadas como um importante modelo experimental para este tipo de câncer, sugerindo, mais uma vez, o potencial de estudo que pode ser aplicado nesta região regulatória, implicando em repercussões clínicas aplicáveis também aos humanos.

Em conjunto, estes resultados apontam para uma regulação diferencial da expressão de *Dact1* entre os grupos de vertebrados amniotos, como os primatas, os bovídeos e as aves, sugerindo haver

uma demanda de expressão deste gene diferente para cada grupo, regulando processos biológicos distintos.

5.5 CONCLUSÕES

1. Existem oito CNEs localizadas no intervalo genômico *KIAA0586-Dact1-Daam1*, que representam potenciais elementos cis-reguladores de *Dact1*.
2. Há seis CNEs que se localizam em segmentos genômicos catalogados como elementos reguladores no *Ensembl Regulatory Build*.
3. Os CNEs que apresentam sobreposição aos elementos reguladores catalogados no *Ensembl Regulatory Build* são ativos em vários contextos celulares nos quais *Dact1* é conhecidamente expresso ou nos quais a modulação da sua expressão está associada com o desenvolvimento de doenças humanas.
4. Os CNEs 1, 2, 3, 4, 5, 7 e 8 apresentam sítios para fatores de transcrição que são conservados evolutivamente.
5. Os CNEs 5, 6, 7 e 8, que não apresentam número de TFBSs suficientemente elevado para possibilitar o estabelecimento de redes de interação, provavelmente divergiram entre humano, camundongo e galinha, mas correspondem a elementos reguladores do humano.
6. Os fatores de transcrição que potencialmente se ligam aos sítios dos CNEs 1, 2, 4 formam redes de interação proteína-proteína com alto valor de significância, demonstrando que estas interações não ocorrem ao acaso.
7. A análise de enriquecimento funcional das proteínas que se ligam aos sítios de ligação do CNE1 identificou que os fatores de transcrição presentes neste segmento estão associados à via KEGG hsa05202 (*Transcriptional misregulation in cancer*), o que é compatível com o envolvimento de *Dact1* em vários tipos de câncer.
8. Os fatores de transcrição que se ligam aos sítios de ligação do CNE2 estão associados ao estabelecimento da linhagem hematopoiética e a doenças das células linfocitárias, corroborando o envolvimento de *Dact1* no desenvolvimento de leucemia mieloide aguda.
9. Os fatores de transcrição que se ligam aos sítios de ligação do CNE4 estão associados à miogênese esquelética e doenças do músculo esquelético, bem como à adipogênese.

10. O CNE4, empregado para validação funcional da metodologia de identificação de elementos cis-reguladores, atua como um *enhancer* por meio de ensaio de *Luciferase* nas células C2C12 em estado proliferativo.
11. Sequências ortólogas ao CNE4 estão presentes no genoma de 12 espécies de vertebrados, em adição ao humano.
12. A análise de conservação do CNE4 entre espécies próximas: primatas (humano e chimpanzé) e bovídeos (vaca, ovelha e cabra) revela elevada conservação intra-clado. Por outro lado, revela um repertório distinto de sítios de ligação para fatores de transcrição quando os dois clados são comparados, indicando que este elemento cis-regulador tem características grupo-específicas.

6. CONCLUSÕES GERAIS E PERSPECTIVAS SOBRE AS APLICAÇÕES DO ESTUDO

Nesta tese foi demonstrado que *Dact1* participa da miogênese esquelética de mamíferos e que este é modulado em diferentes patologias do músculo esquelético em humanos. A descoberta de que *Dact1/Dact1* é amplamente expresso durante toda miogênese em cultura de células C2C12 demonstra que este gene possivelmente atua como um modulador bivalente da via Wnt/ β -catenina, devendo ser responsável pelo *switch* desta sinalização durante o processo de diferenciação destas células. Este achado abre novas possibilidade de estudo sobre a modulação das vias Wnt, por meio de *Dact1*, que estão amplamente relacionadas ao desenvolvimento e reparo do músculo esquelético. Sabendo que *Dact1* tem sua expressão alterada em miopatias humanas, este gene se coloca como um potencial alvo terapêutico nestas doenças, principalmente devido ao fato de que *Dact1* tem uma região semelhante à distrofina e utrofina, proteínas essenciais para a manutenção do músculo esquelético.

Nesta tese também identificamos diversos possíveis elementos cis-reguladores de *Dact1*, incluindo um possível *enhancer* de *Dact1* no contexto miogênico (CNE4), mas também em muitos outros contextos, principalmente relacionados a doenças humanas, como os cânceres. A presença de TFBSs para fatores miogênicos no CNE4 indica que este elemento regulador pode atuar tanto durante a miogênese esquelética, como durante o reparo muscular mediado por células-satélites. Por isso, ensaios de *ChIP-Seq* são necessários para comprovarmos a interação destes segmentos genômicos com os fatores de transcrição preditos por bioinformática em nossas análises. Em adição, a pesquisa por SNPs neste elemento cis-regulador também será de grande valia para avaliarmos se a expressão desregulada de *Dact1*, bem como das vias Wnt que ele modula, se devem a mutações neste elemento, o que contribuiria para o desenvolvimento e progressão de doenças do músculo esquelético.

Os elementos cis-reguladores putativos de *Dact1* identificados nesta tese abrem também novas perspectivas de estudo sobre a modulação deste gene em contextos até então pouco explorados, como as leucemias. Já em outros tipos de câncer, nos quais *Dact1* tem sua expressão conhecidamente alterada, como nos tumores gastrointestinais, de cólon, de fígado, de mama e etc., será de suma importância investigar a ocorrência de SNPs e/ou deleções nestas regiões que compreendem estes elementos identificados. Finalmente, o aprofundamento sobre a conservação dos elementos cis-

reguladores identificados entre os diferentes grupos de vertebrados será importante para compreendermos a regulação da expressão de *Dact1* de maneira grupo-específica e, com isso, poder determinar os processos biológicos distintos que este gene regula. Os achados desta tese também contribuem para a investigação sobre possíveis mutações nestas regiões que podem repercutir em alterações fisiológicas e/ou morfológicas entre diferentes grupos de animais.

7. REFERÊNCIAS BIBLIOGRÁFICAS

- Adams, B., Dörfler, P., Aguzzi, A., Kozmik, Z., Urbánek, P., Maurer-Fogy, I., Busslinger, M., 1992. Pax-5 encodes the transcription factor BSAP and is expressed in B lymphocytes, the developing CNS, and adult testis. *Genes Dev* 6, 1589-1607.
- Altman, B.J., Hsieh, A.L., Sengupta, A., Krishnanaiah, S.Y., Stine, Z.E., Walton, Z.E., Gouw, A.M., Venkataraman, A., Li, B., Goraksha-Hicks, P., Diskin, S.J., Bellovin, D.I., Simon, M.C., Rathmell, J.C., Lazar, M.A., Maris, J.M., Felsher, D.W., Hogenesch, J.B., Weljie, A.M., Dang, C.V., 2015. MYC Disrupts the Circadian Clock and Metabolism in Cancer Cells. *Cell Metab* 22, 1009-1019.
- Alvares, L.E., Winterbottom, F.L., Rodrigues Sobreira, D., Xavier-Neto, J., Schubert, F.R., Dietrich, S., 2009. Chicken dapper genes are versatile markers for mesodermal tissues, embryonic muscle stem cells, neural crest cells, and neurogenic placodes. *Developmental Dynamics* 238, 1166-1178.
- Ambele, M.A., Pepper, M.S., 2017. Identification of transcription factors potentially involved in human adipogenesis in vitro. *Mol Genet Genomic Med* 5, 210-222.
- Andres, V., Nadal-Ginard, B., Mahdavi, V., 1992. Clox, a mammalian homeobox gene related to *Drosophila cut*, encodes DNA-binding regulatory proteins differentially expressed during development. *Development* 116, 321-334.
- Aparicio, S., Morrison, A., Gould, A., Gilthorpe, J., Chaudhuri, C., Rigby, P., Krumlauf, R., Brenner, S., 1995. Detecting conserved regulatory elements with the model genome of the Japanese puffer fish, *Fugu rubripes*. *Proceedings of the National Academy of Sciences* 92, 1684-1688.
- Asfour, H.A., Allouh, M.Z., Said, R.S., 2018. Myogenic regulatory factors: The orchestrators of myogenesis after 30 years of discovery. *Experimental Biology and Medicine* 243, 118-128.
- Aspland, S.E., Bendall, H.H., Murre, C., 2001. The role of E2A-PBX1 in leukemogenesis. *Oncogene* 20, 5708-5717.
- Aspnes, L.E., Lee, C.M., Weindruch, R., Chung, S.S., Roecker, E.B., Aiken, J.M., 1997. Caloric restriction reduces fiber loss and mitochondrial abnormalities in aged rat muscle. *FASEB J* 11, 573-581.
- Astolfi, A., Nannini, M., Pantaleo, M.A., Di Battista, M., Heinrich, M.C., Santini, D., Catena, F., Corless, C.L., Maleddu, A., Saponara, M., Lolli, C., Di Scioscio, V., Formica, S., Biasco, G., 2010b. A molecular portrait of gastrointestinal stromal tumors: an integrative analysis of gene expression profiling and high-resolution genomic copy number. *Lab Invest* 90, 1285-1294.
- Bachner-Melman, R., Dina, C., Zohar, A.H., Constantini, N., Lerer, E., Hoch, S., Sella, S., Nemanov, L., Gritsenko, I., Lichtenberg, P., 2005. AVPR1a and SLC6A4 gene polymorphisms are associated with creative dance performance. *PLoS genetics* 1.
- Bain, G., Maandag, E.C., Izon, D.J., Amsen, D., Kruisbeek, A.M., Weintraub, B.C., Krop, I., Schlissel, M.S., Feeney, A.J., van Roon, M., 1994. E2A proteins are required for proper B cell development and initiation of immunoglobulin gene rearrangements. *Cell* 79, 885-892.

Bejerano, G., Pheasant, M., Makunin, I., Stephen, S., Kent, W.J., Mattick, J.S., Haussler, D., 2004. Ultraconserved Elements in the Human Genome. *Science* 304, 1321-1325.

Bell, D.M., Leung, K.K., Wheatley, S.C., Ng, L.J., Zhou, S., Ling, K.W., Sham, M.H., Koopman, P., Tam, P.P., Cheah, K.S., 1997. SOX9 directly regulates the type-II collagen gene. *Nat Genet* 16, 174-178.

Blau, H.M., Chiu, C.-P., Webster, C., 1983. Cytoplasmic activation of human nuclear genes in stable heterocaryons. *Cell* 32, 1171-1180.

Boeynaems, S., Alberti, S., Fawzi, N.L., Mittag, T., Polymenidou, M., Rousseau, F., Schymkowitz, J., Shorter, J., Wolozin, B., Van Den Bosch, L., Tompa, P., Fuxreiter, M., 2018. Protein Phase Separation: A New Phase in Cell Biology. *Trends Cell Biol* 28, 420-435.

Boström, P., Mann, N., Wu, J., Quintero, P.A., Plovie, E.R., Panáková, D., Gupta, R.K., Xiao, C., MacRae, C.A., Rosenzweig, A., Spiegelman, B.M., 2010. C/EBP β controls exercise-induced cardiac growth and protects against pathological cardiac remodeling. *Cell* 143, 1072-1083.

Brangwynne, C.P., Eckmann, C.R., Courson, D.S., Rybarska, A., Hoege, C., Gharakhani, J., Jülicher, F., Hyman, A.A., 2009. Germline P granules are liquid droplets that localize by controlled dissolution/condensation. *Science* 324, 1729-1732.

Brott, B.K., Sokol, S.Y., 2005a. A vertebrate homolog of the cell cycle regulator Dbf4 is an inhibitor of Wnt signaling required for heart development. *Dev Cell* 8, 703-715.

Brott, B.K., Sokol, S.Y., 2005b. Frodo proteins: modulators of Wnt signaling in vertebrate development. *Differentiation* 73, 323-329.

Bulger, M., Groudine, M., 2010. Enhancers: the abundance and function of regulatory sequences beyond promoters. *Developmental biology* 339, 250-257.

Carroll, S.B., 1995. Homeotic genes and the evolution of arthropods and chordates. *Nature* 376, 479-485.

Chal, J., Pourquié, O., 2017. Making muscle: skeletal myogenesis in vivo and in vitro. *Development* 144, 2104-2122.

Cheyette, B.N., Waxman, J.S., Miller, J.R., Takemaru, K., Sheldahl, L.C., Khlebtsova, N., Fox, E.P., Earnest, T., Moon, R.T., 2002. Dapper, a Dishevelled-associated antagonist of beta-catenin and JNK signaling, is required for notochord formation. *Dev Cell* 2, 449-461.

Chin, E.R., Olson, E.N., Richardson, J.A., Yang, Q., Humphries, C., Shelton, J.M., Wu, H., Zhu, W., Bassel-Duby, R., Williams, R.S., 1998. A calcineurin-dependent transcriptional pathway controls skeletal muscle fiber type. *Genes Dev* 12, 2499-2509.

Christodoulides, C., Lagathu, C., Sethi, J.K., Vidal-Puig, A., 2009. Adipogenesis and WNT signalling. *Trends in Endocrinology & Metabolism* 20, 16-24.

Chung, S.Y., Huang, W.C., Su, C.W., Lee, K.W., Chi, H.C., Lin, C.T., Chen, S.T., Huang, K.M., Tsai, M.S., Yu, H.P., Chen, S.L., 2013. FoxO6 and PGC-1 α form a regulatory loop in myogenic cells. *Biosci Rep* 33.

Church, G.M., Ephrussi, A., Gilbert, W., Tonegawa, S., 1985. Cell-type-specific contacts to immunoglobulin enhancers in nuclei. *Nature* 313, 798-801.

Clarke, J.D., Novak, P., Lake, A.D., Shipkova, P., Aranibar, N., Robertson, D., Severson, P.L., Reily, M.D., Futscher, B.W., Lehman-McKeeman, L.D., Cherrington, N.J., 2014. Characterization of hepatocellular carcinoma related genes and metabolites in human nonalcoholic fatty liver disease. *Dig Dis Sci* 59, 365-374.

Collignon, J., Sockanathan, S., Hacker, A., Cohen-Tannoudji, M., Norris, D., Rastan, S., Stevanovic, M., Goodfellow, P.N., Lovell-Badge, R., 1996. A comparison of the properties of Sox-3 with Sry and two related genes, Sox-1 and Sox-2. *Development* 122, 509-520.

Colosimo, P.F., Hosemann, K.E., Balabhadra, S., Villarreal, G., Dickson, M., Grimwood, J., Schmutz, J., Myers, R.M., Schluter, D., Kingsley, D.M., 2005. Widespread parallel evolution in sticklebacks by repeated fixation of ectodysplasin alleles. *science* 307, 1928-1933.

Consortium, E.P., 2012. An integrated encyclopedia of DNA elements in the human genome. *Nature* 489, 57-74.

Consortium, G., 2013. The Genotype-Tissue Expression (GTEx) project. *Nat Genet* 45, 580-585.

Cox, J.J., Willatt, L., Homfray, T., Woods, C.G., 2011. A SOX9 duplication and familial 46,XX developmental testicular disorder. *N Engl J Med* 364, 91-93.

Daborn, P., Yen, J., Bogwitz, M., Le Goff, G., Feil, E., Jeffers, S., Tijet, N., Perry, T., Heckel, D., Batterham, P., 2002. A single p450 allele associated with insecticide resistance in *Drosophila*. *Science* 297, 2253-2256.

Davis, T.A., Fiorotto, M.L., 2009. Regulation of muscle growth in neonates. *Current opinion in clinical nutrition and metabolic care* 12, 78.

de Wit, E., de Laat, W., 2012. A decade of 3C technologies: insights into nuclear organization. *Genes Dev* 26, 11-24.

Deconinck, A.E., Rafael, J.A., Skinner, J.A., Brown, S.C., Potter, A.C., Metzinger, L., Watt, D.J., Dickson, J.G., Tinsley, J.M., Davies, K.E., 1997. Utrophin-dystrophin-deficient mice as a model for Duchenne muscular dystrophy. *Cell* 90, 717-727.

Dekker, J., Marti-Renom, M.A., Mirny, L.A., 2013. Exploring the three-dimensional organization of genomes: interpreting chromatin interaction data. *Nat Rev Genet* 14, 390-403.

Di Giovanni, S., Molon, A., Broccolini, A., Melcon, G., Mirabella, M., Hoffman, E.P., Servidei, S., 2004. Constitutive activation of MAPK cascade in acute quadriplegic myopathy. *Ann Neurol* 55, 195-206.

Dressler, G.R., 2009. Advances in early kidney specification, development and patterning. *Development* 136, 3863-3874.

Dräger, J., Simon-Keller, K., Pukrop, T., Klemm, F., Wilting, J., Sticht, C., Dittmann, K., Schulz, M., Leuschner, I., Marx, A., 2017. LEF1 reduces tumor progression and induces myodifferentiation in a subset of rhabdomyosarcoma. *Oncotarget* 8, 3259.

Eastman, Q., Grosschedl, R., 1999. Regulation of LEF-1/TCF transcription factors by Wnt and other signals. *Curr Opin Cell Biol* 11, 233-240.

Enerbäck, S., Nilsson, D., Edwards, N., Heglind, M., Alkanderi, S., Ashton, E., Deeb, A., Kokash, F.E.B., Bakhsh, A.R.A., Van't Hoff, W., Walsh, S.B., D'Arco, F., Daryadel, A., Bourgeois, S., Wagner, C.A., Kleta, R.,

Bockenbauer, D., Sayer, J.A., 2018. Acidosis and Deafness in Patients with Recessive Mutations in FOXI1. *J Am Soc Nephrol* 29, 1041-1048.

Eric C, H.L., 2008. Hepatocellular carcinoma: current management and recent advances. *国际肝胆胰疾病杂志: 英文版*, 237-257.

Farmer, A. (2016). The Roles of Nuclear Receptor NR4A1 in Cancer Cell Proliferation and Skeletal Muscle Differentiation. Theses and Dissertations (ETD). Paper 403. doi: 10.21007/etd.cghs.2016.0411

Fisher, D.A., Kivimäe, S., Hoshino, J., Suriben, R., Martin, P.M., Baxter, N., Cheyette, B.N., 2006. Three Dact gene family members are expressed during embryonic development and in the adult brains of mice. *Developmental dynamics* 235, 2620-2630.

Fishilevich, S., Nudel, R., Rappaport, N., Hadar, R., Plaschkes, I., Iny Stein, T., Rosen, N., Kohn, A., Twik, M., Safran, M., Lancet, D., Cohen, D., 2017. GeneHancer: genome-wide integration of enhancers and target genes in GeneCards. Database (Oxford) 2017.

Frazer, K.A., Pachter, L., Poliakov, A., Rubin, E.M., Dubchak, I., 2004. VISTA: computational tools for comparative genomics. *Nucleic Acids Res* 32, W273-279.

Furukawa, T., Morrow, E.M., Cepko, C.L., 1997. Crx, a novel otx-like homeobox gene, shows photoreceptor-specific expression and regulates photoreceptor differentiation. *Cell* 91, 531-541.

Gachon, F., Fonjallaz, P., Damiola, F., Gos, P., Kodama, T., Zakany, J., Duboule, D., Petit, B., Tafti, M., Schibler, U., 2004. The loss of circadian PAR bZip transcription factors results in epilepsy. *Genes Dev* 18, 1397-1412.

Gao, X., Wen, J., Zhang, L., Li, X., Ning, Y., Meng, A., Chen, Y.G., 2008. Dapper1 is a nucleocytoplasmic shuttling protein that negatively modulates Wnt signaling in the nucleus. *J Biol Chem* 283, 35679-35688.

Georgopoulos, K., Bigby, M., Wang, J.-H., Molnar, A., Wu, P., Winandy, S., Sharpe, A., 1994. The Ikaros gene is required for the development of all lymphoid lineages. *Cell* 79, 143-156.

Gerhart-Hines, Z., Rodgers, J.T., Bare, O., Lerin, C., Kim, S.H., Mostoslavsky, R., Alt, F.W., Wu, Z., Puigserver, P., 2007. Metabolic control of muscle mitochondrial function and fatty acid oxidation through SIRT1/PGC-1alpha. *EMBO J* 26, 1913-1923.

Gerstein, M.B., Lu, Z.J., Van Nostrand, E.L., Cheng, C., Arshinoff, B.I., Liu, T., Yip, K.Y., Robilotto, R., Rechtsteiner, A., Ikegami, K., 2010. Integrative analysis of the *Caenorhabditis elegans* genome by the modENCODE project. *Science* 330, 1775-1787.

Gilbert, S., 2010. *Developmental biology* 9th Edition Sinauer Associates Inc.

Gillhouse, M., Wagner Nyholm, M., Hikasa, H., Sokol, S.Y., Grinblat, Y., 2004. Two Frodo/Dapper homologs are expressed in the developing brain and mesoderm of zebrafish. *Developmental dynamics: an official publication of the American Association of Anatomists* 230, 403-409.

Gillis, J.-M., 1996. Membrane abnormalities and Ca homeostasis in muscles of the mdx mouse, an animal model of the Duchenne muscular dystrophy: a review. *Acta physiologica Scandinavica* 156, 397-406.

Gloy, J., Hikasa, H., Sokol, S.Y., 2002. Frodo interacts with Dishevelled to transduce Wnt signals. *Nature cell biology* 4, 351.

Goldstein, J.A., Bogdanovich, S., Beiriger, A., Wren, L.M., Rossi, A.E., Gao, Q.Q., Gardner, B.B., Earley, J.U., Molkentin, J.D., McNally, E.M., 2014. Excess SMAD signaling contributes to heart and muscle dysfunction in muscular dystrophy. *Hum Mol Genet* 23, 6722-6731.

Gompel, N., Prud'homme, B., Wittkopp, P.J., Kassner, V.A., Carroll, S.B., 2005. Chance caught on the wing: cis-regulatory evolution and the origin of pigment patterns in *Drosophila*. *Nature* 433, 481-487.

González-Sánchez, J., Sánchez-Temprano, A., Cid-Díaz, T., Pabst-Fernández, R., Mosteiro, C.S., Gallego, R., Nogueiras, R., Casabiell, X., Butler-Browne, G.S., Mouly, V., Relova, J.L., Pazos, Y., Camiña, J.P., 2018. Improvement of Duchenne muscular dystrophy phenotype following obestatin treatment. *J Cachexia Sarcopenia Muscle* 9, 1063-1078.

Goode, D.K., Callaway, H.A., Cerda, G.A., Lewis, K.E., Elgar, G., 2011. Minor change, major difference: divergent functions of highly conserved cis-regulatory elements subsequent to whole genome duplication events. *Development* 138, 879-884.

Greenbaum, L.E., Li, W., Cressman, D.E., Peng, Y., Ciliberto, G., Poli, V., Taub, R., 1998. CCAAT enhancer-binding protein beta is required for normal hepatocyte proliferation in mice after partial hepatectomy. *J Clin Invest* 102, 996-1007.

Gregor, P.D., Sawadogo, M., Roeder, R.G., 1990. The adenovirus major late transcription factor USF is a member of the helix-loop-helix group of regulatory proteins and binds to DNA as a dimer. *Genes Dev* 4, 1730-1740.

Griswold, M.D., Kim, J.S., 2001. Site-specific methylation of the promoter alters deoxyribonucleic acid-protein interactions and prevents follicle-stimulating hormone receptor gene transcription. *Biol Reprod* 64, 602-610.

Gros, J., Manceau, M., Thomé, V., Marcelle, C., 2005. A common somitic origin for embryonic muscle progenitors and satellite cells. *Nature* 435, 954-958.

Gros, J., Serralbo, O., Marcelle, C., 2009. WNT11 acts as a directional cue to organize the elongation of early muscle fibres. *Nature* 457, 589-593.

Guo, H., Barberi, T., Suresh, R., Friedman, A.D., 2018. Progression from the Common Lymphoid Progenitor to B/Myeloid PreproB and ProB Precursors during B Lymphopoiesis Requires C/EBP α . *J Immunol* 201, 1692-1704.

Guo, Y.L., Shan, B.E., Guo, W., Dong, Z.M., Zhou, Z., Shen, S.P., Guo, X., Liang, J., Kuang, G., 2017. Aberrant methylation of DACT1 and DACT2 are associated with tumor progression and poor prognosis in esophageal squamous cell carcinoma. *J Biomed Sci* 24, 6.

Gutiérrez, N.C., Ocio, E.M., de Las Rivas, J., Maiso, P., Delgado, M., Ferminán, E., Arcos, M.J., Sánchez, M.L., Hernández, J.M., San Miguel, J.F., 2007. Gene expression profiling of B lymphocytes and plasma cells from Waldenström's macroglobulinemia: comparison with expression patterns of the same cell counterparts from chronic lymphocytic leukemia, multiple myeloma and normal individuals. *Leukemia* 21, 541-549.

Gómez-Ferrería, M.A., Rey-Campos, J., 2003. Functional domains of FOXJ2. *J Mol Biol* 329, 631-644.

Hajj, H.E., Nasr, R., Kfoury, Y., Dassouki, Z., Nasser, R., Kchour, G., Hermine, O., de Thé, H., Bazarbachi, A., 2012. Animal models on HTLV-1 and related viruses: what did we learn? *Front Microbiol* 3, 333.

Hamblin, M.T., Di Rienzo, A., 2000. Detection of the signature of natural selection in humans: evidence from the Duffy blood group locus. *Am J Hum Genet* 66, 1669-1679.

Handeli, S., Simon, J.A., 2008. A small-molecule inhibitor of Tcf/ β -catenin signaling down-regulates PPAR γ and PPAR δ activities. *Molecular cancer therapeutics* 7, 521-529.

Harnoš, J., Ryneš, J., Víšková, P., Foldynová-Trantírková, S., Bajard-Ešner, L., Trantírek, L., Bryja, V., 2018. Analysis of binding interfaces of the human scaffold protein AXIN1 by peptide microarrays. *J Biol Chem* 293, 16337-16347.

He, N., Park, K., Zhang, Y., Huang, J., Lu, S., Wang, L., 2008. Epigenetic inhibition of nuclear receptor small heterodimer partner is associated with and regulates hepatocellular carcinoma growth. *Gastroenterology* 134, 793-802.

Helliwell, T., Wilkinson, A., Griffiths, R., McClelland, P., Palmer, T., Bone, J., 1998. Muscle fibre atrophy in critically ill patients is associated with the loss of myosin filaments and the presence of lysosomal enzymes and ubiquitin. *Neuropathology and applied neurobiology* 24, 507-517.

Hemberger, M., Himmelbauer, H., Neumann, H.P., Plate, K.H., Schwarzkopf, G., Fundele, R., 1999. Expression of the von Hippel-Lindau-binding protein-1 (Vbp1) in fetal and adult mouse tissues. *Hum Mol Genet* 8, 229-236.

Herrero, J., Muffato, M., Beal, K., Fitzgerald, S., Gordon, L., Pignatelli, M., Vilella, A.J., Searle, S.M.J., Amode, R., Brent, S., Spooner, W., Kulesha, E., Yates, A., Flicek, P., 2016. Ensembl comparative genomics resources. *Database : the journal of biological databases and curation* 2016, bav096.

Hikasa, H., Sokol, S.Y., 2004. The involvement of Frodo in TCF-dependent signaling and neural tissue development. *Development* 131, 4725-4734.

Hou, J., Yue, Y., Hu, B., Xu, G., Su, R., Lv, L., Huang, J., Yao, J., Guan, Y., Wang, K., Wu, Z., 2019. DACT1 Involvement in the Cytoskeletal Arrangement of Cardiomyocytes in Atrial Fibrillation by Regulating Cx43. *Braz J Cardiovasc Surg* 34, 711-722.

Huang, Y., de Morrée, A., van Remoortere, A., Bushby, K., Frants, R.R., den Dunnen, J.T., van der Maarel, S.M., 2008. Calpain 3 is a modulator of the dysferlin protein complex in skeletal muscle. *Hum Mol Genet* 17, 1855-1866.

Hunger, S.P., Ohyashiki, K., Toyama, K., Cleary, M.L., 1992. Hlf, a novel hepatic bZIP protein, shows altered DNA-binding properties following fusion to E2A in t(17;19) acute lymphoblastic leukemia. *Genes Dev* 6, 1608-1620.

Hunter, N.L., Hikasa, H., Dymecki, S.M., Sokol, S.Y., 2006. Vertebrate homologues of Frodo are dynamically expressed during embryonic development in tissues undergoing extensive morphogenetic movements. *Developmental dynamics* 235, 279-284.

Hutcheson, D.A., Zhao, J., Merrell, A., Haldar, M., Kardon, G., 2009. Embryonic and fetal limb myogenic cells are derived from developmentally distinct progenitors and have different requirements for β -catenin. *Genes & development* 23, 997-1013.

Jenuwein, T., Allis, C.D., 2001. Translating the histone code. *Science* 293, 1074-1080.

Jeong, S., Rokas, A., Carroll, S.B., 2006. Regulation of body pigmentation by the Abdominal-B Hox protein and its gain and loss in *Drosophila* evolution. *Cell* 125, 1387-1399.

Jiang, X., Tan, J., Li, J., Kivimäe, S., Yang, X., Zhuang, L., Lee, P.L., Chan, M.T., Stanton, L.W., Liu, E.T., 2008. DACT3 is an epigenetic regulator of Wnt/ β -catenin signaling in colorectal cancer and is a therapeutic target of histone modifications. *Cancer cell* 13, 529-541.

Kaduthanam, S., Gade, S., Meister, M., Brase, J.C., Johannes, M., Dienemann, H., Warth, A., Schnabel, P.A., Herth, F.J., Sültmann, H., 2013. Serum miR-142-3p is associated with early relapse in operable lung adenocarcinoma patients. *Lung Cancer* 80, 223-227.

Kalinichenko, V.V., Lim, L., Shin, B., Costa, R.H., 2001. Differential expression of forkhead box transcription factors following butylated hydroxytoluene lung injury. *Am J Physiol Lung Cell Mol Physiol* 280, L695-704.

Kamizono, S., Duncan, G.S., Seidel, M.G., Morimoto, A., Hamada, K., Grosveld, G., Akashi, K., Lind, E.F., Haight, J.P., Ohashi, P.S., Look, A.T., Mak, T.W., 2009. Nfil3/E4bp4 is required for the development and maturation of NK cells in vivo. *J Exp Med* 206, 2977-2986.

Kanehisa, M., Furumichi, M., Tanabe, M., Sato, Y., Morishima, K., 2017. KEGG: new perspectives on genomes, pathways, diseases and drugs. *Nucleic Acids Res* 45, D353-D361.

Karolchik, D., Barber, G.P., Casper, J., Clawson, H., Cline, M.S., Diekhans, M., Dreszer, T.R., Fujita, P.A., Guruvadoo, L., Haeussler, M., Harte, R.A., Heitner, S., Hinrichs, A.S., Learned, K., Lee, B.T., Li, C.H., Raney,

B.J., Rhead, B., Rosenbloom, K.R., Sloan, C.A., Speir, M.L., Zweig, A.S., Haussler, D., Kuhn, R.M., Kent, W.J., 2014. The UCSC Genome Browser database: 2014 update. *Nucleic Acids Res* 42, D764-770.

Karweina, D., Kreuzer-Redmer, S., Müller, U., Franken, T., Pieper, R., Baron, U., Olek, S., Zentek, J., Brockmann, G.A., 2015. The Zinc Concentration in the Diet and the Length of the Feeding Period Affect the Methylation Status of the ZIP4 Zinc Transporter Gene in Piglets. *PLoS One* 10, e0143098.

Katoh, M., Katoh, M., 2005. Identification and characterization of rat Dact1 and Dact2 genes in silico. *International journal of molecular medicine* 15, 1045-1049.

Kayserili, H., Uz, E., Niessen, C., Vargel, I., Alanay, Y., Tuncbilek, G., Yigit, G., Uyguner, O., Candan, S., Okur, H., Kaygin, S., Balci, S., Mavili, E., Alikasifoglu, M., Haase, I., Wollnik, B., Akarsu, N.A., 2009. ALX4 dysfunction disrupts craniofacial and epidermal development. *Hum Mol Genet* 18, 4357-4366.

Kim, S.Y., Pritchard, J.K., 2007. Adaptive evolution of conserved noncoding elements in mammals. *PLoS genetics* 3, e147.

Kim, Y.J., Kim, M., Park, H.K., Yu, D.B., Jung, K., Song, K., Choi, Y.L., 2019. Co-expression of MDM2 and CDK4 in transformed human mesenchymal stem cells causes high-grade sarcoma with a dedifferentiated liposarcoma-like morphology. *Lab Invest* 99, 1309-1320.

King, D.P., Zhao, Y., Sangoram, A.M., Wilsbacher, L.D., Tanaka, M., Antoch, M.P., Steeves, T.D., Vitaterna, M.H., Kornhauser, J.M., Lowrey, P.L., Turek, F.W., Takahashi, J.S., 1997. Positional cloning of the mouse circadian clock gene. *Cell* 89, 641-653.

Kivimäe, S., Yang, X.Y., Cheyette, B.N., 2011. All Dact (Dapper/Frodo) scaffold proteins dimerize and exhibit conserved interactions with Vangl, Dvl, and serine/threonine kinases. *BMC Biochemistry* 12, 33.

Kondo, M., Weissman, I.L., Akashi, K., 1997. Identification of Clonogenic Common Lymphoid Progenitors in Mouse Bone Marrow. *Cell* 91, 661-672.

Kumar, S., Hedges, S.B., 1998. A molecular timescale for vertebrate evolution. *Nature* 392, 917.

Kuroda, K., Kuang, S., Taketo, M.M., Rudnicki, M.A., 2013. Canonical Wnt signaling induces BMP-4 to specify slow myofibrogenesis of fetal myoblasts. *Skeletal muscle* 3, 1-13.

Kuss, P., Kraft, K., Stumm, J., Ibrahim, D., Vallecillo-Garcia, P., Mundlos, S., Stricker, S., 2014. Regulation of cell polarity in the cartilage growth plate and perichondrium of metacarpal elements by HOXD13 and WNT5A. *Dev Biol* 385, 83-93.

Ladanyi, M., Lui, M.Y., Antonescu, C.R., Krause-Boehm, A., Meindl, A., Argani, P., Healey, J.H., Ueda, T., Yoshikawa, H., Meloni-Ehrig, A., Sorensen, P.H., Mertens, F., Mandahl, N., van den Berghe, H., Sciot, R., Dal Cin, P., Bridge, J., 2001. The der(17)t(X;17)(p11;q25) of human alveolar soft part sarcoma fuses the TFE3 transcription factor gene to ASPL, a novel gene at 17q25. *Oncogene* 20, 48-57.

Lagathu, C., Christodoulides, C., Virtue, S., Cawthorn, W.P., Franzin, C., Kimber, W.A., Nora, E.D., Campbell, M., Medina-Gomez, G., Cheyette, B.N., Vidal-Puig, A.J., Sethi, J.K., 2009. Dact1, a nutritionally regulated preadipocyte gene, controls adipogenesis by coordinating the Wnt/beta-catenin signaling network. *Diabetes* 58, 609-619.

Lander, E.S., Linton, L.M., Birren, B., Nusbaum, C., Zody, M.C., Baldwin, J., Devon, K., Dewar, K., Doyle, M., FitzHugh, W., Funke, R., Gage, D., Harris, K., Heaford, A., Howland, J., Kann, L., Lehoczky, J., LeVine, R., McEwan, P., McKernan, K., Meldrim, J., Mesirov, J.P., Miranda, C., Morris, W., Naylor, J., Raymond, C., Rosetti, M., Santos, R., Sheridan, A., Sougnez, C., Stange-Thomann, Y., Stojanovic, N., Subramanian, A., Wyman, D., Rogers, J., Sulston, J., Ainscough, R., Beck, S., Bentley, D., Burton, J., Clee, C., Carter, N., Coulson, A., Deadman, R., Deloukas, P., Dunham, A., Dunham, I., Durbin, R., French, L., Grafham, D., Gregory, S., Hubbard, T., Humphray, S., Hunt, A., Jones, M., Lloyd, C., McMurray, A., Matthews, L., Mercer, S., Milne, S., Mullikin, J.C., Mungall, A., Plumb, R., Ross, M., Shownkeen, R., Sims, S., Waterston, R.H., Wilson, R.K., Hillier, L.W., McPherson, J.D., Marra, M.A., Mardis, E.R., Fulton, L.A., Chinwalla, A.T., Pepin, K.H., Gish, W.R., Chissole, S.L., Wendl, M.C., Delehaunty, K.D., Miner, T.L., Delehaunty, A., Kramer, J.B., Cook, L.L., Fulton, R.S., Johnson, D.L., Minx, P.J., Clifton, S.W., Hawkins, T., Branscomb, E., Predki, P., Richardson, P., Wenning, S., Slezak, T., Doggett, N., Cheng, J.F., Olsen, A., Lucas, S., Elkin, C., Uberbacher, E., Frazier, M., Gibbs, R.A., Muzny, D.M., Scherer, S.E., Bouck, J.B., Sodergren, E.J., Worley, K.C., Rives, C.M., Gorrell, J.H., Metzker, M.L., Naylor, S.L., Kucherlapati, R.S., Nelson, D.L., Weinstock, G.M., Sakaki, Y., Fujiyama, A., Hattori, M., Yada, T., Toyoda, A., Itoh, T., Kawagoe, C., Watanabe, H., Totoki, Y., Taylor, T., Weissenbach, J., Heilig, R., Saurin, W., Artiguenave, F., Brottier, P., Bruls, T., Pelletier, E., Robert, C., Wincker, P., Smith, D.R., Doucette-Stamm, L., Rubenfield, M., Weinstock, K., Lee, H.M., Dubois, J., Rosenthal, A., Platzer, M., Nyakatura, G., Taudien, S., Rump, A., Yang, H., Yu, J., Wang, J., Huang, G., Gu, J., Hood, L., Rowen, L., Madan, A., Qin, S., Davis, R.W., Federspiel, N.A., Abola, A.P., Proctor, M.J., Myers, R.M., Schmutz, J., Dickson, M., Grimwood, J., Cox, D.R., Olson, M.V., Kaul, R., Shimizu, N., Kawasaki, K., Minoshima, S., Evans, G.A., Athanasiou, M., Schultz, R., Roe, B.A., Chen, F., Pan, H., Ramser, J., Lehrach, H., Reinhardt, R., McCombie, W.R., de la Bastide, M., Dedhia, N., Blöcker, H., Hornischer, K., Nordsiek, G., Agarwala, R., Aravind, L., Bailey, J.A., Bateman, A., Batzoglu, S., Birney, E., Bork, P., Brown, D.G., Burge, C.B., Cerutti, L., Chen, H.C., Church, D., Clamp, M., Copley, R.R., Doerks, T., Eddy, S.R., Eichler, E.E., Furey, T.S., Galagan, J., Gilbert, J.G., Harmon, C., Hayashizaki, Y., Haussler, D., Hermjakob, H., Hokamp, K., Jang, W., Johnson, L.S., Jones, T.A., Kasif, S., Kasprzyk, A., Kennedy, S., Kent, W.J., Kitts, P., Koonin, E.V., Korf, I., Kulp, D., Lancet, D., Lowe, T.M., McLysaght, A., Mikkelsen, T., Moran, J.V., Mulder, N., Pollara, V.J., Ponting, C.P., Schuler, G., Schultz, J., Slater, G., Smit, A.F., Stupka, E., Szustakowski, J., Thierry-Mieg, D., Thierry-Mieg, J., Wagner, L., Wallis, J., Wheeler, R., Williams, A., Wolf, Y.I., Wolfe, K.H., Yang, S.P., Yeh, R.F., Collins, F., Guyer, M.S., Peterson, J., Felsenfeld, A., Wetterstrand, K.A., Patrinos, A., Morgan, M.J., de Jong, P., Catanese, J.J., Osoegawa, K., Shizuya, H., Choi, S., Chen, Y.J., Consortium, I.H.G.S., 2001. Initial sequencing and analysis of the human genome. *Nature* 409, 860-921.

Larsson, L., Li, X., Edström, L., Eriksson, L.I., Zackrisson, H., Argentini, C., Schiaffino, S., 2000. Acute quadriplegia and loss of muscle myosin in patients treated with nondepolarizing neuromuscular blocking agents and corticosteroids: mechanisms at the cellular and molecular levels. *Critical care medicine* 28, 34-45.

Lassar, A.B., Davis, R.L., Wright, W.E., Kadesch, T., Murre, C., Voronova, A., Baltimore, D., Weintraub, H., 1991. Functional activity of myogenic HLH proteins requires hetero-oligomerization with E12/E47-like proteins in vivo. *Cell* 66, 305-315.

Le Grand, F., Jones, A.E., Seale, V., Scimè, A., Rudnicki, M.A., 2009. Wnt7a activates the planar cell polarity pathway to drive the symmetric expansion of satellite stem cells. *Cell Stem Cell* 4, 535-547.

Le Grand, F., Rudnicki, M.A., 2007. Skeletal muscle satellite cells and adult myogenesis. *Current opinion in cell biology* 19, 628-633.

Le Guen, Y., Leroy, F., Philippe, C., Mangin, J.F., Dehaene-Lambertz, G., Frouin, V., Consortium, I., 2020. Enhancer Locus in ch14q23.1 Modulates Brain Asymmetric Temporal Regions Involved in Language Processing. *Cereb Cortex* 30, 5322-5332.

Lee, W.-C., Hough, M.T., Liu, W., Ekiert, R., Lindström, N.O., Hohenstein, P., Davies, J.A., 2010. Dact2 is expressed in the developing ureteric bud/collecting duct system of the kidney and controls morphogenetic behavior of collecting duct cells. *American Journal of Physiology-Renal Physiology* 299, F740-F751.

Levine, M., Tjian, R., 2003. Transcription regulation and animal diversity. *Nature* 424, 147-151.

Levy, S., Hannenhalli, S., Workman, C., 2001. Enrichment of regulatory signals in conserved non-coding genomic sequence. *Bioinformatics* 17, 871-877.

Li, J., Zhang, M., He, T., Li, H., Cao, T., Zheng, L., Guo, M., 2017a. Methylation of DACT2 promotes breast cancer development by activating Wnt signaling. *Scientific reports* 7, 1-10.

Li, R.N., Liu, B., Li, X.M., Hou, L.S., Mu, X.L., Wang, H., Linghu, H., 2017b. DACT1 Overexpression in type I ovarian cancer inhibits malignant expansion and cis-platinum resistance by modulating canonical Wnt signalling and autophagy. *Sci Rep* 7, 9285.

Li, X., Florez, S., Wang, J., Cao, H., Amendt, B.A., 2013. Dact2 represses PITX2 transcriptional activation and cell proliferation through Wnt/beta-catenin signaling during odontogenesis. *PloS one* 8, e54868.

Lickwar, C.R., Camp, J.G., Weiser, M., Cocchiaro, J.L., Kingsley, D.M., Furey, T.S., Sheikh, S.Z., Rawls, J.F., 2017. Genomic dissection of conserved transcriptional regulation in intestinal epithelial cells. *PLoS Biol* 15, e2002054.

Lim, K.R., Maruyama, R., Yokota, T., 2017. Eteplirsen in the treatment of Duchenne muscular dystrophy. *Drug Des Devel Ther* 11, 533-545.

Lin, F.T., MacDougald, O.A., Diehl, A.M., Lane, M.D., 1993. A 30-kDa alternative translation product of the CCAAT/enhancer binding protein alpha message: transcriptional activator lacking antimitotic activity. *Proc Natl Acad Sci U S A* 90, 9606-9610.

Logan, C.Y., Nusse, R., 2004. The Wnt signaling pathway in development and disease. *Annu. Rev. Cell Dev. Biol.* 20, 781-810.

Loots, G.G., Ovcharenko, I., 2007. *Mulan, Comparative Genomics*. Springer, pp. 237-253.

Lu, D., Zhao, Y., Tawatao, R., Cottam, H.B., Sen, M., Leoni, L.M., Kipps, T.J., Corr, M., Carson, D.A., 2004. Activation of the Wnt signaling pathway in chronic lymphocytic leukemia. *Proceedings of the National Academy of Sciences* 101, 3118-3123.

- Ma, B., Liu, B., Cao, W., Gao, C., Qi, Z., Ning, Y., Chen, Y.G., 2015. The Wnt Signaling Antagonist Dapper1 Accelerates Dishevelled2 Degradation via Promoting Its Ubiquitination and Aggregate-induced Autophagy. *J Biol Chem* 290, 12346-12354.
- Mandal, A., Waxman, J., 2014. Retinoic acid negatively regulates dact3b expression in the hindbrain of zebrafish embryos. *Gene Expr Patterns* 16, 122-129.
- Manickam, R., Oh, H.Y.P., Tan, C.K., Paramalingam, E., Wahli, W., 2018. Metronidazole Causes Skeletal Muscle Atrophy and Modulates Muscle Chronometabolism. *Int J Mol Sci* 19.
- Mansouri, A., St-Onge, L., Gruss, P., 1999. Role of Genes in Endoderm-derived Organs. *Trends Endocrinol Metab* 10, 164-167.
- Marcovitz, A., Jia, R., Bejerano, G., 2016. "Reverse Genomics" Predicts Function of Human Conserved Noncoding Elements. *Mol Biol Evol* 33, 1358-1369.
- Maston, G.A., Evans, S.K., Green, M.R., 2006. Transcriptional regulatory elements in the human genome. *Annu. Rev. Genomics Hum. Genet.* 7, 29-59.
- Mathelier, A., Shi, W., Wasserman, W.W., 2015. Identification of altered cis-regulatory elements in human disease. *Trends Genet* 31, 67-76.
- Mercer, E.M., Lin, Y.C., Murre, C., 2011. Factors and networks that underpin early hematopoiesis. *Seminars in Immunology* 23, 317-325.
- Moon, R.T., Kohn, A.D., De Ferrari, G.V., Kaykas, A., 2004. WNT and β -catenin signalling: diseases and therapies. *Nature Reviews Genetics* 5, 691.
- Nair, S.K., Burley, S.K., 2003. X-Ray Structures of Myc-Max and Mad-Max Recognizing DNA: Molecular Bases of Regulation by Proto-Oncogenic Transcription Factors. *Cell* 112, 193-205.
- Nègre, N., Brown, C.D., Ma, L., Bristow, C.A., Miller, S.W., Wagner, U., Kheradpour, P., Eaton, M.L., Loriaux, P., Sealfon, R., 2011. A cis-regulatory map of the Drosophila genome. *Nature* 471, 527-531.
- Ovcharenko, I., Nobrega, M.A., Loots, G.G., Stubbs, L., 2004. ECR Browser: a tool for visualizing and accessing data from comparisons of multiple vertebrate genomes. *Nucleic Acids Res* 32, W280-286.
- Pahlavan, Y., Mohammadi Nasr, M., Dalir Abdolahinia, E., Pirdel, Z., Razi Soofiyani, S., Siahpoush, S., Nejati, K., 2020. Prominent roles of microRNA-142 in cancer. *Pathol Res Pract* 216, 153220.
- Pardo, P.S., Boriek, A.M., 2011. The physiological roles of Sirt1 in skeletal muscle. *Aging (Albany NY)* 3, 430.
- Park, B., Pyo, S., 2013. The role of Ikaros in adipocyte differentiation of MDI-stimulated 3T3-L1 cells (P3111). *Am Assoc Immunol.*

Park, J.I., Ji, H., Jun, S., Gu, D., Hikasa, H., Li, L., Sokol, S.Y., McCrea, P.D., 2006. Frdo links Dishevelled to the p120-catenin/Kaiso pathway: distinct catenin subfamilies promote Wnt signals. *Dev Cell* 11, 683-695.

Parveen, N., Masood, A., Iftikhar, N., Minhas, B.F., Minhas, R., Nawaz, U., Abbasi, A.A., 2013. Comparative genomics using teleost fish helps to systematically identify target gene bodies of functionally defined human enhancers. *BMC Genomics* 14, 122.

Pelletier, G.J., Brody, S.L., Liapis, H., White, R.A., Hackett, B.P., 1998. A human forkhead/winged-helix transcription factor expressed in developing pulmonary and renal epithelium. *Am J Physiol* 274, L351-359.

Polychronopoulos, D., King, J.W., Nash, A.J., Tan, G., Lenhard, B., 2017. Conserved non-coding elements: developmental gene regulation meets genome organization. *Nucleic acids research* 45, 12611-12624.

Ponting, C.P., Phillips, C., Davies, K.E., Blake, D.J., 1997. PDZ domains: targeting signalling molecules to sub-membranous sites. *Bioessays* 19, 469-479.

Popelka, H., 2020. Dancing while self-eating: Protein intrinsic disorder in autophagy. *Prog Mol Biol Transl Sci* 174, 263-305.

Prendergast, G.C., Ziff, E.B., 1991. Methylation-sensitive sequence-specific DNA binding by the c-Myc basic region. *Science* 251, 186-189.

Prud'homme, B., Gompel, N., Carroll, S.B., 2007. Emerging principles of regulatory evolution. *Proceedings of the National Academy of Sciences* 104, 8605-8612.

Qin, L., Chen, Y., Niu, Y., Chen, W., Wang, Q., Xiao, S., Li, A., Xie, Y., Li, J., Zhao, X., 2010. A deep investigation into the adipogenesis mechanism: Profile of microRNAs regulating adipogenesis by modulating the canonical Wnt/ β -catenin signaling pathway. *BMC genomics* 11, 320.

Raab, J.R., Kamakaka, R.T., 2010. Insulators and promoters: closer than we think. *Nature Reviews Genetics* 11, 439-446.

Ragab, N., Viehweger, F., Bauer, J., Geyer, N., Yang, M., Seils, A., Belharazem, D., Brembeck, F.H., Schildhaus, H.U., Marx, A., Hahn, H., Simon-Keller, K., 2018. Canonical WNT/ β -Catenin Signaling Plays a Subordinate Role in Rhabdomyosarcomas. *Front Pediatr* 6, 378.

Raspopovic, J., Marcon, L., Russo, L., Sharpe, J., 2014. Modeling digits. Digit patterning is controlled by a Bmp-Sox9-Wnt Turing network modulated by morphogen gradients. *Science* 345, 566-570.

Richard, I., Broux, O., Allamand, V., Fougerousse, F., Chiannikulchai, N., Bourg, N., Brenguier, L., Devaud, C., Pasturaud, P., Roudaut, C., 1995. Mutations in the proteolytic enzyme calpain 3 cause limb-girdle muscular dystrophy type 2A. *Cell* 81, 27-40.

Rumfelt, L.L., Zhou, Y., Rowley, B.M., Shinton, S.A., Hardy, R.R., 2006. Lineage specification and plasticity in CD19- early B cell precursors. *J Exp Med* 203, 675-687.

Sabourin, L.A., Rudnicki, M.A., 2000. The molecular regulation of myogenesis. *Clin Genet* 57, 16-25.

Sakabe, N.J., Savic, D., Nobrega, M.A., 2012. Transcriptional enhancers in development and disease. *Genome Biol* 13, 238.

Sandri, M., Sandri, C., Gilbert, A., Skurk, C., Calabria, E., Picard, A., Walsh, K., Schiaffino, S., Lecker, S.H., Goldberg, A.L., 2004. Foxo transcription factors induce the atrophy-related ubiquitin ligase atrogin-1 and cause skeletal muscle atrophy. *Cell* 117, 399-412.

Santolini, M., Sakakibara, I., Gauthier, M., Ribas-Aulinas, F., Takahashi, H., Sawasaki, T., Mouly, V., Concordet, J.P., Defosse, P.A., Hakim, V., Maire, P., 2016. MyoD reprogramming requires Six1 and Six4 homeoproteins: genome-wide cis-regulatory module analysis. *Nucleic Acids Res* 44, 8621-8640.

Schröder, M., Becker-André, M., Carlberg, C., 1994. Thyroid hormone receptor functions as monomeric ligand-induced transcription factor on octameric half-sites. Consequences also for dimerization. *J Biol Chem* 269, 6444-6449.

Schubert, F.R., Sobreira, D.R., Janousek, R.G., Alvares, L.E., Dietrich, S., 2014. Dact genes are chordate specific regulators at the intersection of Wnt and Tgf- β signaling pathways. *BMC Evol Biol* 14, 157.

Shapiro, M.D., Bell, M.A., Kingsley, D.M., 2006. Parallel genetic origins of pelvic reduction in vertebrates. *Proceedings of the National Academy of Sciences* 103, 13753-13758.

Shapiro, M.D., Marks, M.E., Peichel, C.L., Blackman, B.K., Nereng, K.S., Jónsson, B., Schluter, D., Kingsley, D.M., 2004. Genetic and developmental basis of evolutionary pelvic reduction in threespine sticklebacks. *Nature* 428, 717-723.

Shi, Y., Ding, Y., Lei, Y.P., Yang, X.Y., Xie, G.M., Wen, J., Cai, C.Q., Li, H., Chen, Y., Zhang, T., 2012. Identification of novel rare mutations of DACT1 in human neural tube defects. *Human mutation* 33, 1450-1455.

Showalter, C.J., Engel, A.G., 1997. Acute quadriplegic myopathy: analysis of myosin isoforms and evidence for calpain-mediated proteolysis. *Muscle & nerve* 20, 316-322.

Sidhar, S.K., Clark, J., Gill, S., Hamoudi, R., Crew, A.J., Gwilliam, R., Ross, M., Linehan, W.M., Birdsall, S., Shipley, J., Cooper, C.S., 1996. The t(X;1)(p11.2;q21.2) translocation in papillary renal cell carcinoma fuses a novel gene PRCC to the TFE3 transcription factor gene. *Hum Mol Genet* 5, 1333-1338.

Skårn, M., Barøy, T., Stratford, E.W., Myklebost, O., 2013. Epigenetic regulation and functional characterization of microRNA-142 in mesenchymal cells. *PLoS One* 8, e79231.

Staal, F.J.T., Clevers, H.C., 2005. WNT signalling and haematopoiesis: a WNT–WNT situation. *Nature Reviews Immunology* 5, 21-30.

Stauber, M., Weidemann, M., Dittrich-Breiholz, O., Lobschat, K., Alten, L., Mai, M., Beckers, A., Kracht, M., Gossler, A., 2017. Identification of FOXJ1 effectors during ciliogenesis in the foetal respiratory epithelium and embryonic left-right organiser of the mouse. *Dev Biol* 423, 170-188.

Stenson, P.D., Mort, M., Ball, E.V., Howells, K., Phillips, A.D., Thomas, N.S., Cooper, D.N., 2009. The Human Gene Mutation Database: 2008 update. *Genome Med* 1, 13.

Stern, D.L., 1998. A role of Ultrabithorax in morphological differences between *Drosophila* species. *Nature* 396, 463-466.

Strähle, U., Rastegar, S., 2008. Conserved non-coding sequences and transcriptional regulation. *Brain Res Bull* 75, 225-230.

Su, Y., Zhang, L., Gao, X., Meng, F., Wen, J., Zhou, H., Meng, A., Chen, Y.G., 2007. The evolutionally conserved activity of Dapper2 in antagonizing TGF- β signaling. *The FASEB Journal* 21, 682-690.

Sucena, E., Delon, I., Jones, I., Payre, F., Stern, D.L., 2003. Regulatory evolution of shavenbaby/ovo underlies multiple cases of morphological parallelism. *Nature* 424, 935-938.

Suriben, R., Fisher, D.A., Cheyette, B.N., 2006. Dact1 presomitic mesoderm expression oscillates in phase with Axin2 in the somitogenesis clock of mice. *Dev Dyn* 235, 3177-3183.

Suriben, R., Kivimäe, S., Fisher, D.A., Moon, R.T., Cheyette, B.N., 2009. Posterior malformations in Dact1 mutant mice arise through misregulated Vangl2 at the primitive streak. *Nature genetics* 41, 977.

Suzuki, A., Yoshida, M., Hayashi, K., Mizuno, Y., Hagiwara, Y., Ozawa, E., 1994. Molecular organization at the glycoprotein-complex-binding site of dystrophin. Three dystrophin-associated proteins bind directly to the carboxy-terminal portion of dystrophin. *Eur J Biochem* 220, 283-292.

Suzuki, D., Leu, N.A., Brice, A.K., Senoo, M., 2014. Expression analysis of Dact1 in mice using a LacZ reporter. *Gene Expr Patterns* 15, 21-30.

Swart, G.W., van Groningen, J.J., van Ruissen, F., Bergers, M., Schalkwijk, J., 1997. Transcription factor C/EBPalpha: novel sites of expression and cloning of the human gene. *Biol Chem* 378, 373-379.

Szklarczyk, D., Gable, A.L., Lyon, D., Junge, A., Wyder, S., Huerta-Cepas, J., Simonovic, M., Doncheva, N.T., Morris, J.H., Bork, P., Jensen, L.J., Mering, C.V., 2019. STRING v11: protein-protein association networks with increased coverage, supporting functional discovery in genome-wide experimental datasets. *Nucleic Acids Res* 47, D607-D613.

Tajbakhsh, S., Borello, U., Vivarelli, E., Kelly, R., Papkoff, J., Duprez, D., Buckingham, M., Cossu, G., 1998. Differential activation of Myf5 and MyoD by different Wnts in explants of mouse paraxial mesoderm and the later activation of myogenesis in the absence of Myf5. *Development* 125, 4155-4162.

Takahashi, M., Tamura, K., Büscher, D., Masuya, H., Yonei-Tamura, S., Matsumoto, K., Naitoh-Matsuo, M., Takeuchi, J., Ogura, K., Shiroishi, T., Ogura, T., Izpisua Belmonte, J.C., 1998. The role of Alx-4 in the establishment of anteroposterior polarity during vertebrate limb development. *Development* 125, 4417-4425.

Tan, C.K., Chong, H.C., Tan, E.H., Tan, N.S., 2012. Getting 'Smad' about obesity and diabetes. *Nutr Diabetes* 2, e29.

Teran, E., Branscomb, A.D., Seeling, J.M., 2009. Dpr Acts as a molecular switch, inhibiting Wnt signaling when unphosphorylated, but promoting Wnt signaling when phosphorylated by casein kinase Idelta/epsilon. *PLoS One* 4, e5522.

Thurman, R.E., Rynes, E., Humbert, R., Vierstra, J., Maurano, M.T., Haugen, E., Sheffield, N.C., Stergachis, A.B., Wang, H., Vernet, B., 2012. The accessible chromatin landscape of the human genome. *Nature* 489, 75-82.

Torres, M., Gómez-Pardo, E., Gruss, P., 1996. Pax2 contributes to inner ear patterning and optic nerve trajectory. *Development* 122, 3381-3391.

Tournamille, C., Colin, Y., Cartron, J.P., Le Van Kim, C., 1995. Disruption of a GATA motif in the Duffy gene promoter abolishes erythroid gene expression in Duffy-negative individuals. *Nature genetics* 10, 224-228.

Toyota, M., Suzuki, H., Sasaki, Y., Maruyama, R., Imai, K., Shinomura, Y., Tokino, T., 2008. Epigenetic silencing of microRNA-34b/c and B-cell translocation gene 4 is associated with CpG island methylation in colorectal cancer. *Cancer research* 68, 4123-4132.

Trefilov, A., Berard, J., Krawczak, M., Schmidtke, J., 2000. Natal dispersal in rhesus macaques is related to serotonin transporter gene promoter variation. *Behav Genet* 30, 295-301.

Troschel, F.M., Böhly, N., Borrmann, K., Braun, T., Schwickert, A., Kiesel, L., Eich, H.T., Götte, M., Greve, B., 2018. miR-142-3p attenuates breast cancer stem cell characteristics and decreases radioresistance in vitro. *Tumor Biology* 40, 1010428318791887.

Tsivitse, S., 2010. Notch and Wnt signaling, physiological stimuli and postnatal myogenesis. *Int J Biol Sci* 6, 268-281.

Visel, A., Bristow, J., Pennacchio, L.A., 2007. Enhancer identification through comparative genomics. *Semin Cell Dev Biol* 18, 140-152.

Wahlestedt, M., Ladopoulos, V., Hidalgo, I., Sanchez Castillo, M., Hannah, R., Säwén, P., Wan, H., Dudenhöffer-Pfeifer, M., Magnusson, M., Norddahl, G.L., Göttgens, B., Bryder, D., 2017. Critical Modulation of Hematopoietic Lineage Fate by Hepatic Leukemia Factor. *Cell Rep* 21, 2251-2263.

Wang, J., Dai, X., Berry, L.D., Cogan, J.D., Liu, Q., Shyr, Y., 2019. HACER: an atlas of human active enhancers to interpret regulatory variants. *Nucleic Acids Res* 47, D106-D112.

Wang, J., Wynshaw-Boris, A., 2004. The canonical Wnt pathway in early mammalian embryogenesis and stem cell maintenance/differentiation. *Current opinion in genetics & development* 14, 533-539.

Wang, Y., Song, F., Zhang, B., Zhang, L., Xu, J., Kuang, D., Li, D., Choudhary, M.N.K., Li, Y., Hu, M., Hardison, R., Wang, T., Yue, F., 2018. The 3D Genome Browser: a web-based browser for visualizing 3D genome organization and long-range chromatin interactions. *Genome Biology* 19, 151.

Waxman, J.S., Hocking, A.M., Stoick, C.L., Moon, R.T., 2004. Zebrafish Dapper1 and Dapper2 play distinct roles in Wnt-mediated developmental processes. *Development* 131, 5909-5921.

- Webb, A.E., Brunet, A., 2014. FOXO transcription factors: key regulators of cellular quality control. *Trends Biochem Sci* 39, 159-169.
- Webb, B.D., Metikala, S., Wheeler, P.G., Sherpa, M.D., Houten, S.M., Horb, M.E., Schadt, E.E., 2017. Heterozygous Pathogenic Variant in DACT1 Causes an Autosomal-Dominant Syndrome with Features Overlapping Townes–Brocks Syndrome. *Human mutation* 38, 373-377.
- Wen, J., Chiang, Y.J., Gao, C., Xue, H., Xu, J., Ning, Y., Hodes, R.J., Gao, X., Chen, Y.-G., 2010. Loss of Dact1 disrupts planar cell polarity signaling by altering dishevelled activity and leads to posterior malformation in mice. *Journal of Biological Chemistry* 285, 11023-11030.
- Woolfe, A., Goodson, M., Goode, D.K., Snell, P., McEwen, G.K., Vavouri, T., Smith, S.F., North, P., Callaway, H., Kelly, K., Walter, K., Abnizova, I., Gilks, W., Edwards, Y.J., Cooke, J.E., Elgar, G., 2005. Highly conserved non-coding sequences are associated with vertebrate development. *PLoS Biol* 3, e7.
- Wray, G.A., 2007. The evolutionary significance of cis-regulatory mutations. *Nat Rev Genet* 8, 206-216.
- Wu, Z., Rosen, E.D., Brun, R., Hauser, S., Adelmant, G., Troy, A.E., McKeon, C., Darlington, G.J., Spiegelman, B.M., 1999. Cross-regulation of C/EBP alpha and PPAR gamma controls the transcriptional pathway of adipogenesis and insulin sensitivity. *Mol Cell* 3, 151-158.
- Yaffe, D., Saxel, O., 1977a. A myogenic cell line with altered serum requirements for differentiation. *Differentiation* 7, 159-166.
- Yaffe, D., Saxel, O., 1977b. Serial passaging and differentiation of myogenic cells isolated from dystrophic mouse muscle. *Nature* 270, 725.
- Yang, L., Vaitheesvaran, B., Hartil, K., Robinson, A.J., Hoopmann, M.R., Eng, J.K., Kurland, I.J., Bruce, J.E., 2011. The fasted/fed mouse metabolic acetylome: N6-acetylation differences suggest acetylation coordinates organ-specific fuel switching. *J Proteome Res* 10, 4134-4149.
- Yang, Z.Q., Zhao, Y., Liu, Y., Zhang, J.Y., Zhang, S., Jiang, G.Y., Zhang, P.X., Yang, L.H., Liu, D., Li, Q.C., Wang, E.H., 2010b. Downregulation of HDPR1 is associated with poor prognosis and affects expression levels of p120-catenin and beta-catenin in nonsmall cell lung cancer. *Mol Carcinog* 49, 508-519.
- Yau, T.-O., Chan, C.-Y., Chan, K.-L., Lee, M.-F., Wong, C.-M., Fan, S.-T., Ng, I.O.-L., 2005a. HDPR1, a novel inhibitor of the WNT/ β -catenin signaling, is frequently downregulated in hepatocellular carcinoma: involvement of methylation-mediated gene silencing. *Oncogene* 24, 1607.
- Yin, X., Xiang, T., Li, L., Su, X., Shu, X., Luo, X., Huang, J., Yuan, Y., Peng, W., Oberst, M., Kelly, K., Ren, G., Tao, Q., 2013. DACT1, an antagonist to Wnt/ β -catenin signaling, suppresses tumor cell growth and is frequently silenced in breast cancer. *Breast Cancer Res* 15, R23.

Yoshida, N., Yoshida, S., Koishi, K., Masuda, K., Nabeshima, Y.-i., 1998. Cell heterogeneity upon myogenic differentiation: down-regulation of MyoD and Myf-5 generates 'reserve cells'. *Journal of cell science* 111, 769-779.

Yu, Q., Quinn, W.J., Salay, T., Crowley, J.E., Cancro, M.P., Sen, J.M., 2008. Role of beta-catenin in B cell development and function. *J Immunol* 181, 3777-3783.

Yu, Y., Qi, L., Wu, J., Wang, Y., Fang, W., Zhang, H., 2013. Kindlin 2 regulates myogenic related factor myogenin via a canonical Wnt signaling in myogenic differentiation. *PLoS One* 8, e63490.

Yuan, G., Wang, C., Ma, C., Chen, N., Tian, Q., Zhang, T., Fu, W., 2012. Oncogenic function of DACT1 in colon cancer through the regulation of β -catenin. *PLoS One* 7, e34004.

Zerbino, D.R., Wilder, S.P., Johnson, N., Juettemann, T., Flicek, P.R., 2015. The ensembl regulatory build. *Genome biology* 16, 56.

Zhang, C., Yang, P., Chen, Y., Liu, J., Yuan, X., 2018. Expression of DACT1 in children with asthma and its regulation mechanism. *Experimental and therapeutic medicine* 15, 2674-2680.

Zhang, L., Zhou, H., Su, Y., Sun, Z., Zhang, H., Zhang, L., Zhang, Y., Ning, Y., Chen, Y.-G., Meng, A., 2004a. Zebrafish Dpr2 inhibits mesoderm induction by promoting degradation of nodal receptors. *Science* 306, 114-117.

Zhang, P., Iwasaki-Arai, J., Iwasaki, H., Fenyus, M.L., Dayaram, T., Owens, B.M., Shigematsu, H., Levantini, E., Huettner, C.S., Lekstrom-Himes, J.A., Akashi, K., Tenen, D.G., 2004b. Enhancement of hematopoietic stem cell repopulating capacity and self-renewal in the absence of the transcription factor C/EBP alpha. *Immunity* 21, 853-863.

Zhang, Y., Wang, L., 2011. Nuclear receptor small heterodimer partner in apoptosis signaling and liver cancer. *Cancers (Basel)* 3, 198-212.

Zhang, Y., Xu, P., Park, K., Choi, Y., Moore, D.D., Wang, L., 2008. Orphan receptor small heterodimer partner suppresses tumorigenesis by modulating cyclin D1 expression and cellular proliferation. *Hepatology* 48, 289-298.

Zhiying, X., Zhou, P., Zhou, P.Z., Zhang, J., bo Cao, L., Liu, X.C., Li, J., 2013. miR-142-3p inhibits LPS-induced activation of NF- κ B by targeting IRAK1 in colorectal cancer. *Central European Journal of Immunology* 38, 416-420.

Zhou, P., Byrne, C., Jacobs, J., Fuchs, E., 1995. Lymphoid enhancer factor 1 directs hair follicle patterning and epithelial cell fate. *Genes Dev* 9, 700-713.

Zhou, Y., Fu, B., Xu, X., Zhang, J., Tong, X., Wang, Y., Dong, Z., Zhang, X., Shen, N., Zhai, Y., 2020. PBX1 expression in uterine natural killer cells drives fetal growth. *Science Translational Medicine* 12.

Zhou, Y., Song, T., Peng, J., Zhou, Z., Wei, H., Zhou, R., Jiang, S., Peng, J., 2016. SIRT1 suppresses adipogenesis by activating Wnt/ β -catenin signaling in vivo and in vitro. *Oncotarget* 7, 77707.

Zhu, K., Jiang, B., Yang, Y., Hu, R., Liu, Z., 2017. DACT1 overexpression inhibits proliferation, enhances apoptosis, and increases daunorubicin chemosensitivity in KG-1 α cells. *Tumour Biol* 39, 1010428317711089.

Zullo, A., Simone, E., Grimaldi, M., Musto, V., Mancini, F.P., 2018. Sirtuins as Mediator of the Anti-Ageing Effects of Calorie Restriction in Skeletal and Cardiac Muscle. *Int J Mol Sci* 19.

8. APÊNDICES

8.1 Apêndice 1

Sequências FASTA dos CNEs no genoma humano (GRCh37/hg19) identificados com o *ECR Browser* e seus tamanhos em pares de bases (pb).

* Distância relativa ao *Dact1*

CNE1

> CNE1 hg19 chr14:59028673-59029142 470pb | *72 kpb
 CTTTTTCATGAAGGAGGGACTCCAACTTTCCTTGGGAATTATCTAAACAAAGTCACGTGTC
 TCCTTTTCTCTGTTCCATCTTTCACATGAGAAAAGCATAAATCATGTAACCCAACAAATAAA
 CACAATTGACATGTTATTTCTTTTCTATAGAGACATCAGTTCTTTAATCTCCTATTTCT
 TGTAAGCAGGCCTGCTTTGGGGGACTGAAATCCTAATGTATACACAATGGTATTCTGTAA
 GTAAGAAAGAACTCTAATGGAACCTATGTTAGAGAAATTTAATGTAACATGACTATAT
 GAATACCTCTTGCTGTTTTGCTTCTCTCATCCATTTAGCTATAAAAACCAGAAAGCATT
 CCTGCTGAAAGGAAAGGAACAGCTAGCTAAGAGGAAAAATCTAACCGTAAAGTGCTGGCA
 ATGCTCGGTGCCAAGAACCTGAGACTCTCTTTTTTTAAGCATGATGTTGT

CNE2

> CNE2 hg19 chr14:59071439-59071876 438pb | *29,2 kpb
 AGCCTAATGCAGCACATTATCATTTGTTATTTGGGTTTAATAATTTTGACATCTTTTCAC
 TCATACACAAAAAGTCAGAACTGGTTTTATTTACTGTTGATTTCATCCTCCTGTGGATGA
 AATAACTGAAGCCTAAAGGAATGGACTAGTGCTACTGACCTGTTTGTAAATTTATTTT
 GTGTTAATAGTACACTTTGAGTATCTTTTCCACATTACAACTTTCTGAATTATAAATG
 TTTTCCTTACATTATTTAACAATGTACACTTTAAAAAATATAAAAGTTCAAACCTCAGGG
 GTTTCTCAGCAGTCATTAATTGTACATTTTGCACCTAAGCTGTTGCGTTTCTTGTA
 AGATCGCGCTTTGTGCTTCAGTTTGTACCTTTGTAGACTTATTTAATGAAATCATTCAA
 ATAAACCAAACTTGCTTT

CNE3

> CNE3 hg19 chr14:59082951-59083157 207pb | *17,7 kpb
 TCCTCAGTCTTTTCTTGAATGCCTGGATAAAAAGTCTATCATGTGTACTAAGTATGGTGTG
 TGCTTCTCCATTATTCATGGGTGTCAACTGCACAGGCACATTTCAATAGTAATTGTTATT
 ATGTAAATAATCGAATCTGAAATTACTAATACATGTAATTTACCTGAACTAAGTGAAT
 GAATCTTGAGTGTGTAGTTGCATTGAC

CNE4

> CNE4 hg19 chr14:59244279-59244528 250pb | *143,5 kpb
 TGCCGTGAACTGGGAAGTGGAAAGGCAGCAATTGGGCTGTGAGAGCCTCCTGGGTCTCCT
 ACTGCTGACTTCTCTATAAATCAGCCACAGACAGCTGCATCCACTCGCACCATGGGAT
 AGAAAAATAGACCTGTTGCTAGACAGCAACACAGGCTCCACAAACACCATGGGGTCAGA
 GGGCAAGGGGTCTTAAGCCAGGCACTAGCCAAGCAGGCCAGCAGCCCCGCAGATCCATT
 TGTTTATTTT

CNE5

> CNE5 hg19 chr14:59282228-59282497 270pb | *181,5 kpb
 CCTTGGGGATGATGCTGGCCAGAAAGAAATGTGCATGGCTATTATTGTGCCAATTTTATAA
 TTCATCTTTTAAATAATCTGTAAGTGAAGAGCTGAATGGAATTTATCAAACCTGCCTTGT
 TTCTCACTGGGGAACCTTTCTGTGTTTTGTTACTGAGCCAAGGCTGAATGCAGCTGTTGGA
 GCTGTCACCAGCATTTCTGGTTCACCCAAATATTGATAATAATTACTTCTGCTTCACATGG
 AAAGAATTTATTTAAATATCTGTAATGGAA

CNE6

> CNE6 hg19 chr14:59298964-59299212 249pb | *198 kpb
 AGGCAGCAACTCTGTGCCAGGCATCGGACTTCCAGGCTCATAAATGTGGAATCTGTCAGG
 CTGAGAACATAAGTTGCTCCAGCCCATTATTAGCCAGTAGAAAGCTTATAAAAAAGGACT
 TTATAAACCAAGGCACGAGTACATAATTTTCATGAGCAAAGCTGCAGTGTGGCAGTTCTTG
 GTTTAAAGCTGAAAGCCTTTCTAATGTGTTGCCAGAGGATTAAACGCTATCACCATGTTT
 CAGGCAGTG

CNE7

> CNE7 hg19 chr14:59331774–59331996 223pb | *231 kpb

TGGGGCTTTTGCCAAAAAAGCCAGAACGCCCTTAATCTCCTCTTTCCTCATGAAACTCAC
AGCTTTGGTTCGTCACAGATCTTGCAGATCTTGGCGGTGCTCTCCTACTCCCCAGTTTTG
TCTCAGCACCATATGGCTTTTGACAAAGTCAGGCCACAGCTTCAACCTAAACAGCTTAAC
ATTATCGAAAGCCCAGCCCACTGAGGAAATAACAGTTCTCGTT

CNE8

> CNE8 hg19 chr14:59349427–59349727 301pb | *248,7 kpb

ATAACACTTGTAATTTGCGATGTCATCAGCTAAAAATAAGAGACAATTATTACAAATCATA
TTGCTAATATTTTGAGAAACTGAGATCTAATGTTGCTAAATCAAATTCCTGTGAATTCCA
CAGGGCGCTCAGCGCATTATAAGATTACCCGTGTGCTCTAAGCATGTCCTCATTAAGGCTG
CTGGCTACTGTGTTTTTATTGCTCCCCCTGGGAGGTTTGTGAGTTAATGCTGCCAGTAGG
AAGTCGTGTTTTCTTAGCCTCCCTGGTATACAGGCCTCCCAAATAGGTCTGCAGCCAATG
A

8.2 Apêndice 2

Sequências FASTA das regiões estendidas (1 kb) dos CNEs no genoma humano (GRCh38/hg38) e suas sequências correspondentes em camundongo e galinha identificadas com *Ensembl*.

CNE1 (Promoter Flanking Region)

Human

>14 dna:chromosome chromosome:GRCh38:14:58561691:58562692:1

```
TTTTAAATGGATACTTTAAACTCAAATTGCATTGTTCCTTTTTCACATCTATATGAACATA
AAGTTTCATGAAAGGGCCCTTAAATTTCCCTGTTTCATGGGAGTCTTCCTTCATTTCACTT
TTTTTCTTTTTCTCTTAGAATGCTAAATAGCCTCCTCCTCTGGAGACTAAATTAGATGC
ATCTTGAATTCATTTTCGTACACAAGAAATCTTTTTCTTCCCCATAAGCCGTCAGTACCAT
TTATATTTTAAATGAAAGGCAACCCTTTTCATGAAGGAGGGACTCCAACCTTTCCCTGGGA
ATTATCTAAACAAAGTCACGTGTCTCCTTTTTCTCTGTTCCATCTTCACATGAGAAAGCAT
AAATCATGTAACCCAAACAAATAAACACAATTGACATGTTATTTCTTTTTCTATAGAGACA
TCAGTTCCTTAATCTCCTATTTCTTGTAAGCAGGCCTGCTTTGGGGGACTGAAATCCTAA
TGTATACACAATGGTATTCTGTAAGTAAGAAAGAACTCTAATGGAACCTCTATGTTAGAG
AAATTTAATGTAACATGACTATATGAATACCTCTTGCTGTTTTGCTTCTCTCATCCCATT
TAGCTATAAAACCAGAAAGCATTCCTGCTGAAAGGAAAGGAACAGCTAGCTAAGAGGAA
AAATCTAACCGTAAAGTGCTGGCAATGCTCGGTGCCAAGAACCTGAGACTCTCTTTTTTTT
AAGCATGATGTTGTATTGTGAAGGTTTTATAAGAAGTTAACTTTAACCTATTAAAAAAA
GAGTGATTTGGAAATAAAGGAAAAGGAGCAATTTTGCCCCCTTATTCCTCTCTAGTATAA
ACAAGGTATGAGGCTAATGCTTTCTTGGCATAACATGAACTCTGAACTAACAAAAGAGG
TGTAATTAAAGCTCACCTGGCCTCAGAGGTCTTCTTTCCATAAAAAGTTAAGCCATAATGG
CTGGCACTCCGTGGCCAAGATTGGCCCGAGTTATTGGCTGTG
```

Mouse

>12 dna:chromosome chromosome:GRCm38:12:71248929:71249930:1

```
GCTGGAAGATGTGAGGTAAAGAACTAAAAGTGGATGCTCTGAGCTAAAACCACGCTGTCC
CTTTTTACATCTCATGAACATAAAAGTTTTGCGAAAGGACCCCTCAATTCCATTTCGTATGA
TGTCCTTCCTTAATCCCATTTTTTTTTCTTTCTGCCTTAGAATGCTAAATATTCTACCTCT
CTAGAGACTCAATTAGGTGTGTCTTGAATTTATACTTTAAGACATAAGAAATCTTTGTTT
CCTCCATAAGCTGTGCAACCATTTATAGTTCAAATGAAAGGCAACAGTCTCCCCAGGAA
GAGACTCAAACCTCTCTTTGGGGATTATCTAAACAATGTCACGTGTCCCCTTTTTCTGGT
CCATCTTCACATAAGAAAGCATAAATCATGTGACCCAAACAAATAAACACAACGTGACATGT
TATTTCTTTTTCTTTAGAGACATCGGTCTTTTAACTCTCCTATTTCTTAGAAAAAATGCT
GTCTTGAAAGCACTGAACTCCTAATGTATATACAGTGGTATTGTGTGTGTAGGACGGGAG
AGAAAAGCCATGTAACCTGGACCACATGATCGTCCCTGCTGGTTTGCCTCTCTTGTCTGT
TCTGATATAAAACCAGAAAAGCGCTGCTGCTGACTGGGAGAATGAGCAGAGCTGAAAAGC
ACAGCCCTGTGCTTGCTGCAGGCAGAGCTCAGCGCCAGGAACCGGGTACTGGCTTGCAAG
CACGCTGTTTTATTGTTGAAGTTTAATACGTTTGCTTTACTGGTTTAGAAATAGACAGAA
GGCACAATGTTTCCCCATTCATCCTTCTTCAGTATAAACACCAAGTGGAGATAAAGCTCT
CTTGGCATAACACGAGGCTCTGAACTAACAAAGAGAAGCCATTAAAGCTCACCTGGCCCCA
AAAGGCTTCTTACCATAAAAGCTAAGCGGTTATGGCTGGCACTCTGTGGCCAAGATTAGT
TCTGGGTATTGGCAGGGGCTTTTAATTGGTCTCTGCTACAGAG
```

Chicken

>5 dna:chromosome chromosome:GRCg6a:5:55327908:55328909:1

```
TACGGCTCTTGGCTGCACACCCTTAGCTGCAGCGAGGCATTGGCAGAGCTCAGCGCACTA
ATTCACCTCCATGTAAACATGATTATTTTAAATGAATTCCAGACGAGGTTTAACTGTAAA
AACATCTCCAGTTTTTTGAAGTGTCTTAAGTGCTCAGCTCGGAGTTGTTTCTGTAAAGGC
AAGGACTTACCCCTTTCCTCGGTTTACAAAGAATATTTTTCAGAATTAATGATGTCCAGCA
CCCCCTTCCCCCTCCCCCCCCACCATAAAATGGTGCTCATAAGAAAGTCACAGGTCTGAC
AATTAAGCTGCACCAGGATTTTCACGGTTCAGATTTTCCCTTTCTTAGCTGGGCTGTTTCAT
TTCTTTACAAGCAGGAATGCTGTTTGCTCTCCAAGCCGAGACGAAGAAGAGAAGCGAGAA
CAGTAAGAGGTATTCAATTTGGCTGCATTACTTTAAATGCTTCTGGCACAGAGCTCCCATT
AGAGCTCGCGGGGAGAAGTGGCAGAAGAACATTGTGTATACATTGTAGTGATTGTACTTT
CTCTGAGCAGACACGCTTGCAAGGAGGCGGCAGATTAAAGAGCCGGTGTACTTCAGACGG
AGAAAAATAACATCTTGCCAGTTGTGTTTATTTGCAAGGACGCATGATTTATGCTTTCCC
ATTTGTAGAGTGTTGAGAGCAATGGATCACGTGGCTCTGTTTAGATAATCCCCGAGGAA
```

ATTCGGGCTGCATCGGGCAGGGAGAGATAGGGTTTGTCTCCAAGACGAGCAATGCTGGG
 AAAGGGCAGGGAGATTTCTCAAGTTCTAGGAATGAAGTTGAAGGCTCTGATCTTCTGAA
 AAGATTGAGTGTGCGTTGTGGGAATGGAGCAATTAAGATAACAACAAGGAGGGCCAGGTA
 TGCCGGGATGAGCGGTCCCTTCTGTTATTTACTGACCCAAAGCAGAGAGAAAACAGTGC
 TTCCCTAAGGGACCTTTACCCGTCCAACAGTGCAGCAAGAAG

CNE4 (*Enhancer*)

Human

>14 dna:chromosome chromosome:GRCh38:14:58777185:58778186:1
 TTGAGAAGACAAAGGCGCTTTGTATGTATTATTTCAGGTGTGTATTTAAAGCCAAGGCTGA
 TCCCAATAAAGTACAATAACCTAAAAGATTGATACTTTACAGACTGAAGAAAAATTTTC
 CACTTTAATCCCAAGCCCCAGACCTACCCCTCTCCCTTGAAAAAGCTTAGAAATGTTGTCT
 CTAGAGCTGGGCACTGAGAAGGACGGAAGATATGCTCAGCTCAGAGGACCTAAATTCTGA
 TTAGCTGCAGAGAAAATGACCAGCTCTGCAGCTTCTGGGGCACATTTGAGGGGCATCTCT
 CCAGGAGGACATAGCACCAAGCATCATTTGATTTACTGATCTTCCCGCTTCCACCCCTGC
 CCAGGGTTCCCATGGCTGCCGTGAACTGGGGAAGTGGAAAGGCAGCAATTGGGCTGTGAGA
 GCCTCTGGGTCTCCTACTGCTGACTTCCTCCTATAAATCAGCCACAGACAGCTGCATCC
 ACTCGCACCATGGGATAGAAAAATAGACCCGTGTGCTAGACAGCAACACAGGCTCCACAA
 ACACCATGGGGTCAGAGGGCAAGGGGTCTTAAGCCAGGCACTAGCCAAGCAGGCCCAGCA
 GCCCCGAGATCCATTTGTTTATTTTCACAGTGGCATTGCCTGTCTGTACTACACGAAGG
 AGGGAGACTCCCTTGCCCCCTCTGCTCTGCATGTGTAGGCCACTTGCTGCTCACCACCGGA
 AAGTCCCGCTGACCTGCTGGCTAGTTTATTTTAGATGCGTGTAGACTGTCTGCTAG
 GTCTAAGTTTTTGAAGCTGATGGGGAAGTCATCAAGTCTAGTACAGAGTATTTAGCAATTG
 AATTTTATCTGGTCCATTGAATATCTGATTGAACAAGAAAGGGAAACCTTGTACTTTGTA
 CTGACGTATTACAAACCTCCCCCTCCTTCAGCACCTCCCTCCAGGGCCTTCCTTTCAAATC
 AGCATCTGACCCACCAAGAGAAAATAGGTTTATGATGGCTGTG

Mouse

>12 dna:chromosome chromosome:GRCm38:12:71437751:71438752:1
 GGCCAGGAGTGTGTGAAGTAGCTATTGGGGAAGGACTGTAGGGCCACTACTCCTGAACGA
 CTCAGGGAGTGCTACATTTAGGTATATGTTTGGCCAGAGCCACGAAGGTGGTAATGCTG
 AAGTACTCTAACAAACAAAAGAAATATCTCCCAAGTTTTTGTTCGAAGTTCCACCCACAC
 CCTCCTTTGAGAAGGGCTAGAAAACAACATCTCCCAAGCTGGAGAACTATAAAGGATGGGA
 AATGTGTGTGGTGCATGGGATCTAAATCTCTCCAGGGGCCTCACTGAGGGTCTGCTTCCC
 AGGATGATGGAGTAGCAACACACACTGCCTGGTTTACTGTTCCCCAAACCCAGAGTTCCC
 ATGCCTGTCAATTCTGGGCTGAGAGAGCCTTCTGGGTCTCCTGGGTGGTGGCTTCCTTCTA
 TAAATCAACTCCAGACAGCTGCAGCCACTTGCTATCCTGGGATAGAAAAATAGACCCTGTT
 GCTAGACATCAACCTAGGCTCCACAAACATGGCGGGATCACAGGGCAGGGGGCACCTAGC
 CCAGGCAGGCACTAGGCAAGCAGGCCCACCAGCCGGGTAGCTCCGTTTGTATTTTCTC
 AGTGGTATTGCTGTCTCTACCAGGAGAAGGGGGGCACCCTCCGCCCCCTGCTGGGCTGAT
 GTGGAGGCCACTTACTCTTTACCACCGGAAAGGCCTGCTGACTGCTGGCTGATTTATTTA
 TTTTAGTTGCACATAGACTGTCTGCTAGGTCTAAGTTTGGAAAGCCAACAGGAAGTCATCC
 AATCTAACACCAAGTGTTCAGTGCTTGTGTTCTGTTTGGGCAAACACATATTAGATTGCG
 CAAGAAAGGGAAGTCTTCTGTTGGTATATTCCCAACTCCTTCCCCCTCCACCTCTCCCATC
 CCTCACCCCCCCCCACCCCTGCATTGTCAATTTACCTCAGGCCAGCTGTGAGATTGATTTCT
 TACATTTGACAACAACCACACTCTTCCAAAATAAGTTTTTTCA

Chicken

>5 dna:chromosome chromosome:GRCg6a:5:55230502:55231503:1
 TCCAGAGAAGAGCTCCGCAGTTCTGGGAAGTCTGGGGAGGGATTCAACGAGGGCACTGAG
 CTTTCTCAACAGGCAAACAGGAGGCCAGGCTGGGAGACCAGAGGAGCCTGGATGGGCAGA
 AAGGAAACAAGATGTTTTTACCACCCCTGGGACTAACCTTAGCTATCGATGCCCTCCCAACA
 ACAACTTGTTTGGGCAAGGCTCGTGGTTGGGTTGTGTCAGGAGAAGGAAGAGCAGCCAAA
 GCATTAACACCTTCCCTCTTGCCGTGCTGGTCCCTTGGAAGTGCCTCCTGGTGCAGACTTG
 GCAGCGGCCAGCAGGATGCTGAATGAAAGGGACATTTAGAGCTCAGGCGCAGGACAAACG
 ATGCTAATGATAAATCCATCCCCCGACGCTGCGCCGACACGGATGTCCACAGCTAAAATA
 AACAAATGGTCCGGGCGAGCTGACAGGCTTGTTTAGATTCTGCTCGGGATAAATCGCAGT
 GCCCTTTGATCCCGTGATGTTTTTGGCGCCTGTGTTGCTGTCTAGCAGCAGGGTATATTT
 TTCTGTCCCGAGGAGCAAGTGGATGCAGCTGCCTATAGCTGATTTATAGTAGGAAGCGAT
 TTGCAGGAAATCAAAGCTTCTCACACAGTCAGAAATTGCTGTCTTTTACCTCCACCATCG
 ATGCAATGATTAGAGCTCTAGGGACTTTGGTGGGGAAGATCAGCAAAGCGTTCTGTGCTG
 AGCATAAATCACTTTCTGCTGTGACTCACGAAGGATGCTCGCAGTGGGCTGCACCTGTCT

GTACATCCTGGTGTGCAGGCACTGACAGTATGTCCATTCAATCTCAATCCAGAGATTTGA
AACCCATGTTTCTGTGCTCTCTGCAATGTGCTGGACAGCAAAATGACGGCAGTAGGCAAG
AATACCAAATACCATAACCAAAACGTGCTTATCACACACACACTGCATTACCAGACATGTA
GTTTAGGAGCTGGTTACTGCTTGCACTTTCTGATAACTGCTC

CNE5 (Promoter Flanking Region)

Human

>14 dna:chromosome chromosome:GRCh38:14:58815144:58816145:1

TTTGCTGAATTCCACTTAGAATCCACAAGAACATACTGCCAAAACCCCTTTGTCTGGTGAG
GAAAGCATGCAGGACACTGTCTCAGTCCAAGGCCTCTGCTTAAAAATGTTGAAAGCATT
GTCTGCCTGCCTCCATCTCTGAGTCCCTGTCCCATGGACCAGAAAAAGGTGATGCTCAGA
CCCTTATGCCATCTTGTGCTATCTTTGTCCCATCAGCAACCGCATGTGGCTGTCTGTGAT
AAGTAATCGAACAATTGCTTAGGGAGTACATGTGCTAATGGTGCATGGCAGCTTTACAA
ACACTGTCTGTTGAGTTAGACCTGAATTGTCCTCTGATTTAAAGATTCCACTGTGATTA
TCAGAGCCTTGGGGATGATGCTGGCCAGAAAGAATGTGCATGGCTATTATTGTGCCAATT
TTATAATTCATCTTTTAATAATCTGTAAGTGAAGAGCTGAATGGAATTTATCAAACCCCTG
CCTTGTTTCTCACTGGGGAACCTTTCTGTGTTTTGTTACTGAGCCAAGGCTGAATGCAGCT
GTTGGAGCTGTCACCAGCATTCTGGTTCACCCAAATATTGATAATAATTACTTCTGCTTC
ACATGGAAAGAATTTATTTAAATATCTGTAATGGAAAGACCGTTATTATTGGTATTACAA
CTTTACAGTGAGTACATCAAAATGTAGATATGACATTCATAAACTGTGTTTGCAAAAAGC
TGTATCTGTCTCCTGCATAGTACACAGGACCCCTGGGCTAGTATTGCTGCTCATCTGTCT
TCATAGAAATAGGGAGAGTTATTACTAGAGAATAATTAAACCCACAGAAACGTCCAAAT
GTTGGCAGGCTGCAGCCATCCCCAACACCGCACAAACAGCCTGTTATGATCAGCGAAATCC
AATAGCCGGAGCTTTGAGCCTGAGGGGGTAAATAGCATAATGAATTTTCTTTACCCCTGTC
ATTACTCCTCTCCTGGCTTTCTTAGATTACAGTCACAAGCAG

Mouse

>12 dna:chromosome chromosome:GRCh38:12:71481174:71482175:1

AGAGTCAGCATTAATATATTATTAATTTTACCCAGAATTCACAACCATACTGCCAAATC
CCGTGTCTGGCTAGGAAAACACCTTAGTCCAAAGCCACTCGTTAGATATTCTGAAAGTAC
TGCCTACCTGCCTTCATCTCACGTCCATCAGGATAGGGTGATGCCAGACCTTTGTGTC
ATCTAAATTGTCTTGGTGCCATCAGTAACCATGGCTGTCTGTGGTAAGTAATGACACCTC
CGCTCAGGGAAGCTCACGTGCTAATGGTGTGTGGCAGCTTTACAAACACTGTCTGTTTGA
ATTAGAGAGGCGAATTGTTCTGTGATTTGAAGATTCCCTCAGTGAGTATCAGAGCCTTGGA
GATGACGCTGTCCAGAAAAGATGTGTATGGCCACTATTGTGCCAATTTTATAATTCATCT
TTTAATAATCTGTAAGTGAAGAGCCGAGTGGAACCTCATCAAACCCCTGCCTTGTTTCTCAC
TGGGGAGCTTTTCTGTGTTTTGTTTTGAGCCAAAGCTGAACGCAGCTGTTGGAGCTGTC
ACCAGGATTCTTGTTACCCAAATATTGATAATAATGACTCCTGCTTCACATGGAAAGCA
TTTATTTAAATATTTGTAATGGAAAGACTGTTATTATTGATATTAGAGCTTTGCAGTGAG
AATGTAATAACGTGGACATGACATTCATAAAACCCGTGTTTGGAAGAAGCTGTACCCGCCT
CCTGCGTGGCACACAGGGCCTTGAGCTGGTACTGTTGATCATCTGTTCTTCATAGAAATA
GGGAGAAGTTATTACCAGGGAATAATTAAACCCACAGAAAGGGCCAAATGCTGGCAGGCT
GCAACCATCCGCATCACCACACAACAGCCTGTTATGATCAGTGAAATCAAACCTGCGGGGA
CTTTGCGTCTGAGGGGGTGGGGCGGTGTAATGAATATACCTTACTCTCTCCCTCCCTGA
CTCGCCTGGATTACAGTAAAAAGCAAGCCCTCGCACGTTTTT

Chicken

>5 dna:chromosome chromosome:GRCg6a:5:55218724:55219725:1

AAGTTCCCTAATATTTAAACAGAAATTTCTCTGTTGAAACACGCACCCAGAGGAAGCATG
GGCATTGCTAAAAACGACCTTAGAGGTGGGAAGCATTTCCAGCCCCAGCGCAGTTTGC
TCTCCAGGGTGGGTGAGAGAAAACCTTACAACAGGGCAGTGCCAGCATCCCTGTTGGGCC
CTTCCTGCTGATCTCAGCAAGCTGCTGTGTGGTCTGTGTGGAGAGGTTGCAGCCTGCCAA
CACTTTGCTCTTTTCAGCTGCTTTAATTATTCTCTTGTAAATAACTCTTCCCTTCTTCTA
TGAAGAACAGATGATGGACACTGATCACAGGGTGTTCAAACACTGCTCTATCGAGAGAA
CACTTTCCAACACCTTCCCCACCTCCTTGATCAATGGGGGATGAGCTTGCAGCACCAG
CTCCAGAGCCCTAACTTCAGTGGTGACAGCACTGCTAACACTGATATTAAAGTAAATGCT
TTCCGTGTCGGGGCATGAGTAACAATAGTCAATATTTTGGTGAGGCTGAATTACAGGCAG
CAGTTGCAACAGATTGCCTTCAGTTTGGCTCTAACACCCCGCTCCCGAGTGAGGTACAAC
ACACCATCTGATTAATTCCACACTCTATTCTTCGGTTACAGTTTATTAAAGATGGATTA
AAACATTGGCACAAACAGTGGCCACAAGAATATTTTCCAGCAACATCATCCAGAAGGCT
TTAGAGACGTAAACTGTGTCTAATTTGGGGCACCATTTCATATGCAGTTTGAACCGGTGT
TCAGAAAACCAATGCTCACTGTTAAAGCAGTATTTGAAAAAGTAAGAAGCACAAACCGTGT

CCTCCTTCGACTCAAGGGATATACATTATTGCAAGTGTGATAAAATCTGAGAAGCACCTT
TGGGGCTTTAAAGTCAAAGGAAATACACCAGCAAGACATTGGAAAGACTTTGTTTGCCAG
CAGCTGGCAGGCACTCATCATGAGTGAGCAGTAAGCTTAGTA

CNE6 (Enhancer)

Human

>14 dna:chromosome chromosome:GRCh38:14:58831870:58832871:1
CTCTTAAGTATATATTTGTTTCTTTGTAATCTGCTTTTACTTTTAATTTGAACATCAAAC
CAATATCCTCTCCCCACTCTTTTTCATATTCACAGAAGAGGAGTCAGCCCTGTTGTTGAC
ACTAATTAATGACCACTCATATTTGGTTGGCAGAGAGATTTATCCCATGTCCTGTGAGTT
CTCTTCTCCTTTTTTTTGCACCTCTGCACCCTAGACATCTGGGCCACGATTGCCAGAAGCCT
CTGAGTGCTCCATCTCACTTTTCTCCCTCCCAAGCCTCCCCATGTGTCATCTCTCTGGTCA
CAGTGGTGTCCCAACCAGTGGTTTCTGCAGTTTCTTGACACTGAAACCTGCTCAGCCAGA
TGGCATAACTCCCTCCAGGCAGCAACTCTGTGCCAGGCATCGGACTTCCAGGCTCATAAA
TGTGGAATCTGTGAGGCTGAGAACATAAGTTGCTCCAGCCCATTATTAGCCAGTAGAAAG
CTTATAAAAAAGGACTTTATAAACCAAGGCACGAGTACATAATTTTCATGAGCAAAGCTGC
AGTGTGGCAGTTCTTGGTTTAAAGCTGAAAGCCTTTCTAATGTGTTGCCAGAGGATTAAA
CGCTATCACCATGTTTCAGGCAGTGATTTTTTATTGCATGTTTCATTTCAGGATTTGGCACT
GGAAAAAGAAGGAATTCTGCACACAGCAAATTGAAATTAGAAGGCAGTGAGTTGCTATTTG
AAGCAGGGATGCTATTGTTTTGAGGACATTATCTTCTGTTGAGGTGATTCCCATCTTGTG
CACAATCTCTCCATTTTTTCAATAGATATTTACCATTTTTTCATTCTTGATAAATCCTTTCT
GTAATGCCTGGTACTTTTTCAAAATTATTCTTCAAATAGACAAATAGAATCATTCTTGACG
AATAGAATGATGTGTTGCTTAACGATGGGGCTGTGTTCTGAGAAATGCTTCATTTAGGTG
ATTTTGTCTTGTGTGATTATTGTAGGGTGACTTACACAAA

Mouse

>12 dna:chromosome chromosome:GRCh38:12:71494521:71495522:1
CTTCTCATTGTAGCTTTATCCTTGCTAAGAGATAAAATTGCGAATGAATAAATACATATTA
AAGTCTCCAAATTACTTTGGTCTAAGTATATAATGTATGGTGTAAATCTGCTTCTATTT
CTAATTTGTTTCAGTAAGCCACCACCAAGCTTTCCAGTTTCCCTTGTGACAAGGGAACAG
TGTAATGCTAATGAATGATCGCTGGTATGTAGTGACAGATTTATTTCCCTCTGTGCTGAC
TCTTCTCTCCTTCTCCTCACTCCACAATCTCACCTTCTCCTCCCTCCAGGCCCTCCGTGCCAT
CCCTCTGGTTCAGCAGAAATCCAACCTCAGTCCCCATAGCCTCCTCCTGCACCTGAAAG
CCACTCGGTTAGCCACTAGCATGGCTCCTTCCGGGCGGCCATTCTGCCAAGATCGGACTT
CAGCCTCATAAATGTAGGATCTGTGACCTGAGAACATAAGTGCTCCAGCCCGCTATTAG
CCGATAGAAAAGCTTATAAAAAAGGACTTTATACACCAAGGCATGAATAAATAATTTTCAGGA
GCAGAGCAGCCGTGTGGCAGTTTGGGTTTTAAAGCTGAAAGCCTTTCTAGTGTGTTGCCA
GGGGATTGAATGCTATCACCGTGTTCAGGCAGTGATTTTTTATTGCATGTTTCATTTCAGG
TTTTTGGCATTGGGAGAGAAGGAATCTGCACACATCACACTGAAATTAGAAGGCAGCGTG
TTGCTATTTAAAGCCAGAGTGCTATTGTCTTGAAGAGATTATCTTTTTTGCTCCCTGGTTT
GTGTGTGTGTATGTGTGTGGGTGCGTGTGTGCGCTCTTGAGTGTCTGCCTACATTTTTTG
AAATTTACCACCTTTCAATTTCCATAAATAATCGTTCCTTATGCTTTCAAATTTATGCT
TCACAATAACAAGCACAATGGTGTGCTCACAAATACCAGGAGAGCAACGGCTGTACACAGA
AAGCCAGAGAATTCAATATAGAGTCGAATAGCTATGAGAACT

Chicken

>5 dna:chromosome chromosome:GRCg6a:5:55213577:55214578:1
TATATATATCTCATGTCTGACCACTAAACTCATCATACCCTTGCAGCCACTCTGTGCCC
ATTTAAGTTGGGATGAGGTTAAGGCTTAAGGCTGCCTTCAATCTGCACTGATGGCTAGGA
GCTCTGAGTTAAAGCCTGATGCAGTGTCAAAGCCCCACGACCTGCAAGCTTTGGCAGAA
GGATAATGCATAAAATAGATACTGTGCCATGCTTGGCTCACTTCAGTTGTCCCAATTCAC
TGTCAGCCCATTTGCTCCCAGAGCTGCTGTACACATGTGGAACCTTCTCTCTGCGTGTCTGA
GCAACCAAACCTCCTCACCAACCCACTAAAGAGAAACCACACGTACCCCTTGGCACACTGC
ACAGTGCTGAAAAACAGTGATAGCATTTAATCCTCTTGGCAGCACATTACAGCAGCTTT
CAGACTCCAAACAAGCTGTACACCGCCACAGCCTTCTTGTAAAATTTATTTATCATGGAT
TGGTTTATAGAGTCTTTTATTAGCTTCATATTGGCTAATAACGGCCTGGAGGTCTTGCAT
CCATCGGCTTGACAGATTCCGTGTTTATGAGGCAGGAGCTTGACTCCTGGAAGAGATTT
GCAGTCTGGAGTCACACTGTCTGCTGGACCTGGGTTTCAGGGGATATGCTGCAGTGAGCCA
TACGTGGCATTTGGGTTGGTTTGGTGGGACTGGGTTGGGTGCCTGTGTGCTTGTGCTCTTG
CTCTGGTGAGGACTTTTCCAGGAGGAATGAGATGTGCTTGAATCTGACAGTGTCCCCAAA
CAAAGCCATGTTATAAGTATTGTGAAGCAGTGTTGTCTTAAGCTTTGGATGGTAGCCAGC
CCTTTGGGAATCACTGGGGGCTTTTCAGGAGTTTGCCAGTTGTGATCTGTCCTGGATTCCC

AGCCCTGAGAGCATCCCTTTCTCCACTGCCCATGAAGATTATGGGTACACACTCCTTTC
TTCAAATCCCCAGGAAATGGGGGTAAAGCCAAGCCAAAGAAA

CNE7 (*Promoter Flanking Region*)

Human

>14 dna:chromosome chromosome:GRCh38:14:58864667:58865668:1
CCATTTATTTCTAGTAAAAGTTATGTAGCTTTTTTTGTTCTGATGGAATTCATAATGA
CTAAGTCACTGAAACACTTGCACTCTCTTAGGAGGGACGTGTGGAATTTTTTTCACAAA
GAAAGAGCATGTTGTTATAAATGCTTCATGTAACCAGTTTAATACATCACTGAGTAAGAG
ACACACAGAAAAAATTCCTCTTATTTGGTCCCTGGGAGACTGGGCATATGCCTCAGTGTA
AAGGTGGGAGGCTTGTGAGTGCTTATTTCTGGAAGTTGAGAAGCGTCTTTAGTCTACTG
TGTTGACAGCAAAGCCAGGATCAACAGATTGTTTTATTAATTTGTTGTAGGTTGAGTTGC
TTTTGCCTCTGGGCATTCTGTGTTCAACTGGGGCTTTTGCCAAAAAGCCAGAACGCCC
TTAATCTCCTCTTCTCATGAACTCACAGCTTTGGTTCGTACAGATCTTGCAGATCT
TGGCGGTGCTCTCTACTCCCCAGTTTTGTCTCAGCACCATATGGCTTTTGACAAAGTCA
GGCCACAGCTTCAACCTAAACAGCTTAACATTATCGAAAGCCCAGCCACTGAGGAAATA
ACAGTTCTCGTTTATTTCTGTAAGTCATCTGGCTGGGAGTGGGAAGAGGGTAGGGAAGTC
AGTGTGGTGAGGGATGGGGAACATGTGAAAGATGCAAGGAAAATAGGTGGATGTGGGAGT
GGGGAGGGGGCAGTTCTAACAGATAAGATGTAGTGGATTCTGAGCAATATTTTCCAAGTC
CAAGAAAAATACATTGCTCATATTCATTTTTTCCCCCTAAGGATAATGAACTTTCACCTG
GATTTAAAAAGGAGGAGCCAAAAGTCTGGTTTCAATTAACATGTTTTTACAATATACGGGG
TGTTTTCCATATTGCTTCCGTAGCTTGCAATGGATGCTTAGGACATAACGATGTCTTACA
TACTCGGCTGTGGGAATGTGATGTCAGAATCCTCAGGGCTGG

Mouse

>12 dna:chromosome chromosome:GRCh38:12:71526145:71527146:1
GTCCCAGGAAGATAATATCGCTCTTTTGTATGATTTTAATATTATATCAGCTAACCCAT
TCTCGGGTGGGATACATAAAATTCCTTATGAAGCAAGGGACAGCTGATGGGTGACTTAT
GTTTCCAATTTAACACACCACTGATAATGAGACCCTAGTAATTCATACTGTCTGGTCCCA
GGAGTCTGGGCACACACCTAAGTGTAATAATCCGGAGGCTTGCATGTGCTGATTTTCTGCA
AGTGGAGCTGTGTTTGGTCTACTGTGTTGACAACAGAGCCAGAACCAATGGGCAGTTGTA
TTGATGATTGTAAGTCAAGTTGTTTTTGCCTCTGGGCATTCTGTGTTCAACTGGTGTT
TTGCCAAAAAAGAAAAAGTCAGTGTTGCCCTTCATCTCCTCTCTACACTTGTGC
AGCTTCGGTTCGTACAGATCTTGTGATCTTGGTGGGTTTGCTCCCACTCCTTTGCTTTT
CTCTCAGAACCATATGGCTTTTTGACAAAGTCAGATCACAGCTCCAACCTAAATGGCCAG
CATCATCAAAGCCTGGCTTACTGAGGAAATGACCCCTGTTTTGTTGCTGGAGGTCATATGG
TTGCAAAGGGAATAGGATCTCAAAGGCAGCTTGGCGGAGACAGAGCATGAGGAAGAATGT
GTGCGTGTGTGCGTGTGTGCATGTGCACGTGCACGCTTGCACTTGTGCTTGTGGTATGCA
TATGCGTGGTGTGTGTGTGAACATGTGCACCTGGTCATGTGGTCAAATGCAGCAGATTG
TGAGTGACATTTTCCAAGTCCAAGAAACATCCATTTCTTTCTAAGGACAAGGGACTCTCA
CCGGCTTTAGAAAAAGGACTAGCCCTATCTGGTTTCAATCAGTGCCATTGAGAAGACAGC
GAGTTTCCCGTGTTTTCTTGGTTCACAGTGAGTGTGTAAGACACAACCTGAATCTTGTACG
CTTCAGAGCGGAGGAAGTTGAATGCCAGAGTCTCTCCTGGCTG

Chicken

>5 dna:chromosome chromosome:GRCg6a:5:55204687:55205688:1
ATAGAGACCATCAAGTGTTCTGCAGCTTGGAGGTGTATAACATTCTCTCACTGAACTTA
AACATCAAATAAATTAATTATATATGGCAACAATAGGAAATAAACGTAGTCTGTGTATAT
AAGGCACCCCTAATTGAAACCAGACCTTGGCTCTCACTTCCTTTGTGGTCCAGGTGGACG
CTCGTTGTCTATGAACATAACGAGGGTGCAACAAAGGCTTTTCATGAGCCCTGCAGCCC
CTGCTCAAGATCAGAGAGGCTTGTCTAGTTTGTGAGAGCTTTCTTCCTCCTGTTATAAATG
CAGGAAGAGGAAGCAGGGGTTGTTATTGCTGCAGCGAGCTTGGCTTTTGATAATATTA
CTGTTTAGGTTGGATGTGAAACCTGACTTTGTGCGAGAACCATATGGTGCTTAAATGATGT
CGCTGCCTTGCAAGTCTGCAGAGTGGAAAGCAGGGCCAGATTCCACAATATCTGTAAATT
ACCAAAGCTGTGAATGTAATGAAGAAGGAAGAAAGATTAAGGTAGCTCTGTTTTGATGGC
AAAGGTATAACAGAGTATGGGGATACCCAAAGGGAGAGAGGAGGTTACCCACAGTAAATT
ACAGCATCCTAGATATGGAATTAACCATCAGCATTAAGGATGGAGACTTGGAGAGCGTG
AGAGCCACAGACCTTATACAGACACACTCATCTTTCTTAGGTTATGGGGAACAACGTC
CCATCAATCTTCCAACAGCAGTAATGATATTTTAAAGCTTCTTTTGCACCTCATCTGCTG
CATCATGACTCAGTTCTTTGTAGGTGACACTCCATCCAGAGAGCGCAATACACCGTAAGG
GTCTCTGCTGGGAAAAGGAGATGATGTCTTGGAGCACTCAGCGCTTTCTGCCTCTTGAG
CAGTACTGTACAGAGCTGATGAAGTGCTCCAGCATCCTTCAGAATAAAAAGTGTATATA

AGTGTGAAATAATACTGATAAATCTGAGAGAAAAATAAGCAC

CNE8 (Promoter Flanking Region)

Human

>14 dna:chromosome chromosome:GRCh38:14:58882359:58883360:1

CATACTTGTATTCTTTCTTGGCCATTTTAAATTTAGAAGGTAGTAGGGCCAACCTAACACT
TGGGGCTTTAGAGGTTTGCTGCAGACAAAATATGCACAATTGAAGAGATTTTTGGAGTGA
AAGAGATGGTTTCATTAAAAATATTATAATGATAGAGTAGAGGTTGTCTTTTTTCATTGTT
CTGCAGAAAAAGGCTGCTCATGTGTCTTAATATATGTAACATTCTGATTAAAGGCCTTGA
GGAAAAATTTGGCATGTGCTGTATTAAACAAAGGCAAACCTCCACTTACATCTGCCAACAGC
TGAAAGGCAGATTCTTGTGGTAAGATCTCTGAACTATAAAACAATGCCATAATAACACTTG
TAATTTGCGATGTCATCAGCTAAAAATAAGAGACAATTATTACAAATCATATTGCTAATAT
TTTGAGAACTGAGATCTAATGTTGCTAAATCAAATTCCTGTGAATTCACAGGGCGCTC
AGCGCATTATAAGATTACCTGTGCTCTAAGCATGTCTCATTAAGGCTGCTGGCTACTG
TGTTTTTATTGCTCCCTGGGAGGTTTGTGTAGTTAATGCTGCCAGTAGGAAGTCGTGTT
TTCCTAGCCTCCCTGGTATACAGGCCTCCCAAATAGGTCTGCAGCCAATGATGTGCTAGA
TTAGGTTATGATGATTTGTCTTGCCATCTCTGTGAAAAGTGAAGGCAATCAGAAATCAGT
GTGAAGTACGGTCTAGCCATGTAAATATCTAAATAATCTCATCAGAAAAAGGAATGCTAT
AGTAGGTCCAGGATGAAAGTATTGAAAGATAAGTAAATCAGTAGCATTAAACGTTTTAGAT
TGGTGCTCATAGAGTTTGTGCTTTTGTGCATCACTATCACTATTGTTTAATTACGGTTAGA
AAGGGCTACTGAGCACTTACTGTGAGCTACGCACTGGGCAAAATGCTTCATGTGCATTAT
TTGATGATCCTTACCAAACCTTATTAATCTAGGTGCTATTTT

Mouse

>12 dna:chromosome chromosome:GRCm38:12:71542595:71543596:1

CCAGCTGGGGCTGTTAGAAGTTTGACTAGAGAGCAAACATTTTGAATAACTGGAGAGCTT
TGTGGAAAAACATGGGGACAAGGAATTTATGTGTGTTAGAATGATCAAGTAGACATTGCCT
TTTAGGTTGTTATGTGGAAACACACTTTGCGTGTGTTCTAATGTATGTAACAGTCTGATA
AAAGGCCTTAAGGAAAAACGTGGCGTGTGCTATATTAACAAAAGCAGACTCCACACGCATC
CTCCAACAGCTGAAAAGCAGATTCTGTGGTAACATCGCTGAATTATAAACACCGCCACA
ATAATGCTTGCAATTCTCAATGTCATCAGCTAAAATCAGAGACAATTATTGTAAATTACA
TTGCTAATATTGAGAAACGGAGATCTACTGTTACCAAATCTAATTCCTGTGAATTCCAA
GGGGCCCTAAGCATAATATAAGATTACACTGGGCTCCGAGCATGTCTCATTAAGGCAGC
GGCAACCACATTTTTTAAAAATCACTCTCTCTGGGAGGTTTGTGTGACTTAATGGTGCCAGGA
AGAAGGCATGTTTTGTTTTGTTA
GTGTCCATAGTGACAGGCTTCCAAAAGAGGGCTGTGGCCAGTGACACACTGAGTCAGGG
TGTGGCTGTTCAACTTTGCCACTTTGATGCAAAGTGAAGGCAGTTAGAAGCCAATGTGAA
ATGTGGTACCAATATTAACAGTGTGAGTACATAACCAGACTGAAATAAAGCATTTGTAAG
TGATGCTGTGATAGACTCCGGATGACGATGCATTGGGTAGGAGCATCAGGAGCCTTAATG
TTTAGACTAATGTTTTCTTCTTTACAACCATGGTCATTTACCTTAATTACTGTTATCAAT
GACTGCTGAGCACTTCTTGGGAACGTGCACAGGGCAAAATACTTTCTGTGTGTTATTTG
ATCATCCTTACCAAATCCTATTAAACTGGCAGCTATTATTGA

Chicken

>5 dna:chromosome chromosome:GRCg6a:5:55196150:55197151:1

ATGGAAATCATATTCTTCTGATTGTTTGCAGCACAGCTCGTCTCAAGGTCACTCTCATA
CATAAATTTCAAGTATCATCATATTAGCACCCTATTTCAGGCATTCAAAAACAAATCTAA
ACGGGAGAGCTCATCAATTTAGCTGTCTCATGGCTAAACCTCACCATTACGCATGGCTTC
CTTTCTTACTGATTACATGACAGTAGATTCACTACTATTTTTTTTTTTTATTTTAAATTTG
GTTAAATACAATGATGTGGTATCTTTTAATTGCCTTTTTTTTTTTCTTTCTTCCAACCTGA
GCTGCCAAGGGACGTAAGCAAATTATATGGCTTTACAGTTCTGCTTGCTGCACTCAGACT
TGTGTGGCCAGCGTACCAAAGACACTGTGAAAACACTACCTTTTACTGGAAGTGTAACCT
CAATAAACCCCCAGGAGAGCAATAAAAAATGCAGTAGCCAGCAGATTTAACAAGGACATGC
TCAGTGACAGTGTAATCTTATTATGTGCTTAATGCCCAGTGGAATTACAGGAATTTGAT
TTAGCAACATCAGATTTTCAGTTTCTCAAATTATACTGACTGCTTATAAATTAAGACTGCG
ATTTGGGATAATTGCTTTTCATTTGCATTGCCGAGCCCATCTATTATATTTATTTCTCAC
CACACTGTTTGTATTTTTGAAGTGCCATGGAGCATTGCCTTTGGGACATAAAGTGATGTG
ATGGAGAGCTTTCTTTTGCACTCAGTGACACTCAAGTTCCTTCCCATTAACCTGGGTAA
ACGTTCTACCTGAGCGATCTTTATATGCTTGAATTGTCTACTTTATGGTTTCCCAAACCT
GACTGAACACAGTGAGATATGAGCGGTGAAAGACTCTGTCTTATCCATCATTTTGAGCTG
GGACAGCATTTGCTGGGCTGCCCCGTGATCTACAGATGTGCAGAGCGAAGGTGAGTGCAACA
GTGTCTCACAGCCATACCCTGTGTCTGCTGAGTGGTTTTTG

8.3 Apêndice 3

Tabela 1 – Principais funções relatadas dos fatores de transcrição (TF) para os quais os oito CNEs identificados possuem sítios de ligação.

TF	Principais funções	CNE
AP4	Fator de transcrição da família <i>helix-loop-helix-zipper</i> (bHLH-ZIP) que contém um domínio básico, que é usado para ligação ao DNA, e os domínios HLH e ZIP, que são usados para oligomerização. O fator de transcrição AP4 ativa os genes virais e celulares ligando-se à sequência simétrica de DNA CAGCTG.	4;5
CDXA	Cdx pertence ao grande grupo de genes <i>homeobox</i> , que foram originalmente descritos como reguladores-chave do desenvolvimento do corpo no embrião. Os genes Cdx e seus homólogos <i>caudal</i> em <i>Drosophila</i> e <i>Xcad</i> em <i>Xenopus</i> pertencem ao <i>cluster ParaHox</i> , que é considerado um antigo parálogo do cluster de genes <i>Hox</i> .	1;4
COMP1	Interage com a proteína miogenina. Também foi relatado que COMP1 e HNF4 juntos atuam como principais fatores de transcrição na indução de miR-150, que então reprime p27 em células cancerosas.	4
FOX	Proteína de ligação ao RNA, homólogo de <i>fox-1</i> (<i>C. elegans</i>) 1; regula eventos de <i>splicing</i> alternativo por ligação a elementos 5'-UGCAUGU-3'. Regula o <i>splicing</i> alternativo de exons tecido-específicos e de exons diferencialmente <i>spliced</i> durante a eritropoiese (também conhecido como FOX1).	4
FOXD3	Forkhead box D3; liga-se à sequência consenso 5'-A [AT]T[AG]TTTGTTT-3' e atua como um repressor transcricional. Também atua como um ativador transcricional. Promove o desenvolvimento de células da crista neural a partir de progenitores do tubo neural. Restringe as células progenitoras neurais à linhagem da crista neural enquanto suprime a diferenciação interneurônio. Necessário para a manutenção de células pluripotentes nos estágios de pré-implantação e peri-implantação da embriogênese.	4
FOXF2	Forkhead box F2; provável ativador de transcrição para uma série de genes específicos do pulmão (também conhecido como FREAC2).	2;4
FOXO1	Forkhead box O1; fator de transcrição que é o principal alvo da sinalização da insulina e regula a homeostase metabólica em resposta ao estresse oxidativo. Liga-se ao elemento de resposta à insulina (IRE) com a sequência de consenso 5'-TT[G/A] TTTTG-3' e ao elemento de ligação da família Daf-16 (DBE) relacionado com a sequência de consenso 5'-TT[G/A]TTTAC-3'. Tem atividade suprimida pela insulina. Principal regulador do equilíbrio redox e do número de osteoblastos, controlando a massa óssea. Orquestra a função endócrina do esqueleto na regulação do metabolismo da glicose.	1;4

Continuação

TF	Principais funções	CNE
FOXO3	Forkhead box O3; ativador transcricional que desencadeia apoptose na ausência de fatores de sobrevivência, incluindo morte de células neuronais por estresse oxidativo. Reconhece e liga-se à sequência de DNA 5'-[AG]TAAA[TC]A-3'. Participa da regulação pós-transcricional de MYC: após a fosforilação por MAPKAPK5, promove a indução da expressão de miR-34b e miR-34c, dois reguladores pós-transcricionais de MYC que se ligam à 3' UTR do transcrito de MYC e impedem sua tradução.	4
FOXO4	Forkhead box O4; fator de transcrição envolvido na regulação da via de sinalização da insulina. Liga-se a elementos de resposta à insulina (IREs) e pode ativar a transcrição de IGFBP1. Regula negativamente a expressão de HIF1A e suprime a ativação transcricional induzida por hipóxia de genes modulados por HIF1A. Também está envolvido na regulação negativa do ciclo celular. Envolvido no aumento da atividade do proteassoma em células-tronco embrionárias (ESCs) por meio da ativação da expressão de PSMD11 em ESCs, levando a uma maior montagem do proteassoma 26S, seguido por maior atividade do proteassoma.	4
FOXJ2	Forkhead box J2 é um ativador transcricional da família <i>forkhead</i> , sua expressão começa muito cedo no desenvolvimento embrionário e é amplamente distribuída no adulto.	1;4
FOXQ1	Forkhead box Q1; desempenha papel na diferenciação do folículo piloso (também conhecido como HFH1).	4
HFH3	Forkhead box I1; ativador transcricional necessário para o desenvolvimento da audição normal, senso de equilíbrio e função renal. Necessário para a expressão de SLC26A4/PDS, JAG1 e COCH em um subconjunto de células epiteliais e para o desenvolvimento do sistema endolinfático no ouvido interno. Também necessário para a expressão de SLC4A1/AE1, SLC4A9/AE4, ATP6V1B1 e para a diferenciação de células intercaladas no epitélio dos túbulos renais distais (também conhecido como FREAC6; FOXI1).	1;4
HFH4	Forkhead box J1; fator de transcrição necessário para ciliogênese móvel e desenvolvimento do epitélio dos pulmões e rins.	1;4
HFH8	Forkhead box F1; fator de transcrição da morfogênese pulmonar. É expressa em células endoteliais e musculares lisas no pulmão embrionário e adulto (também conhecido como FOXF1).	1;4
HNF4	Hepatocyte Nuclear Factor 4 alpha; fator de transcrição controlado transcricionalmente. Liga-se a sítios de DNA necessários para a transcrição de alfa 1- antitripsina, apolipoproteína CIII, genes da transtirretina e HNF1-alfa. Pode ser essencial para o desenvolvimento do fígado, rim e intestino.	4

Continuação

TF	Principais funções	CNE
IK1	IKZF1 (DNA-binding protein Ikaros); regulador da transcrição da diferenciação de células hematopoiéticas. Liga ao DNA satélite gama. Desempenha um papel no desenvolvimento de linfócitos, células B e T. Liga e ativa o <i>enhancer</i> (elemento delta-A) do gene CD3-delta. Repressor do gene TDT (fikzfterminal desoxinucleotidiltransferase) durante a diferenciação de timócitos. Regula a transcrição por meio da associação com complexos dependentes e independentes de HDAC.	7
LBP1	Upstream binding protein 1 (LBP-1a); funciona como um ativador transcricional de dependente do contexto do promotor. Modula a expressão placentária de CYP11A1. Envolvido na regulação do gene da alfa-globina em células eritróides. A ativação do promotor da alfa-globina nas células eritroides é via interação sinérgica com TFCEP2. Liga-se fortemente às sequências em torno do sítio de iniciação HIV-1 e fracamente ao TATA-box. Reprime a transcrição do HIV-1 ao inibir a ligação de TFIID à TATA-box.	4;5
MYOD	MYOD1 - Myogenic differentiation 1; atua como um ativador transcricional que promove a transcrição de genes-alvo específicos do músculo e desempenha um papel na diferenciação muscular. Junto com MYF5 e MYOG, co-ocupa a região central do promotor de genes músculo-específicos durante a miogênese. Induz fibroblastos para se diferenciarem em mioblastos.	4
MYOGENIN	Myogenin (myogenic factor 4); atua como um ativador transcricional que promove a transcrição de genes-alvo específicos do músculo e desempenha um papel na diferenciação muscular, saída do ciclo celular e atrofia muscular. Essencial para o desenvolvimento da diferenciação funcional da fibra esquelética embrionária. No entanto, é dispensável para o crescimento do músculo esquelético pós-natal; a fosforilação por CAMK2G inibe sua atividade transcricional em resposta à atividade muscular. Necessário para o recrutamento do complexo FACT para regiões promotoras específicas do músculo, promovendo assim o início da expressão gênica.	4
PPAR	Peroxisome proliferator-activated receptor alpha; fator de transcrição ativado por ligante. Regulador chave do metabolismo lipídico. Ativado pelo ligante endógeno 1-palmitoil-2-oleoil-sn-glicerol-3-fosfolina (16:0/18:1-GPC). Ativado por oleetanolamida, um lipídio que ocorre naturalmente e regula a saciedade. Receptor para proliferadores de peroxissoma, como drogas hipolipemiantes e ácidos graxos. Regula a via de oxidação beta peroxissômica dos ácidos graxos. Atua como ativador da transcrição para os genes ACOX1 e P450.	4
PPARG	Peroxisome proliferator activated receptor gamma; receptor nuclear que se liga a proliferadores de peroxissoma, como drogas hipolipemiantes e ácidos graxos. Uma vez ativado por um ligante, o receptor nuclear se liga a elementos de resposta PPAR específicos do DNA (PPRE) e modula a transcrição de seus genes-alvo, como a acil-CoA oxidase. Portanto, controla a via de oxidação beta peroxissômica dos ácidos graxos. Regulador chave da diferenciação de adipócitos e homeostase da glicose.	4

SMAD2 - SMAD family member 2; SMAD regulado por receptor (R-SMAD) que é um transdutor de sinal intracelular e modulador transcricional ativado por TGF-beta e receptores quinases. Liga ao elemento TRE na região promotora de muitos genes que são regulados por TGF-beta e, na formação do complexo SMAD2/SMAD4, ativa a transcrição. Pode atuar como supressor de tumor no carcinoma colorretal. Regula positivamente a atividade da quinase PDPK1, estimulando sua dissociação da proteína 14-3-3 YWHAQ, que atua como um regulador negativo.

SMAD3 - SMAD family member 3; possui as mesmas funções que SMAD2 e também pode formar um complexo SMAD3/SMAD4/JUN/FOS no sítio AP-1/SMAD para regular a transcrição mediada por TGF-beta.

SMAD4 - SMAD family member 4; na fisiologia muscular desempenha um papel central no equilíbrio entre atrofia e hipertrofia. Quando recrutado por MSTN, promove a resposta à atrofia via SMAD2/4 fosforilado. A diminuição de MSTN causa liberação de SMAD4 e subsequente recrutamento pela via BMP para promover hipertrofia via SMAD1/5/8 fosforilado. Age sinergicamente com SMAD1 e YY1 na expressão gênica cardíaca-específica mediada por proteína morfogenética óssea (BMP). Liga-se a elementos ligantes de SMAD (SBEs) (5'-GTCT/AGAC-3') dentro do elemento de resposta BMP (BMPRE) das regiões de ativação cardíaca.

SMAD

4

SMAD5 - SMAD family member 5; modulador transcricional ativado por BMPs. SMAD5 é um SMAD regulado por receptor.

SMAD6 - SMAD family member 6; atua como mediador da atividade anti-inflamatória TGF-beta e BMP. Suprime a sinalização de IL1R-TLR por meio de sua interação direta com PEL1, evitando a ativação de NF-kappa-B, o transporte nuclear e expressão mediada por NF-kappa-B de genes pró-inflamatórios. Pode bloquear a via de sinalização BMP-SMAD1 competindo com SMAD4 pela ligação SMAD1 ativada por receptor. Liga-se a elementos reguladores em regiões promotoras alvo.

SMAD7 - SMAD family member 7; antagonista da sinalização por membros da superfamília do receptor tipo 1 TGF-beta; foi mostrado inibindo TGF-beta e a sinalização de ativina ao se associar a seus receptores, evitando assim o acesso ao SMAD2. Funciona como um adaptador para recrutar SMURF2 para o complexo receptor TGF-beta. Também atua recrutando o complexo PPP1R15A-PP1 para TGFBR1, que promove sua desfosforilação.

SMAD9 - SMAD family member 9; modulador transcricional ativado por BMPs. SMAD9 é um SMAD regulado por receptor (R-SMAD).

SRY

Sex determining region Y; regulador transcricional que controla o desenvolvimento masculino. É necessário e suficiente para iniciar a determinação do sexo masculino, direcionando o desenvolvimento de precursores celulares de suporte (células pré-Sertoli) em células de Sertoli ao invés de células da granulosa. No cérebro adulto masculino está envolvido na manutenção das funções motoras dos neurônios

4

dopaminérgicos. Envolvido em diferentes aspectos da regulação gênica, incluindo ativação ou repressão do promotor.

SOX4 - SRY (sex determining region Y)-box 4; ativador transcricional que se liga com alta afinidade ao *motif* 5'-AACAAAG-3' do *enhancer* em células T.

SOX9 - SRY (sex determining region Y)-box 9; desempenha um papel importante no desenvolvimento normal do esqueleto. Pode regular a expressão de outros genes envolvidos na condrogênese, agindo como um fator de transcrição para esses genes.

SOX30 - SRY (sex determining region Y)-box 30; ativador transcricional que liga-se à sequência de DNA 5'-ACAAT-3' e mostra uma preferência por resíduos de guanina em torno do *core* deste *motif*.

SOX8 - SRY (sex determining region Y)-box 8; deve desempenhar um papel no sistema nervoso central, membros e desenvolvimento facial. Pode estar envolvido na determinação do sexo masculino. Liga ao *motif* consenso 5'-[AT][AT]CAA[AT]G-3'.

SOX17 - SRY (sex determining region Y)-box 17; atua como regulador da transcrição que se liga ao DNA do promotor alvo e dobra o DNA. Liga-se às sequências 5'-AACAAAT-3' ou 5'-AACAAAG-3'. Modula a regulação da transcrição via WNT3A, inibindo a sinalização Wnt. Promove a degradação de CTNNB1 ativado. Desempenha um papel fundamental na regulação do desenvolvimento embrionário. Necessário para *loop* normal do tubo cardíaco embrionário. Necessário para o desenvolvimento normal da endoderme intestinal definitiva. Provável ativador transcricional nas células germinativas pré-meióticas.

SOX7 - SRY (sex determining region Y)-box 7; liga-se e ativa o promotor CDH5, portanto, desempenha um papel na regulação da transcrição de genes expressos no endotélio do tecido sanguíneo e bloqueia a diferenciação terminal em precursores sanguíneos. Pode ser necessário para a sobrevivência de precursores hematopoiéticos e endoteliais durante a especificação. Compete com GATA4 pela ligação e ativação do promotor FGF3. Reprime a transcrição estimulada por Wnt/beta-catenina, provavelmente por direcionar CTNNB1 para a degradação proteossomal.

SOX14 - SRY (sex determining region Y)-box 14; atua como um regulador negativo da transcrição de genes.

SOX11 - SRY (sex determining region Y)-box 11; Fator de transcrição envolvido na neurogênese embrionária. Também pode ter um papel na modelagem do tecido durante o desenvolvimento.

SOX2 - SRY (sex determining region Y)-box 2; fator de transcrição que forma um complexo trimérico com OCT4 no DNA e controla a expressão de vários genes envolvidos no desenvolvimento embrionário, como YES1, FGF4, UTF1 e ZFP206. Crítico para a embriogênese inicial e para a pluripotência de células-tronco

embrionárias. Deve funcionar como um *switch* no desenvolvimento neuronal. Mantém as células neurais indiferenciadas, neutralizando a atividade das proteínas proneurais.

SOX6 - SRY (sex determining region Y)-box 6; ativador transcricional. Liga-se especificamente à sequência de DNA 5'-AACAAT-3'. Desempenha um papel fundamental em vários processos de desenvolvimento, incluindo neurogênese e formação do esqueleto.

SOX1 - SRY (sex determining region Y)-box 1; ativador transcricional. Pode funcionar como um *switch* no desenvolvimento neuronal. Mantém as células neurais indiferenciadas, neutralizando a atividade das proteínas proneurais e suprime a diferenciação neuronal.

SOX18 - SRY (sex determining region Y)-box 18; liga-se à sequência consenso 5'-AACAAAG-3' e é capaz de transativar a transcrição através deste sítio.

SOX12 - SRY (sex determining region Y)-box 12; liga-se à sequência 5'-AACAAT-3'.

SOX10 - SRY (sex determining region Y)-box 10; fator de transcrição que parece funcionar sinergicamente com a proteína do domínio POU TST-1/OCT6/SCIP. Pode conferir especificidade celular para a função de outros fatores de transcrição na glia em desenvolvimento e madura.

SOX15 - SRY (sex determining region Y)-box 15; liga-se à sequência 5'-AACAAT-3'.

SOX13 - SRY (sex determining region Y)-box 13; liga-se à sequência 5'-AACAAT-3'.

SOX3 - SRY (sex determining region Y)-box 3; fator de transcrição necessário durante a formação do eixo hipotálamo-hipófise. Pode funcionar como um *switch* no desenvolvimento neuronal. Mantém as células neurais indiferenciadas, neutralizando a atividade das proteínas proneurais e suprime a diferenciação neuronal. Necessário também dentro do epitélio faríngeo para morfogênese craniofacial. É necessário para iniciar a determinação do sexo masculino.

SOX21 - SRY (sex determining region Y)-box 21; deve desempenhar papel como ativador da transcrição de OPRM1.

Continuação

TF	Principais funções	CNE
XFD1	XFD1 forkhead related gene em <i>Xenopus laevis</i> . Essencial no desempenho da função de organizador de Spemann durante a formação do eixo dorsal e na indução neural.	1
XFD2	XFD2 forkhead related gene em <i>Xenopus laevis</i> . Atua na sinalização ventral durante a embriogênese.	4
XVENT1	Xvent-1 atua em uma via de sinalização ventral que mantém o estado mesodérmico ventral e antagoniza o organizador de Spemann, controlando a padronização dorsoventral do mesoderma de <i>Xenopus</i> .	4
ARNT	Necessário para a atividade do receptor Ah (dioxina). Esta proteína é necessária para que a subunidade de ligação ao ligante seja translocada do citosol para o núcleo após a ligação do ligante. O complexo então inicia a transcrição dos genes envolvidos na ativação dos procarcinogênicos da HAP. O heterodímero com HIF1A ou EPAS1/HIF2A funciona como um regulador transcricional da resposta adaptativa à hipóxia.	1
CLOCK-BMAL	As proteínas CLOCK e BMAL1 são membros da superfamília de fatores de transcrição <i>helix-loop-helix</i> PAS (PER-ARNT-SIM) que se ligam ao <i>motif</i> E-box do DNA. Eles são necessários para o alto nível de expressão de genes do relógio circadiano (<i>per</i> e <i>cry</i>).	1
CRX	Cone-rod homeobox containing gene; fator de transcrição que se liga e ativa a sequência 5'-TAATC[CA] -3' que é encontrada <i>upstream</i> de vários genes específicos de fotorreceptores, incluindo os genes <i>opsina</i> . Age sinergicamente com outros fatores de transcrição, como NRL, RORB e RAX, para regular a transcrição de genes específicos de células fotorreceptoras. Essencial para a manutenção de fotorreceptores de mamíferos.	1
EBOX	Elemento promotor reconhecido por fatores de transcrição da classe <i>helix-loop-helix</i> . Sua sequência é reconhecida e ligada por fatores de transcrição para iniciar a transcrição do gene.	1
FREAC4	Forkhead box D1; fator de transcrição necessário para a formação da identidade posicional na retina em desenvolvimento, regionalização do quiasma óptico e morfogênese do rim. Pode neuralizar células ectodérmicas diretamente.	1
MAX	A proteína MAX é o componente de dimerização mais conservado da rede MYC-MAX-MXD1 de fatores básicos de transcrição <i>helix-loop-helix leucine zipper</i> (bHLHZ) que regulam a proliferação, diferenciação e apoptose celular. Enquanto a heterodimerização de MAX com MYC ativa e medeia sua atividade de transcrição e transformação, a heterodimerização de MAX com membros da família MXD1 antagoniza a transformação	1

celular dependente de MYC por repressão transcricional das mesmas sequências de DNA E-box-alvo.

MYCMAX	O protooncogene MYC codifica um fator de ligação ao DNA que pode ativar e reprimir a transcrição. Por meio desse mecanismo, o MYC regula a expressão de vários genes-alvo que controlam funções celulares importantes, incluindo o crescimento celular e a progressão do ciclo celular. MYC também tem um papel crítico na replicação do DNA. A expressão de MYC desregulada resultante de vários tipos de alterações genéticas leva à atividade de MYC constitutiva em uma variedade de cânceres e promove a oncogênese.	1
NMYC	Fator de transcrição pertencente à família MYC. É expresso principalmente em embriões em desenvolvimento normal e é considerado crítico no desenvolvimento cérebro e em outros tecidos neurais.	1
PAX2	Promove a indução da nefrogênese no desenvolvimento dos rins; no embrião humano, a PAX2 é expressa na vesícula ótica e posteriormente na retina, nos canais semicirculares do ouvido interno e nos mesonefros, metanefros, supra-renais, medula espinhal e rombencéfalo.	1
PBX1	Pre B cell leukemia homeobox 1; desempenha um papel na regulação dependente de cAMP da expressão do gene CYP17 por meio de sua sequência reguladora de cAMP (CRS1) 5'-ATCAATCAA-3'. Atua como um ativador transcricional de PF4 em complexo com MEIS1. Pode ter um papel na esteroidogênese e, subsequentemente, no desenvolvimento e diferenciação sexual. A isoforma PBX1b como parte de um complexo PDX1: PBX1b: MEIS2b em células acinares pancreáticas está envolvida na ativação transcricional do <i>enhancer</i> de ELA1.	1

Continuação

TF	Principais funções	CNE
POU6F1	POU domain, class 6, transcription factor 1; fator de transcrição que se liga preferencialmente a uma variante do <i>motif</i> do octâmero 5'-ATGATAAT-3'.	1
SREBP1	Sterol regulatory element binding transcription factor 1; ativador transcricional necessário para a homeostase lipídica. Regula a transcrição do gene do receptor de LDL, bem como do ácido graxo e, em menor grau, a via de síntese do colesterol. Se liga ao elemento regulador do esterol 1 (SRE-1) (5'-ATCACCCCAC-3'). Tem uma especificidade de ligação tanto para o motif E-box (5'-ATCACGTGA-3') quanto para SRE-1 (5'-ATCACCCCAC-3'). A isoforma SREBP-1A é mais ativa que a isoforma SREBP-1C no estímulo da transcrição de promotores contendo SRE-1.	1
STRA13	F; componente de ligação ao DNA do complexo central FA envolvido no reparo de danos ao DNA e manutenção do genoma. Componente do complexo CENP-T-W-S-X heterotetramérico que se liga e compacta o DNA e desempenha um papel importante na montagem do cinetocoro. Desempenha um papel importante na progressão mitótica e na segregação cromossômica.	1
USF	Upstream stimulatory factor (USF); fator de transcrição celular expresso de forma ubíqua. USF purificado consiste em 2 polipeptídeos relacionados, USF1 e USF2. USF1 pertence à família MYC de proteínas de ligação ao DNA.	1
CBF	Cebpz - CCAAT/enhancer binding protein zeta; estimula a transcrição do promotor HSP70.	2
E4BP4	Nfil3 - Nuclear factor, interleukin 3, regulated; atua como um regulador transcricional que reconhece e se liga à sequência 5'- [GA]TTA[CT]GTAA[CT]-3', uma sequência presente em diversos promotores celulares e virais. Reprime a atividade do promotor em osteoblastos. Reprime a atividade transcricional de PER1. Reprime a atividade transcricional de PER2. Ativa a transcrição do promotor da interleucina-3 em células T.	2
HLF	Member of the PAR bZIP (proline and acidic amino acid-rich basic leucine zipper) transcription factor family. HLF é expresso no fígado adulto e em uma linhagem de carcinoma hepatocelular e é expresso com níveis mais baixos no rim e no pulmão de adultos. O HLF é expresso no fígado fetal em um nível significativamente menor do que o observado no fígado adulto. Nenhuma expressão foi detectada em células mononucleares de sangue periférico normal ou em qualquer linhagem celular hematolinfóide. HLF se liga ao DNA especificamente como um homodímero ou como um heterodímero com outros fatores PAR.	2
ISRE	Elemento de resposta estimulada por interferon (ISRE) do promotor de MHC classe I.	2

Continuação

TF	Principais funções	CNE
LEF1	Lymphoid enhancer binding factor 1; participa da via de sinalização Wnt. Ativa a transcrição de genes alvo na presença de CTNNB1 e EP300. Pode desempenhar um papel na diferenciação das células ciliadas e na morfogênese do folículo. TLE1, TLE2, TLE3 e TLE4 reprimem a transativação mediada por LEF1 e CTNNB1. Regula a função do <i>enhancer</i> do receptor alfa de células T. PIASG antagoniza a ativação dependente e independente de Wnt por LEF1.	2
NFAT	NFAT (nuclear factors of activated T cells) transcription complex; um fator de transcrição de múltiplas subunidades que consiste em pelo menos 3 componentes de ligação ao DNA, incluindo homodímeros ou heterodímeros de proteínas da família FOS e JUN.	2
PAX5	Paired box gene 5; pode desempenhar um papel importante na diferenciação de células B, bem como no desenvolvimento neural e na espermatogênese. Envolvido na regulação do gene CD19, um gene alvo específico para o linfóide B.	2
SOX5	SRY-box containing gene 5; se liga especificamente a sequência de DNA 5'-AACAAAT-3'. Ativa a transcrição de COL2A1 e AGC1 <i>in vitro</i> .	2
VBP	No desenvolvimento embrionário de camundongos, a Vbp1 é expressa principalmente no sistema nervoso central, retina e fígado. Em tecidos de camundongo adulto, a Vbp1 é expressa de forma ubíqua. A análise <i>in situ</i> de vários tecidos adultos mostrou que na maioria dos tecidos os transcritos foram distribuídos uniformemente. No entanto, a Vbp1 é expressa em tipos específicos de células, como cérebro, olhos, rins e intestinos.	2
VMYB	MYB liga-se ao sítio e recruta o parceiro de ligação à acetilase H3K27, CBP, bem como os componentes principais de um importante complexo de transcrição leucemogênico. Além disso, a maioria dos <i>superenhancers</i> endógenos encontrados nas células de leucemia linfoblástica aguda são ocupados por MYB e CBP, o que sugere um papel geral para MYB na ativação do <i>superenhancer</i> .	2;5
ZEC	Proteína Zinc finger expressa em células embrionárias e em alguns órgãos no adulto. As proteínas Zinc finger (ZNFs) se ligam ao DNA e regulam a transcrição.	2
CDPCR3	CDP, CCAAT displacement protein; identificado pela primeira vez no ouriço-do-mar como um possível repressor de um gene histona H2b específico do esperma. Acredita-se que o CDP atue evitando a ligação de fatores CCAAT de ação positiva aos promotores, embora haja pouca evidência experimental para isso.	3
HOXD13	Homeobox (HOX); são genes que controlam a padronização, a diferenciação e a morfogênese durante o desenvolvimento. HOXD13 é um regulador-chave da morfogênese esquelética do autopode.	3

Continuação

TF	Principais funções	CNE
----	--------------------	-----

NRF2	Nfe2l2 - Nuclear factor, erythroid derived 2, like 2; ativador de transcrição que se liga a elementos de resposta antioxidante (ARE) nas regiões promotoras dos genes alvo. Importante para a regulação positiva coordenada de genes em resposta ao estresse oxidativo. Pode estar envolvido na ativação transcricional de genes do agrupamento da beta-globina por meio da mediação da atividade do <i>enhancer</i> do promotor da beta-globina.	3
TFE	MITF, TFE3, TFEB, e TFEC compreendem uma família de fatores de transcrição (MiT) que regula as principais vias de desenvolvimento em várias linhagens celulares. Como MYC, os membros MiT são fatores básicos <i>helix-loop-helix-leucine zipper</i>. Os membros do MiT compartilham homologia virtualmente perfeita em seus domínios de ligação ao DNA e se ligam a um <i>motif</i> comum de DNA. As translocações de TFE3 ocorrem em subconjuntos específicos de carcinomas de células renais humanas e em sarcomas alveolares de partes moles. A maioria dos genes lisossômicos exibe comportamento de transcrição coordenado e são regulados pelo fator de transcrição EB (TFEB).	3
GR	Nr3c1 - Nuclear receptor subfamily 3, group C, member 1; receptor de glicocorticóides (GC). Tem um duplo modo de ação: como fator de transcrição que se liga aos elementos de resposta aos glicocorticóides (GRE), tanto para o DNA nuclear quanto mitocondrial, e como modulador de outros fatores de transcrição. Afeta as respostas inflamatórias, proliferação e diferenciação celular nos tecidos-alvo. Envolvido na remodelação da cromatina.	5
KAI1	Membro da família BTB/POZ (Broad Complex, Tramtrak, Bric à brac/pox virus and zinc finger) de fatores de transcrição <i>zinc finger</i>, que estão fortemente implicados no desenvolvimento e no câncer. A maioria dessas proteínas funcionam como repressores transcricionais e são caracterizadas por um domínio POZ de interação proteína-proteína N-terminal e um domínio de ligação ao DNA C-terminal que consiste em um ou mais <i>Kruppel-like C₂H₂ zinc fingers</i>. Contribui para a repressão de genes-alvo da via de sinalização Wnt.	5
PAX4	Paired box gene 4; desempenha um papel importante na diferenciação e desenvolvimento das células beta das ilhotas pancreáticas. Repressor transcricional que compete com PAX6 na ligação a um elemento comum nos promotores de glucagon, insulina e somatostatina	5

Continuação

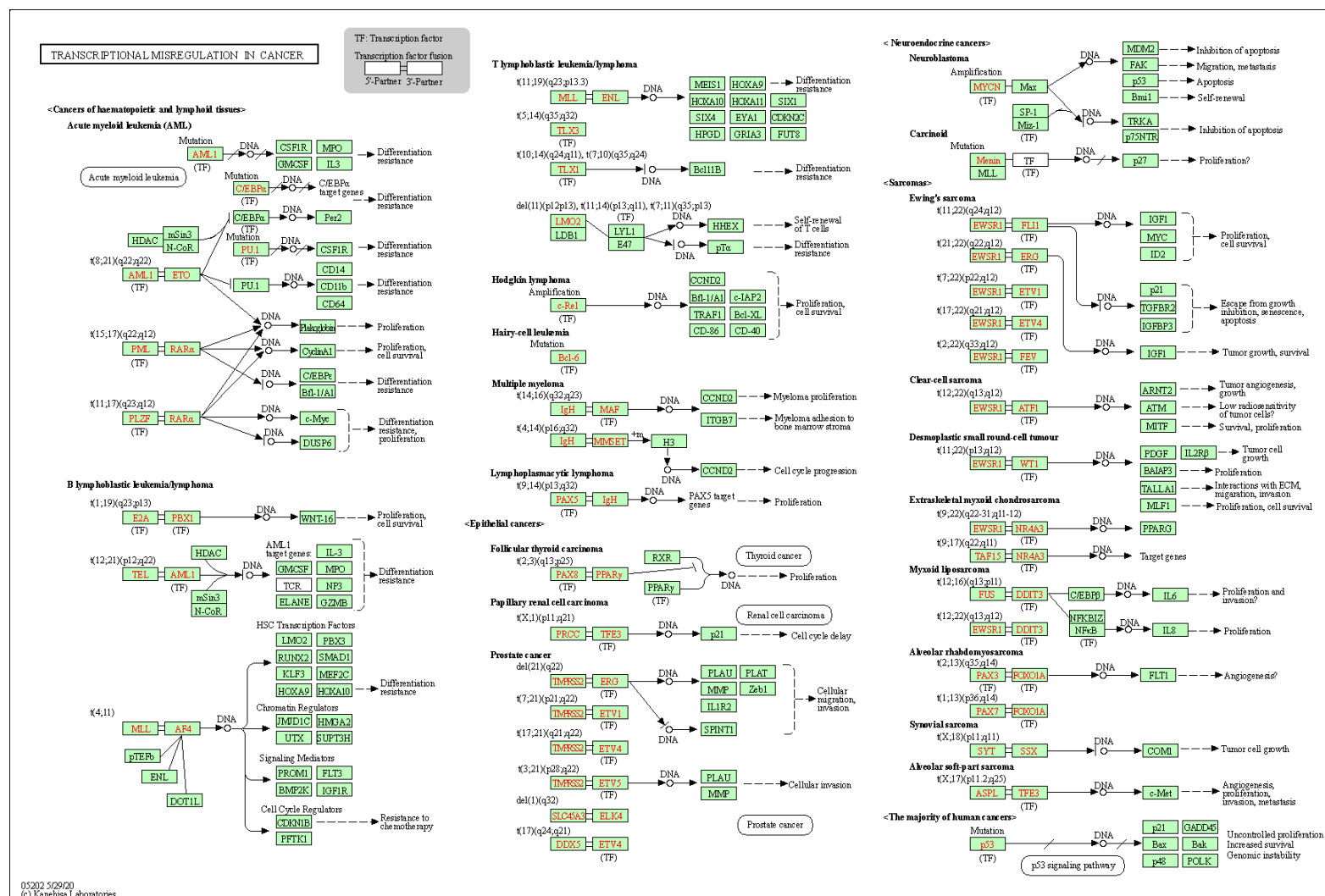
TF	Principais funções	CNE
TEF	O padrão de expressão gênico na glândula pituitária anterior do rato corresponde temporal e espacialmente ao início da expressão gênica do TSHB (hormônio estimulador da tireoide). Em contraste com esse padrão restrito de expressão durante a embriogênese, os transcritos de TEF apareceram em vários tecidos do organismo adulto. Também ativa o promotor de Telokin de maneira específica para músculo liso e dependente de cálcio.	5
CEBPA	Cebpa - CCAAT/enhancer binding protein (C/EBP), alpha; fator de transcrição que coordena a parada da proliferação e a diferenciação de progenitores mieloides, adipócitos, hepatócitos e células do pulmão e da placenta. Liga-se diretamente à sequência consenso de DNA 5'-T[TG]NNGNAA[TG]-3', atuando como um ativador em genes-alvo distintos. Durante a embriogênese inicial, desempenha funções essenciais e redundantes com CEBPB. Essencial para a transição de progenitores mieloides comuns (CMP) para progenitores de granulócitos/monócitos (GMP).	2;8
CEBPB	Cebpb - CCAAT/enhancer binding protein (C/EBP) beta; importante fator de transcrição que regula a expressão de genes envolvidos nas respostas imunológicas e inflamatórias. Também desempenha um papel significativo na adipogênese, bem como na via gliconeogênica, regeneração hepática e hematopoiese.	8
CEBPG	Cebpg - CAAT/enhancer binding protein (C/EBP) gamma; serve como um regulador e reservatório tampão contra as atividades de transcrição dos outros membros da família CEBP. CEBPG é o principal fator de transcrição responsável pela regulação dos níveis de antioxidantes e de reparo de DNA em células epiteliais brônquicas humanas normais.	2
ALX4	Aristaless-like homeobox 4; fator de transcrição envolvido no desenvolvimento do crânio, membros e folículos capilares.	5
P53	Trp53 - Transformation related protein 53; atua como supressor tumoral em muitos tipos de tumor; induz a parada do crescimento ou apoptose dependendo das circunstâncias fisiológicas e do tipo de célula. Envolvido na regulação do ciclo celular como um transativador que atua regulando negativamente a divisão celular, controlando um conjunto de genes necessários para esse processo. Um dos genes ativados é um inibidor de quinases dependentes de ciclina. A indução da apoptose parece ser mediada pela estimulação da expressão dos antígenos BAX e FAS ou pela repressão da expressão de Bcl-2.	8
T3R	Nuclear thyroid hormone (3,5,3'-triiodo-L-thyronine [T3]) receptor. Superfamília de receptores nucleares do hormônio tireoidiano (T3), e que age como um fator de transcrição.	8

Continuação

TF	Principais funções	CNE
----	--------------------	-----

CDP	Cux1 - Cut-like homeobox 1; deve estar envolvido no transporte retrógrado intra-Golgi	5
CLOX	Mammalian homeobox gene related to <i>Drosophila cut</i> ; codifica proteínas regulatórias de ligação a DNA diferencialmente expressas durante o desenvolvimento. As proteínas Clox são expressas em vários tecidos neurais e não neurais.	5
E47	E47 pertence à família de fatores de transcrição <i>helix-loop-helix</i> e formado pelo <i>splicing</i> alternativo do gene E2A. Tem sido associado à regulação da expressão do gene da imunoglobulina, mas também é crucial na diferenciação inicial de células B.	5

Via KEGG hsa05202, de desregulação transcricional em cânceres, relacionada aos sítios de ligação para fatores de transcrição do CNE1.



8.5 Apêndice 5

Sequências estendidas (1 kb) correspondentes ao CNE4 em treze espécies distintas de amniotos (humano à galinha).

```
>Human 14 dna:chromosome chromosome:GRCh38:14:58777185:58778186:1
TTGAGAAGACAAAGGCGCTTTGTATGTATTATTCAGGTGTGTATTTAAAGCCAAGGCTGA
TCCCAATAAAGTACAATAACCTAAAAGATTGATACTTTACAGACTGAAGAAAAATTTTC
CACTTTAATCCCAAGCCCCAGACCTACCTCTCCCTTGAAAAAGCTTAGAAATGTTGTCT
CTAGAGCTGGGCACTGAGAAGGACGGAAGATATGCTCAGCTCAGAGGACCTAAATTCTGA
TTAGCTGCAGAGAAAATGACCAGCTCTGCAGCTTCTGGGGCACATTTGAGGGGCATCTCT
CCAGGAGGACATAGCACCAAGCATCATTTGATTTACTGATCTTCCCGCTTCCACCCCTGC
CCAGGGTTCCCATGGCTGCCGTGAACTGGGAAGTGGAAAGGCAGCAATTGGGCTGTGAGA
GCCTCCTGGGTCTCCTACTGCTGACTTCCTCCTATAAATCAGCCACAGACAGCTGCATCC
ACTCGCACCATGGGATAGAAAAATAGACCCTGTTGCTAGACAGCAACACAGGCTCCACAA
ACACCATGGGGTCAGAGGGCAAGGGGTCTTAAGCCAGGCACTAGCCAAGCAGGCCCAGCA
GCCCCGCAGATCCATTTGTTTATTTTACAGTGGCATTGCCTGTCTGTACTACACGAAGG
AGGGAGACTCCCTTGCCCCCTCTGCTCTGCATGTGTAGGCCACTTGCTGCTCACCACCGGA
AAGTCCCCTGACCTGCTGGCTAGTTTATTTATTTTAGATGCGTGTAGACTGTCTGCTAG
GTCTAAGTTTTGAAGCTGATGGGGAAGTCATCAAGTCTAGTACAGAGTATTTAGCAATTG
AATTTTATCTGGTCCATTGAATATCTGATTGAACAAGAAAGGAAACCTTGTACTTTGTA
CTGACGTATTCACAACCTCCCCCTCTTCAGCACCTCCCTCCAGGGCCTTCCTTTCAAATC
AGCATCTGACCCACCAAGAGAAATAGGTTTATGATGGCTGTG
```

```
>Chimpanzee 14 dna:chromosome
chromosome:Pan_tro_3.0:14:43034200:43035302:1
TTTCTAGCTGGTCACACCTAGAAAGGTGGTCACAGGTGGTCACCTGTGTCTTGAGAAGA
CACAGGCACTTTGTATGTATTATTCAGGTGTGTATTTAAAGCCAAGGCTGATCCCAATAA
AGTACAGTAACCTAAAAGATTGATACTTTACAGACTGAAGAAAAATTTTCCACTTTAAT
CCCAAGCCCCAGACCTACCTCTCCCTTGAAAAAGCTTAGAAATGATGTCTCTAGAGCTG
GGCGCTGAGAAGGACAGAAGATATGCTCAGCTAAGAGGACCTAAATTCTGATTAGCTGCA
AAGAAAAATGACCAGCTCTGCAGCTTCTGGGGCACATTTGAGGGGCATCTCTCCAGGAGGA
CATAGCACCAAGCATCACTTGATTTACTGATCTTCCCGCTTCCACCCCTGCCCGGAGTTC
CCATGGCTGCCGTGAACTGGGAAGTGGAAAGGCAGCAATTGGGCTGTGAGAGCCTCCTGG
GTCTCCTACTGCTGACTTCCTCCTATAAATCAGCCACAGACAGCTGCATCCACTCGCACC
ATGGGATAGAAAAATAGACCCTGTTGCTAGACAGCAACACAGGCTCCACAAACACCATGG
GGTCAGAGGGCAAGGGGTCTTAAGCCAGGCACTAGCCAAGCAGGCCCAGCAGCCCCGCAG
ATCCATTTGTTTATTTTACAGTGGCATTGCCTGTCTGTACTACACAAAGGAGGGGAGACT
CCCTTGCCCCCTCTGCTCTGCATGTGTAGGCCACTTGCTGCTCACCACCGGAAAGTCCCGC
TGACCTGCTGGCTAGTTTATTTATTTTAGATGCGTGTAGACTGTCTGCTAGGTCTAAGTT
TTGAAGCTGATGGGGAAGTCATCAAGTCTAGTACAGAGTATTTAGCAATTGAATTTTATC
TGGTCCATTGAATATCTGATTGAACAAGAAAGGAAACCTTGTACTTTGCACTGACGTAT
TCACAACCTCCCCCTCTTCAGCACCTCCCTCCAGGGCCTTCCTTTCAAATCAGCATCTGA
CCCACCAAGAGAAATAGGCTTATGATGGCTGTGAGTTTGAGCTCTTCTTCTTCTTTGAC
TTGAATCAATCTCTTTGCAAATA
```

```
>Dog 8 dna:chromosome chromosome:CanFam3.1:8:34074699:34075700:1
CTAAGTCAGGGCTGCTTCTTTGAAGGCACAGAAATGTATAAGAGTGTTACTTTAATAAAC
TGAAGAAAAAAGATGTCCATTTTTATCCCACTCCTACACCCTCCCTGGAAAAAATATC
TAGAAATAACATCTGTAGAGCGGGACACGAAGAAGGATGGAGAATGTGCATGACAAACAG
GATCTAAATCCAATCAGTTGCAGAGAAAATGACTAGCCTGATAGCCTTCTGGGGCTTCTT
TGGGGGTGGGGGAGTGTTTCCCTGAAGGATGGACCCCCCAGGACCCCCCAGCTTACTGAA
TACCCACCCCTGAGTTCCCAGGGCTACTGTGAACGGGGAAGTGAAAATGAAGATGGGGG
CTGTGAGAGCCTTCTGGGGCTCCTACTGGTGACTTCCTCCTATAAATCAGCCATAGACAG
CTGCATCCACTTGCGCCATGGGATAGAAAAATAGACCCTGTTGCTAGACAGGGATACAGG
```


CTCCACAAACATTGAGGGATCAGAGGGCAAGGGGCTTAGCCCAAGCGCTGGCCAAACAGC
 CCCACAGCCCCACAGATCCATTTGTTTATTTTCACAGTGGCATTGCCTGTCTGTGCCACA
 GGAAGGAGGAAGACCCCCCTGCCCCACTCTGCTCTGCATATGTAGACCTACTTATGCTCA
 CCGCTGCAGAGGCCTGCAGTGAGGCCTGCTGACCTGCTAGCTAATTTATTTATTTGAGTT
 GCACACAGTTTCAAAGCTAATGGGGAAGTGCTCAGGCGGTGGAATATTCAGCCATTGGAT
 TTTATTAGGTCTGCTGAATATCTGATTGAGCAAGAAAGGGAAGCCTTGTACTGACATACC
 TGCCCTCCTCTCTTCCCAACAAGAACCTACCATCCCCACCCCTGATTCTGTCATCAGGG
 TCTTCACTTTCCAGTTGGCATCTGACTCACCCAAATAAACAGGCTATGGCAGCTGCGAGG
 TTAAGCTTCCTTCCTTTGACTTGAATCAATGTCTCCTAAAAT

>Cow 10 dna:primary_assembly primary_assembly:ARS-

UCD1.2:10:71048740:71049741:1

AAAGTCAGGGCTGATCCCAATAAGGCACAGAAGCATATAAGAATGCTACTTTAATAAAAC
 TGAAGAAAAAAAATTTATTTTCATCCCAAGCCCCACCCCTCATGCTCCCAAGAAGAAGC
 CTAGAAATAACATCTCTGGAGCTAAGGGACCGAGAGGGATGGAAAATGTGGGTGGCAAAC
 AGGATCTATATCCAGTTGGCTGCAGAGAAAATGACCAGCTGGGCAACCTTCTGGGGTTTC
 TGTAGGGGGTGTCTCCCTGAAAGATAAAGCTCCATGCACCCCTGGTGTAAATGATCACAC
 TGGCAAACCCCTCACCTAGTCCTCATGGCTACCAGGAGCTGGGGAATTGAGATGAAGACT
 TTGGGCTGTGAGCGTCCTGACCCTCATACTGGTGACTTCCTCCTGTAAATCAGCCCATGA
 CAGCCGCATCCACTTGCACCGCCCGAGGAGAGGAAAATAGACCCTGTTGCTAGACAGCAA
 CACAGGCTCCACAAACATTGTAGGATCAGAGGGCAGGGGCTCAGCCCAGGCACTAGGCAA
 ACAGGCCTGCAACCCCAAGCATGCATGTTTATTTTCACAGTGGCAATGCCTGTCTGTGCCA
 CAGGAAGGAGGAAGACCCCCCTGCCCCAGACCCCTTTTCTCTGCATATGTAGACCACTT
 ACTGGTCATCACACAGAGGCCACAACCTGCAGTGAGGTCCACTGACCTGCTGGCTAATTT
 ATTTATTTTAGTTGCACATAGAAGCTGCCTGCTAGGTATGAGTTTCCAAGCTGGTGGGAAG
 TGCATTGTAGTCTCATAACAGAAATATTCAGCGATTGAATTCTATTTGAATATCTGATGAAA
 CAAGAAAGGCTAACCTTGTGCTGACATGTTACACCCCTCACTCCCCATGCACCTGCA
 GTCCCCCAGCTCTCACACTTCCATCCAGGGCCTTCACTTTGCACTCAGTATCTGATCCAC
 CCAACATGATAAACAGGCCATGGCGGCTGTGAGGCTGAGCTT

> Sheep 7 dna:primary_assembly

primary_assembly:Oar_rambouillet_v1.0:7:73579129:73580130:1

CTCCCAGGAAGAAGCCTAGAAATAACATCTCTGGAGCTAAGGGACTGAGAGGGATGAAAA
 ATGTGCATGGCAAACAGGATCTATATCCAGTTGGCTGCAGAGAAAATGAGCAGCTGGGCA
 ACCTTCTGGGGTTTTTGTAGTGGGTGTCTCCGTGAAAGATAAAGCTCCATGCACCCCTG
 ATGTAATGATCACACTGGCAAACCCCAACCTAGTCCTCATGGCTACCAGGAGTCAGGAA
 GTTGAGATGAAGACTTTGGGCTGTGAGCCTCCTGACACTCATACTGGTAACTTCCTCCTG
 TAAATCAGCCCTGACAGCTGCATCCACTTGACCGCCCTAGGAGAGAAAAATAGACCCT
 GTTGCTAGACAGCAACACAGGCTCCATAAACATCGTGGGATCAGAGGGCAGGGGCTCAGC
 CCAGGCACTAGCCAAACAGGCCTGCAGCCCCACCATGCATCTGTTTATTTTCACAGTGGC
 AATGCCTGTCTGTGCCACAGGAAGGAGGAAGACCCCCCAGATCCCTTTTCTCTGCATATG
 TAGACCACTTACTGGTCATCACCGCAGAGGCCACAACCTGCAGTGAGGCCCACTGACCTGC
 TGGCTAATTTATTTATTTTAGTTGCACATAGACTGCCTGCTAGGTATGAGTTTCCAAGCT
 GGTGGGAAGTGCAATTGCAGTCTCATATAGAATATTCAGCAATTGAATTCTATTTGGTCTG
 CTGAATATCTGATGAAGCAAGAAAGGCTAACTTTGTGCTGACATGTTACACCCCTCAT
 TCCCCATGCACCTGCACTCTCCAGCTCTCAGACTTCCATCCAGGGTCTTCACTTCGCA
 TTCAGCATCTGATCCACCAACATGATAAACAGGCCGTGGCAGCTGTGAGGCTGAGCTTT
 CCTTCCTTTGACTTGAATCCAGCTCCCCGAAAATAAAAGCAATTTTCCAACATTCCATTA
 TAAAAATTTCAGTATGTTAAAGTAGAAAGATTTTGTACAGTG

>Goat 10 dna:chromosome chromosome:ARS1:10:32171403:32172404:1

ACTGTACAAAATCTTTCTACTTTTAACATACTGAAATTTTATAATAGAATGTTGGAAAAT
 TGCTTTTATTTTAGAGGAGCTGGATTCAAGTCAAAGGAAGGAAAGCTCAGCCTCACAGCT
 GCCACGGCCTGTTTATCATGTTGGGTGGATCAGATGCTGAATGCAAAGTGAAGACCCTGG
 ATGGAAGTGTGAGAGCTGGGAGACTGCAGGGTGCATGGGGAGTGAGGGGGTGTGAACATG
 TCAGCACAAAGTTAGCCTTTCTTGCTTCATCAGATATTCAGCAGACCAAATAGAATTCAA

TTGCTGAATATTCTATATGAGACTGCAATGCACTTCCCACCAGCTTGAAACTCATACCT
 AGCAGGCAGTCTATGTGCAACTAAAATAAAATAAATTAGCCAGCAGGTCAGTGGGCCTCAC
 TGCAGTTGTGGCCTCTGCAGTGATGACCAGTAAGTGGTCTACATATGCAGAGAAAAGGGA
 TCTGGGGGGGTCTTCCTCCTTCTGTGGCACAGACAGGCATTGCCACTGTGAAAATAAAC
 AGATGCATGGTGGGGCTGCAGGCCTGTTTGGCTAGTGCCTGGGCTGAGCCCCTGCCCTCT
 GATCCACGATGTTTATGGAGCCTGTGTTGCTGTCTAGCAACAGGGTCTATTTTTCTCTC
 CTAGGGCGGTGCAAGTGGATGCAGCTGTCACGGGCTGATTTACAGGAGGAAGTTACCAGT
 ATGAGTGTGAGGAGGCTCACAGCCCAAAGTCTTCATCTCAACTCCCTGGCTCCTGGTAGC
 CATGAGGACTAGGGTGGGGGTTTGCCAGTGTGATCATTACATCAGGGGGTGCATGGAGCT
 TTATCTTTCACGGAGACACCCACTACAGAAACCCAGAAAGGTTGCCAGCTGGTCATTTT
 CTCTGCAGCCAACTGGATATAGATCCTGTTTGCCACGCACATTTTCCATCCCTCTCAGTC
 CCTTAGCTCCAGAGATGTTATTTCTAGGCTTCTTCATGGGAG

>Pig 1 dna:chromosome chromosome:Sscrofa11.1:187854618:187855619:1
 GTGCATGACTCTTCAATGTCAGGACTGATCCCAGTAAGGCAGAGAAATATTATGCGAGTG
 TTAGTTTAATAAACCGAAGAAAAAGATTCCATTTTCATTCCAAGCCGCACCCAGTATGC
 TCCTTGGAAGAAAGCTTGGAATAACATCTCTGGAGCTAGGGACTGAGAAGGGTGGAAAAT
 ACCTGGTGAAGATCTAAATCCAATGAGCTGCAGAGAAGATGACCAGCCAGGCAGCCTTCT
 GGGGTTTCCGTGGGGTGCATCTCCCTGAGGATAAAGCTCCCAGCATCACCCGGTTTCCTG
 ACTGGCGTCCCCCACCCTGACCCTGGAGTCTCCCCTGGCTGCCCCAACCTGGGAGTAGA
 ACTGAAAACACTGGGTGGTGAGCCTCCCTGTGCTCCTACTGGTGACTTCCTGCTATAAAC
 CTGCAGCAGACAGCGCCCTCCACTTGCACCACCTTGGGAAAGAAAAATAGACCCTATTGC
 TAGACAGCAACGCAGGCTCCACAAACATCGTGGGATCAGAGGGCAGGGGCTCAGCCAGGG
 CACTAGCAAAACTGGCCCACAGCCCCACCACCGTCTGTTTATTTTCACAGTAGCATTGCC
 TGTCTGTACCTCAGGAAGGAGAAAGACCTCCATCCCCACCCCTGCCCCAGCCCTCCTCT
 GCTCTGCACGTGTAGATCACTTAGGGCCTACTGCTGCAAAGGCCACAACGGCGCTGACCT
 GCTGGCTCATTTATTTATTTTAGTTGCACGTAGACTGTCTTCTGGGTCTAAGTTACAAAG
 CTGGTGGGGAAGCACACAGTCTCCTACAAGATATTAGCCATTGGGTCTATTTGGTCTG
 CAGAAAATCCTATTTAATCAAGAAAGGGAGAGAGTTCCCATTGTGGAGCAGTAGGTTAAG
 AATATGACTGCAGTGGCTTAGGTGGCTGTGGAGTTTCCAGGAACTTCCATATTCTCTCA
 ATGCAGCCAAAAAAGGAAACTTGCACTGACATGTAC

>Horse 24 dna:primary_assembly
 primary_assembly:EquCab3.0:24:6261570:6262571:1
 GTACCCTTCTGTGTGTGTTTACAGTCAGGGCTGATCCCCTAAAACACAGAAACGTGTAA
 GAGTGTACTTCAATAAACCGAAGAAAAATTTCCGTTTTATCCCAAACCCACCCTTC
 CACCCTCCCTTGAAAAAGCCTAGAAATAGCTTTTCTAGAGCTCAGAACTAAGAGGGATGG
 AAATGTGCTCACCAAACAGGATCTAAATTCATCAGCGGTGGAGAAAACAACCAGCCTGG
 CAGCCTTCGGGGGCTTCTTTGGGGCGCGTCTCCCAGAAGGATAACGCCCGGCACCATCCA
 GCTTACTGGGCACCCACACCCCGCCCGGAGTTCTCGTGGCTGCGGTGAACCGGGAAG
 TGGAAGTGAAGCCATGAGGCTGTGAGAGCCTCCGGGGCTCCTACTGGTGACTTCCTCCTA
 TAAATCAGCCCCGGACAGCTGTGTCCACTTACGCCATGGGAGAGAAAAATAGACCTTGCT
 GCTAGACAGCAGTGCAGGCTCCACAAACATCGGGGATCAGAGGGCAGGGGGCTTAGCCAG
 GCACTAGGCAAACAGGCCCCAGCCCCACGGATCCGTTTGTGTTTTCACAGTGGCATTG
 CCTGTCTGTGCCACAGGAAGGAGGGAGATGCCCCAGCCCCCTCTGCTCTGCATGTGCAGA
 CCACTTACTGCTCACCGCCGCAAAGGGCGTGACAGTAGCGGGCCCACTGACCTGCGGGC
 TCGTGTATTCAATTTAGCTGCACATAGACTGTCTGCGAGATCTAAGTTTCCAAGCTAACG
 GCGAAGCACGCGGTCTAGCACAGAACGTTGGGGTGGATTTTACTCGGTTTGCCGAATATT
 TGATGGAGCAAGAAAGGGAAACCTTGTACTGACTTATTCACACCTCGTGCCTCCACGCAT
 CCCTGCCACCTCCTCCCTACACTTCTATCCAGGGTCTCACTTTCCAGTCAGCATCTGAC
 CCACACAGATAAACAGGCTGTGCCAGCTGTGAGGCTGACCAC

```
>Cat B3 dna:chromosome  
chromosome:Felis_catus_9.0:B3:107913493:107914494:1  
TGGAAAGACTCTGGATTCTTTGTATACTTACGTACGTATTTGAAAGGCAGGACTGCTTCC  
ATTAAGGTACACAAATGTAGAAGAGTGTTACTTTAATAAACTGAAGAAAAAAGATGTCCA  
TTTTTACCCCAAGCCCCACCTCTAAACCCCTCCCCTGAAGGGGGAAATAGCATCTCTAGAG  
CTAGAGACGAAGAACAATGAAAAATGTGCATGGCAAACAGTATCTAAATCCAATCAGTTG  
CAGAGAAAATGACTAGTCTGACAGCCTTCTGGGGCTTCTTTGGGAGTGAGGGGTGCCTCC  
CTGAAGGATAAAGCCCCAACACCCCTCAGTTCAGTACGTACCCACCCCTGAGTTCCCA  
TGGCTGCTTGTGAACTGGGAAGTGAAAATGAAGATGGGGGCTGTGGGGCTCCTACTGGTG  
ACTTCCTCCTATAAATCAGCCACAGACAGCTGCATCCACCAGTGCCATGGGATAGAAAAA  
TAGACCCTATTGCTAGACAGTAACATAGGCTCCACAAACATTGTGGGATCAGAGGGCAGG  
GGGCTTAGCCCAAGCATAGCCAAACAGCCCCCTCAACCCACAGATCCATTTGTTTATTTT  
CACAGCAGCTTTGCCTGTCTGTGCCACAGGAAGGAGGAAGAAGACCCCCCCCCCCCCCTT  
GCTCCACTCTGCGCTGCACATGTAGACCACTTACTGCTCACCCTGCAAAGACCTGCTAG  
CTAATTTATTTGTTTGAGTTGCCCACAGACTGTCTGCTAGGTCTAAGTTTCAAAGCTAAT  
GGGAAGTGCTCAGTCTATGGAATATTCAGCTGTTGGATTTCAATTTGGTCTGCCGAATAC  
CGGATTGAGCAAGAAAGGGGAAGCCTTGTCCTGATATATTCGGCACCCCTGCTCTTCCCAC  
CATGCACCCTACATTCCCCCACCCCTGATACTTTTCATCGGGGTTTTTCACTTTCCAGCTGG  
CCTCTGACCCACCCAGAGAAACCGGCTGTGG
```

```
>Rat 6 dna:chromosome chromosome:Rnor_6.0:6:94149174:94150175:1  
AGAGCCAGCACTGAGCCCACCAAGGTGTAATGCCGAAGAACGAAAGCACATGAAAGAAAT  
AGCTTCCATTTTTTTATCCCAAGTCCCACCCTACACCCTCCTTTGAGAAAGGCTAGAAATA  
GCATCTCCCAAGCCTGGGACTAGAAAGGATGGGAAATGTGTGTGGTGCATGGGATCTAAA  
TCTCTCCAGGGGCTCACTGAGGGGCATCTCCCCGGGATGATGGAGTAGCAGCACACACT  
GCCTGTTTTTACTGTTCCCCAAATCCAGAGTTCCCATGCCTGTCATTCTGGACTGAGAGAG  
CCTTCTGGGTCTCCTGGGTAGTGGCTTCTTCTATAAATCAACTCCAGACAGCTGTGGCC  
ACTTGCATCCTGGGATAGAAAAATAGACCCTGCTGCCAGACATCAACCTAGGCTCCACAA  
ACATGTATTTATCAGAGGGCAGGGGCACCGAGCCCAGGCAGGCACTAGCCAAGCAGGCCC  
AACAGCCGGGTAGCTCCATTTGTTTATTTTCTCAGTGTTACTGCTGTCTCTACCACAAGA  
AGGGGGGCGCCCTCCGCCCCCTGCTGGGCTGATGTGGAGGCCACTTACTCTTTACCACCG  
GAAAGGCCTGCCAACTGCTGGCTGATTTATTTATTTTAGTTGCACATAGACTGTCTGCTG  
GGTCTAAGTTTGGGAAGCCAACAGGAAGTCATCCAGTCTAACACCAAGTGTTCAATGCTTG  
TGTTCTGTTTGGGCAAACATATTAGACTGAGCAAGAAAGGGAAGCCTTCCGTTGGTGTAT  
TCCCAGCTCATTCCCCTCCACCCCTCCACCCCTCACCTCCCTTTATTCCCCCTCCCTG  
CCCCAACCCCTCCTGCCTTGCCATCTAGCACAGGCCACCTGTGAGGTTGACTTCTGCCT  
TTGACTTCAGGGGCTCTCTTCTAAAATACGTTTTTCAACATTTTTTTTCCCTGTGAAAAT  
GTAATAGCAGTGTTGAGGGGGGGGGATCTCTACCTAGAATCT
```

```
>Mouse 12 dna:chromosome chromosome:GRCm38:12:71437819:71438820:1  
GTGCTACATTTAGGTATATGTTTGGCCAGAGCCCACGAAGGTGGTAATGCTGAAGTACTC  
TAACAAACAAAAGAAATATCTCCCAAGTTTTTTGTTCCAAGTTCCACCCTACACCCTCCTTT  
GAGAAGGGCTAGAAACAACATCTCCCAAGCTGGAGAACTATAAAGGATGGGAAATGTGTG  
TGGTGCATGGGATCTAAATCTCTCCAGGGGCTCACTGAGGGTCGTCTTCCCAGGATGAT  
GGAGTAGCAACACACACTGCCTGGTTTACTGTTCCCCAAACCCAGAGTTCCCATGCCTGT  
CATTTCTGGGCTGAGAGAGCCTTCTGGGTCTCCTGGGTGGTGGCTTCTTCTATAAATCAA  
CTCCAGACAGCTGCAGCCACTTGCATCCTGGGATAGAAAAATAGACCCTGTTGCTAGACA  
TCAACCTAGGCTCCACAAACATGGCGGGATCACAGGGCAGGGGGCACCTAGCCCAGGCAG  
GCACTAGGCAAGCAGGCCCACCAGCCGGGTAGCTCCGTTTGTATTTTCTCAGTGGTAT  
TGCTGTCTCTACCAGGAGAAGGGGGGCACCCCTCCGCCCCCTGCTGGGCTGATGTGGAGGC  
CACTTACTCTTTACCACCGGAAAGGCCTGCTGACTGCTGGCTGATTTATTTATTTTAGTT  
GCACATAGACTGTCTGCTAGGTCTAAGTTTGAAGCCAACAGGAAGTCATCCAATCTAAC  
ACCAAGTGTTTCAAGTGTGTTCTGTTTGGGCAAACACATATTAGATTGCGCAAGAAAG  
GGAAGTCTTCTGTTGGTATATTCCCAACTCCTTCCCCTCCACCTCTCCCATCCCTCACC  
CCCCCACCCCTGCATTGTCAATTTACCTCAGGCCAGCTGTGAGATTGATTTCTACATTTG  
ACAACAACCACACTCTTCCAAAATAAGTTTTTCAACATTTTTTTTTTCCCTGTGAAAATGTA
```

ACAGCAGTGTTTCAGGGGAATCCCCTACCTAGATTCTACCGT

>Turkey 5 dna:chromosome

chromosome:Turkey_2.01:5:57458957:57459958:1

CTTCCAGAGAAAAGCTTTGCAGTTCTGGGAAGTCTGGGGAGGGATTTCAGAGAGGGCATTG
GGCTCTCTCAACAAGCAAACAGGAGGCCAGACTGGGAGACCAGAGGATCCTAGATAGGGA
GAAATGAAACAAGAGGTTTTTACCACCGTGGGACAAACCATGGCTATCAGTGTTTCCTGA
CAATAACTTTGTTTGGACAAGGCTCGTGGTTGGGTTGTGTCAGGAGAAGAAGGAAGAGCAG
CTGAAGCATTAACTGTCTCTTGCCTGTGCTGGTCCCTTGGAAAGTGCAGACTTGGCAG
CAGCTAGCAGGATGCTGAATGAAAGGGACATTTAGAGCTCAGGCACGGGACAAACGATGC
TAATGATAAATCCATCCCCCGACGCTGCGCGGACACGGATGTCCACAGCTAAAATAAACA
AATGGTCCGGGCGAGCTGACAGGCTTGTTTAGATTCTGCTCGGGATAAATCGCAGTGCCC
TTTGATCCCGTGATGTTTTTGGCGCCTGTGTTGCTGTCTAGCAGCAGGGTATATTTTTCT
GTCCCAGAAAGCAAGTGGATGCAGCTGCCTATAGCTGATTTATAGTAGGAAGCGATTAC
AGGAAATCAAAGCTTCTCACACAGTCAGAATCGCTGTCTTTTCACCTCCACCATCTATGC
AATGATTAGAGCTCTAGGGACTTTGGTGGGGAAGATCAGCAAAGCGTTCTGTGCTTAGCA
TAAATCATTTTTCTGCTGTAACTCATGAAGGTTGCTCGCAATGGGCTGCACCTGCCTGTAC
ATCCTGGTGCACAAGCACTGACAGCGTGTTTGTTCATCTCAATCCAGAGGTTTGAAATC
CACATTTCTGTGCTCGTTGCAGCAATGTGCTGGGCAGCAAAATGAAGGCAATATGCAAGA
ATACCAAATACCATATCAAATGTGCTTATCACACACGCACTGCATTACCAGACGTAATT
TAGGAGCTGGTTACTGCTTGCAGTTTCTGATAACTGCTCACA

>Chicken 5 dna:chromosome chromosome:GRCg6a:5:55230575:55231576:1

CAAACAGGAGGCCAGGCTGGGAGACCAGAGGAGCCTGGATGGGCAGAAAGGAAACAAGAT
GTTTTTACCACCCTGGGACTAACCTTAGCTATCGATGCCTCCCAACAACAACCTTGTTTGG
GCAAGGCTCGTGTTGGGTTGTGTCAGGAGAAGGAAGAGCAGCCAAAGCATTAACACCTT
CCTCTTGCCGTGCTGGTCCCTTGGAAAGTGTCTCTGGTGCAGACTTGGCAGCGGCCAGCA
GGATGCTGAATGAAAGGGACATTTAGAGCTCAGGCGCAGGACAAACGATGCTAATGATAA
ATCCATCCCCCGACGCTGCGCCGACACGGATGTCCACAGCTAAAATAAACAATGGTCCG
GGCGAGCTGACAGGCTTGTTTAGATTCTGCTCGGGATAAATCGCAGTGCCCTTTGATCCC
GTGATGTTTTTTGGCGCCTGTGTTGCTGTCTAGCAGCAGGGTATATTTTTCTGTCCCGAGG
AGCAAGTGGATGCAGCTGCCTATAGCTGATTTATAGTAGGAAGCGATTTGCAGGAAATCA
AAGCTTCTCACACAGTCAGAATTGCTGTCTTTTCACCTCCACCATCGATGCAATGATTAG
AGCTCTAGGGACTTTGGTGGGGAAGATCAGCAAAGCGTTCTGTGCTGAGCATAAATCACT
TTCTGCTGTGACTACGAAGGATGCTCGCAGTGGGCTGCACCTGTCTGTACATCCTGGTG
TGCAGGCACTGACAGTATGTCCATTCAATCTCAATCCAGAGATTTGAAACCCATGTTTCT
GTGCTCTCTGCAATGTGCTGGACAGCAAAATGACGGCAGTAGGCAAGAATACCAAATACC
ATACCAAACGTGCTTATCACACACACACTGCATTACCAGACATGTAGTTTAGGAGCTGG
TTACTGCTTGCAGTTTCTGATAACTGCTCACATCAGCTCAGAAGCTGATGAAGCACGTGT
TTATTACTGCAAAAACCTTGAGACCCCGTGGGACGTTTCAGCAA

8.6 Apêndice 6

Análise de conservação do *core* do CNE em treze espécies distintas de amniotos (humano à galinha).

```

::: block no. 1, score=891104.0, strand:+, base position:413-626
Mouse      CTG---AGAGAGCCTTCTGGGTCTCCTGGGTGGTGGCTTCTTCTATAAAATCAACTCCAG
Rat        CTG---AGAGAGCCTTCTGGGTCTCCTGGGTAGTGGCTTCTTCTATAAAATCAACTCCAG
Turkey    CTG---TGTGAGAAGCTTTGATTTCCTGTG-AATCGCTTCCTACTATAAAATCAGCTATAG
Chicken    CTG---TGTGAGAAGCTTTGATTTCCTGCA-AATCGCTTCCTACTATAAAATCAGCTATAG
Pig        CTGGGTGGTGAGCCTCCCTGTGCTCCTACT-GGTGACTTCCTGCTATAAACTGCAGCAG
Dog        CTG---TGAGAGCCTTCTGGGGCTCCTACT-GGTGACTTCCTCCTATAAAATCAGCCATAG
Cow        TTGGGCTGTGAGCGTCTGACCCCTACTACT-GGTGACTTCCTCCTGTAAATCAGCCCATG
Cat        CTG---TGAGAGCCTCCTGGGGCTCCTACT-GGTGACTTCCTCCTATAAAATCAGCCACAG
Sheep      TTGGGCTGTGAGCCTCCTGACACTCATACT-GGTAACCTCCTCCTGTAAATCAGCCCCTG
Goat       TTGGGCTGTGAGCCTCCTGACACTCATACT-GGTAACCTCCTCCTGTAAATCAGCCCCTG
Horse      CTG---TGAGAGCCTCC-GGGGCTCCTACT-GGTGACTTCCTCCTATAAAATCAGCCCCTG
Chimpanzee CTG---TGAGAGCCTCCTGGGTCTCCTACT-GCTGACTTCCTCCTATAAAATCAGCCACAG
Human      CTG---TGAGAGCCTCCTGGGTCTCCTACT-GCTGACTTCCTCCTATAAAATCAGCCACAG 468
          **      *  ***      ** *      *  ***** ** ***** *  *  *

Mouse      ACAGCTGCAGCCACTTG---CATCCTGGGATAGAAAAATAGACCCTGTTGCTAGACATCA
Rat        ACAGCTGTGGCCACTTG---CATCCTGGGATAGAAAAATAGACCCTGCTGCCAGACATCA
Turkey    GCAGCTGCATCCACTTG---CTTCTCGGGACAGAAAAATATACCCTGCTGCTAGACAGCA
Chicken    GCAGCTGCATCCACTTG---CTCCTCGGGACAGAAAAATATACCCTGCTGCTAGACAGCA
Pig        ACAGCGCCTCCACTTGCACCCACTTGGGAAAGAAAAATAGACCCTATTGCTAGACAGCA
Dog        ACAGCTGCATCCACTTG---CGCCATGGGATAGAAAAATAGACCCTGTTGCTAGACAGGG
Cow        ACAGCCGCATCCACTTGCACCGCCCGAGGAGAGGAAAAATAGACCCTGTTGCTAGACAGCA
Cat        ACAGCTGCATCCACCAG---TGCCATGGGATAGAAAAATAGACCCTATTGCTAGACAGTA
Sheep      ACAGCTGCATCCACTTGCACCGCCCTAGGAGAGAAAAATAGACCCTGTTGCTAGACAGCA
Goat       ACAGCTGCATCCACTTGCACCGCCCTAGGAGAGAAAAATAGACCCTGTTGCTAGACAGCA
Horse      ACAGCTGTGTCCACTTA---CGCCATGGGAGAGAAAAATAGACCCTGCTGCTAGACAGCA
Chimpanzee ACAGCTGCATCCACTCG---CACCATGGGATAGAAAAATAGACCCTGTTGCTAGACAGCA
Human      ACAGCTGCATCCACTCG---CACCATGGGATAGAAAAATAGACCCTGTTGCTAGACAGCA 525
          ****      ****      *  *** ** ***** ** *  *** *****

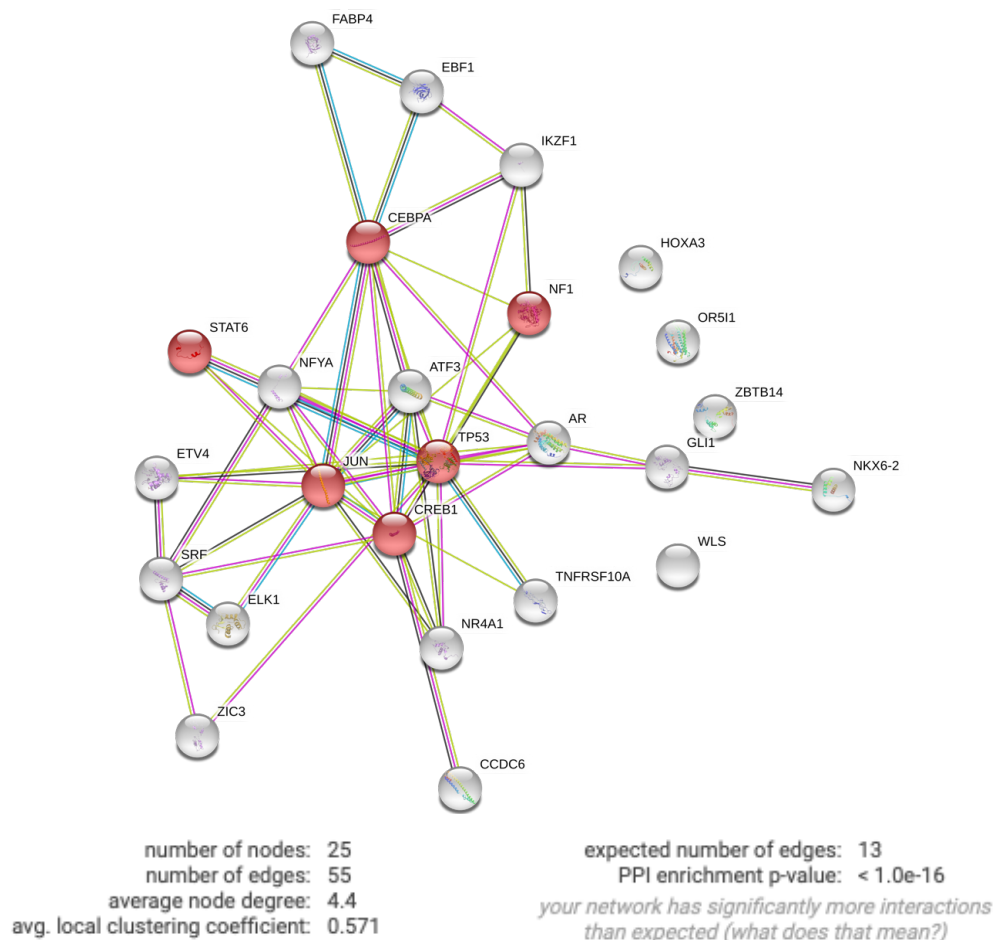
Mouse      ACCTAGGCTCCACAAACATGGCGGGATCACAGGGCAGGGGGCACCTAGCCCAGGCAGGCA
Rat        ACCTAGGCTCCACAAACATGTATTATCAGAGGGCAGGGG-CACCGAGCCAGGCAGGCA
Turkey    ACACAGGCGCCAAAAACATCAGCGGATCAAAGGGCACTGC-GATTTATCCCGAGCAGAAT
Chicken    ACACAGGCGCCAAAAACATCAGCGGATCAAAGGGCACTGC-GATTTATCCCGAGCAGAAT
Pig        ACGCAGGCTCCACAAACATCGTGGGATCAGAGGGCAGGGG---CTCAGCCAGGGCA----
Dog        ATACAGGCTCCACAAACATGTAGGATCAGAGGGCAAGGG-G-CTTAGCCCAAGCG----
Cow        ACACAGGCTCCACAAACATGTAGGATCAGAGGGCAGGGG---CTCAGCCAGGGCA----
Cat        ACATAGGCTCCACAAACATGTGTGGATCAGAGGGCAGGGG-G-CTTAGCCCAAGCA----
Sheep      ACACAGGCTCCATAAACATCGTGGGATCAGAGGGCAGGGG---CTCAGCCAGGGCA----
Goat       ACACAGGCTCCATAAACATCGTGGGATCAGAGGGCAGGGG---CTCAGCCAGGGCA----
Horse      GTGCAGGCTCCACAAACATCG-GGGATCAGAGGGCAGGGG-GCTT--AGCCAGGGCA----
Chimpanzee ACACAGGCTCCACAAACACCATGGGGTGAGAGGGCAAGGG-GTCTTAAGCCAGGGCA----
Human      ACACAGGCTCCACAAACACCATGGGGTCAGAGGGCAAGGG-GTCTTAAGCCAGGGCA---- 580
          ****      *** *****      *** ***** *      *  **

Mouse      CTAGGCAAGCAGGCCCACAGCCGGGTAGCTC-CGTTTGTTTATTTT
Rat        CTAGCCAAGCAGGCCCCAACAGCCGGGTAGCTC-CATTTGTTTATTTT
Turkey    CTA----AACAAGCCTGTCAGCTCGCCCGGAC-CATTTGTTTATTTT
Chicken    CTA----AACAAGCCTGTCAGCTCGCCCGGAC-CATTTGTTTATTTT
Pig        CTAGCAAAACTGGCCCA-CAGCC--CCACCAC-CGTCTGTTTATTTT
Dog        CTGGCCAAACAGCCCCA-CAGCCCCACAGATC-CATTTGTTTATTTT
Cow        CTAGGCAAAACAGGCCTG-CAACC--CCAGCATGCA--TGTTTATTTT
Cat        -TAGCCAAACAGCCCT-CAACCCACAGATC-CATTTGTTTATTTT
Sheep      CTAGCCAAACAGGCCTG-CAGCC--CCACCATGCATCTGTTTATTTT
Goat       CTAGCCAAACAGGCCTG-CAGCC--CCACCATGCATCTGTTTATTTT
Horse      CTAGGCAAAACAGGCC--CAGCCCCACGGATC-CGTTTGTTTATTTT
Chimpanzee CTAGCCAAGCAGGCCCAGCAGCCCCGAGATC-CATTTGTTTATTTT
Human      CTAGCCAAGCAGGCCCAGCAGCCCCGAGATC-CATTTGTTTATTTT 626
          *      *  *  **      ** *      *  ***** *****

```

8.7 Apêndice 7

Rede de interações proteína-proteína para os sítios de ligação para fatores de transcrição preditos para o CNE8 humano. Linhas em cor rosa indicam interações determinadas experimentalmente e, em azul turquesa, interações conhecidas em banco de dados curados; linhas em cor verde claro indicam correlações presentes na literatura científica; e em verde escuro, co-expressão. Em vermelho destaca-se os fatores relacionados ao estudo “*Co-expression of MDM2 and CDK4 in transformed human mesenchymal stem cells causes high-grade sarcoma with a dedifferentiated liposarcoma-like morphology*”, de Kim et al., 2019).

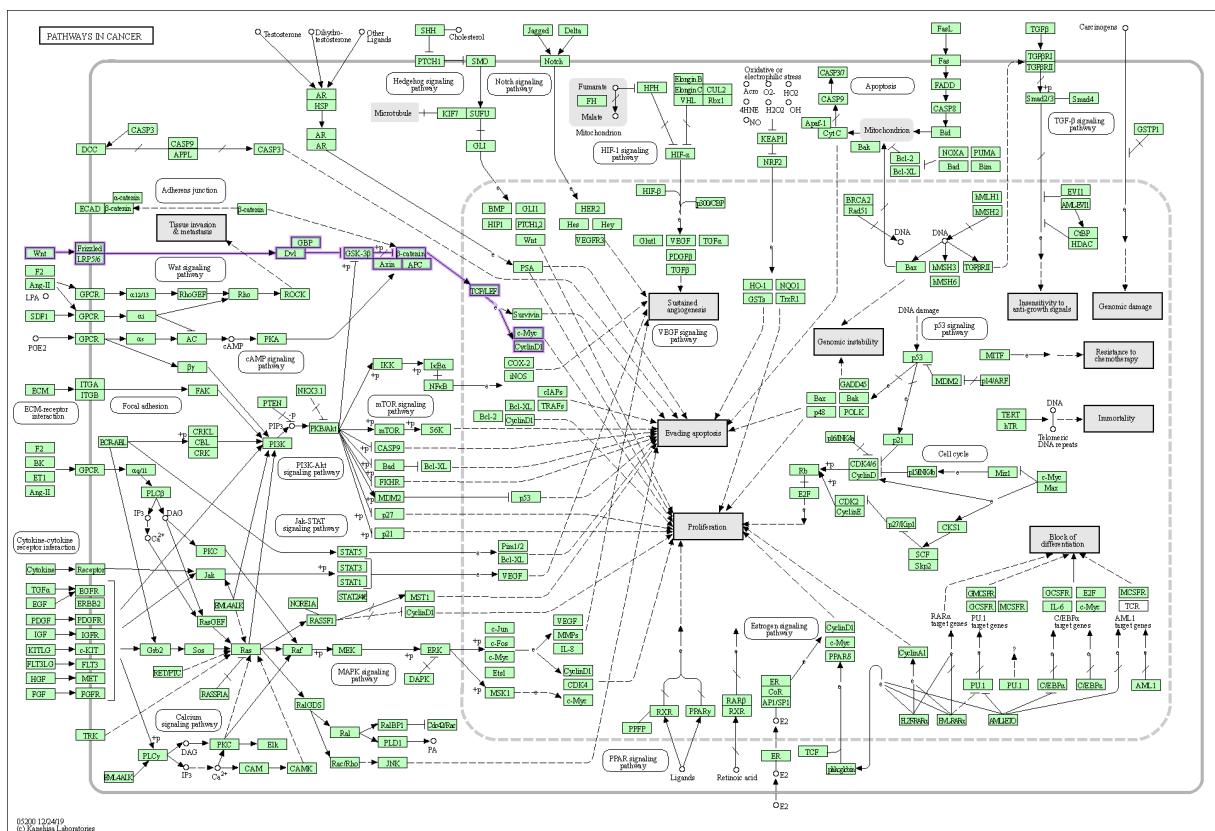


Reference publications			
publication	(year) title	count in gene set	false discovery rate
PMID:27833527	(2016) Intrinsic Axonal Growth and the Drive for Regenerati...	7 of 57	2.04e-07
PMID:24130129	(2013) Towards an understanding of cell-specific functions ...	7 of 52	2.04e-07
PMID:20233023	(2010) Regulation of hepatic ABCC transporters by xenobiot...	8 of 109	2.04e-07
PMID:24457012	(2014) Regulation of the latent-lytic switch in Epstein-Barr vi...	6 of 29	2.55e-07
PMID:28546992	(2017) Identification of transcription factors potentially invo...	7 of 85	1.02e-06
PMID:31160689	(2019) Co-expression of MDM2 and CDK4 in transformed h...	6 of 42	1.27e-06
PMID:23023699	(2012) Comparison of targeted maximum likelihood and shr...	5 of 15	1.38e-06
PMID:24066765	(2013) Molecular mechan (2019) Co-expression of MDM2 and CDK4 in		1.75e-06
PMID:16674813	(2006) Reductionoxidatio transformed human mesenchymal stem cells		1.82e-06
PMID:31000713	(2019) ZFP30 promotes a causes high-grade sarcoma with a		2.24e-06
PMID:30390616	(2018) Norepinephrine tri dedifferentiated liposarcoma-like morphology.		2.39e-06

8.8 Apêndice 8

Via KEGG hsa05200, de vias relacionadas ao câncer, na qual o CNE8 possui um enriquecimento de sítios de ligação para fatores de transcrição envolvidos. Em lilás destaca-se a via Wnt canônica.

KEGG Pathways			
<i>pathway</i>	<i>description</i>	<i>count in gene set</i>	<i>false discovery rate</i>
hsa05200	Pathways in cancer	8 of 515	2.02e-05
hsa05166	HTLV-I infection	6 of 250	3.85e-05
hsa05161	Hepatitis B	5 of 142	3.97e-05
hsa04010	MAPK signaling pathway	6 of 293	4.75e-05
hsa05203	Viral carcinogenesis	4 of 183	0.0020



9. ANEXOS

9.1 Artigos publicados em colaboração



Article

Injured Achilles Tendons Treated with Adipose-Derived Stem Cells Transplantation and GDF-5

Andrea Aparecida de Aro ^{1,2,*}, Giane Daniela Carneiro ¹, Luis Felipe R. Teodoro ¹,
Fernanda Cristina da Veiga ³, Danilo Lopes Ferrucci ¹, Gustavo Ferreira Simões ¹,
Priscyla Waleska Simões ⁴, Lúcia Elvira Álvares ³, Alexandre Leite R. de Oliveira ¹,
Cristina Pontes Vicente ¹, Caio Perez Gomes ⁵, João Bosco Pesquero ⁵,
Marcelo Augusto M. Esquisatto ², Benedicto de Campos Vidal ¹ and Edson Rosa Pimentel ¹

¹ Department of Structural and Functional Biology, Institute of Biology, State University of Campinas–UNICAMP, Charles Darwin, s/n, CP 6109, 13083-970 Campinas, SP, Brazil; gianedc@gmail.com (G.D.C.); teo.luisfelipe@gmail.com (L.F.R.T.); daniloferrucci@yahoo.com.br (D.L.F.); gfsimoes2@gmail.com (G.F.S.); alroliv@unicamp.br (A.L.R.d.O.); crpvicente@gmail.com (C.P.V.); camposvi@unicamp.br (B.d.C.V.); pimentel@unicamp.br (E.R.P.)

² Biomedical Sciences Graduate Program, Herminio Ometto University Center–UNIARARAS, 13607-339 Araras, SP, Brazil; marcelosquisatto@uniararas.br

³ Department of Biochemistry and Tissue Biology, Institute of Biology, State University of Campinas–UNICAMP, Charles Darwin, s/n, CP 6109, 13083-970 Campinas, SP, Brazil; fernandaveiga6@gmail.com (F.C.d.V.); lealvare@unicamp.br (L.E.Á.)

⁴ Engineering, Modeling and Applied Social Sciences Center (CECS), Biomedical Engineering Graduate Program (PPGEBM), Universidade Federal do ABC (UFABC), Alameda da Universidade s/n, 09606-045 São Bernardo do Campo, SP, Brazil; pritsimoes@gmail.com

⁵ Department of Biophysics, Federal University of São Paulo–Unifesp, Pedro de Toledo, 699, 04039-032 São Paulo, SP, Brazil; caiopgomes@hotmail.com (C.P.G.); jbpesquero@unifesp.br (J.B.P.)

* Correspondence: andreaaro80@gmail.com; Tel.: +55-19-3543-1423

Received: 19 July 2018; Accepted: 23 August 2018; Published: 31 August 2018



Abstract: Tendon injuries represent a clinical challenge in regenerative medicine because their natural repair process is complex and inefficient. The high incidence of tendon injuries is frequently associated with sports practice, aging, tendinopathies, hypertension, diabetes mellitus, and the use of corticosteroids. The growing interest of scientists in using adipose-derived mesenchymal stem cells (ADMSC) in repair processes seems to be mostly due to their paracrine and immunomodulatory effects in stimulating specific cellular events. ADMSC activity can be influenced by GDF-5, which has been successfully used to drive tenogenic differentiation of ADMSC in vitro. Thus, we hypothesized that the application of ADMSC in isolation or in association with GDF-5 could improve Achilles tendon repair through the regulation of important remodeling genes expression. Lewis rats had tendons distributed in four groups: Transected (T), transected and treated with ADMSC (ASC) or GDF-5 (GDF5), or with both (ASC+GDF5). In the characterization of cells before application, ADMSC expressed the positive surface markers, CD90 (90%) and CD105 (95%), and the negative marker, CD45 (7%). ADMSC were also differentiated in chondrocytes, osteoblast, and adipocytes. On the 14th day after the tendon injury, GFP-ADMSC were observed in the transected region of tendons in the ASC and ASC+GDF5 groups, and exhibited and/or stimulated a similar genes expression profile when compared to the in vitro assay. ADMSC up-regulated *Lox*, *Dcn*, and *Tgfb1* genes expression in comparison to T and ASC+GDF5 groups, which contributed to a lower proteoglycans arrangement, and to a higher collagen fiber organization and tendon biomechanics in the ASC group. The application of ADMSC in association with GDF-5 down-regulated *Dcn*, *Gdf5*, *Lox*, *Tgfb1*, *Mmp2*, and *Timp2* genes expression, which contributed to a lower hydroxyproline concentration, lower collagen fiber organization, and to an improvement of the rats' gait 24 h after the injury.



Article

Transected Tendon Treated with a New Fibrin Sealant Alone or Associated with Adipose-Derived Stem Cells

Katleen Frauz¹, Luis Felipe R. Teodoro¹, Giane Daniela Carneiro¹,
Fernanda Cristina da Veiga², Danilo Lopes Ferrucci¹, André Luis Bombeiro¹,
Priscylla Waleska Simões³, Lúcia Elvira Alvares², Alexandre Leite R. de Oliveira¹,
Cristina Pontes Vicente¹, Rui Seabra Ferreira Jr.⁴, Benedito Barraviera⁴,
Maria Esméria C. do Amaral⁵, Marcelo Augusto M. Esquisatto⁵, Benedicto de Campos Vidal¹,
Edson Rosa Pimentel¹ and Andrea Aparecida de Aro^{1,5,*}

¹ Department of Structural and Functional Biology, Institute of Biology, University of Campinas–UNICAMP, Charles Darwin, s/n, CP 6109, 13083-970 Campinas, SP, Brazil; kafrauz@hotmail.com (K.F.); teo.luisfelipe@gmail.com (L.F.R.T.); gianedc@gmail.com (G.D.C.); daniloferrucci@yahoo.com.br (D.L.F.); aobombeiro@gmail.com (A.L.B.); alroliv@unicamp.br (A.L.R.d.O.); crpvicente@gmail.com (C.P.V.); camposvi@unicamp.br (B.d.C.V.); pimentel@unicamp.br (E.R.P.)

² Department of Biochemistry and Tissue Biology, Institute of Biology, University of Campinas–UNICAMP, Charles Darwin, s/n, CP 6109, 13083-970 Campinas, SP, Brazil; fernandaveiga6@gmail.com (F.C.d.V.); lealvare@unicamp.br (L.E.A.)

³ Engineering, Modeling and Applied Social Sciences Center (CECS), Biomedical Engineering Graduate Program (PPGEBM), Universidade Federal do ABC (UFABC), Alameda da Universidade s/n, 09606-045 São Bernardo do Campo, SP, Brazil; pritsimoes@gmail.com

⁴ Center for the Study of Venoms and Venomous Animals (CEVAP), São Paulo State University (UNESP – Universidade Estadual Paulista), Botucatu, SP, St. José Barbosa de Barros, 1780, Fazenda Experimental Lageado, 18610-307 Botucatu, SP, Brazil; rui.ead@gmail.com (R.S.F.); bbviera@jvat.org.br (B.B.)

⁵ Biomedical Sciences Graduate Program, Herminio Ometto University Center–UNIARARAS, Av. Dr. Maximiliano Baruto, 500, Jd. Universitário, 13607-339 Araras, SP, Brazil; esmeria@fho.edu.br (M.E.C.d.A.); marcelosquisatto@fho.edu.br (M.A.M.E.)

* Correspondence: andreaaro80@gmail.com; Tel.: +55-19-3543-1423

Received: 26 November 2018; Accepted: 7 January 2019; Published: 16 January 2019

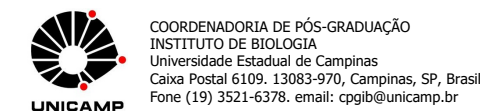


Abstract: Tissue engineering and cell-based therapy combine techniques that create biocompatible materials for cell survival, which can improve tendon repair. This study seeks to use a new fibrin sealant (FS) derived from the venom of *Crotalus durissus terrificus*, a biodegradable three-dimensional scaffolding produced from animal components only, associated with adipose-derived stem cells (ASC) for application in tendons injuries, considered a common and serious orthopedic problem. Lewis rats had tendons distributed in five groups: normal (N), transected (T), transected and FS (FS) or ASC (ASC) or with FS and ASC (FS + ASC). The in vivo imaging showed higher quantification of transplanted PKH26-labeled ASC in tendons of FS + ASC compared to ASC on the 14th day after transection. A small number of Iba1 labeled macrophages carrying PKH26 signal, probably due to phagocytosis of dead ASC, were observed in tendons of transected groups. ASC up-regulated the *Tenomodulin* gene expression in the transection region when compared to N, T and FS groups and the expression of *TIMP-2* and *Scleraxis* genes in relation to the N group. FS group presented a greater organization of collagen fibers, followed by FS + ASC and ASC in comparison to N. Tendons from ASC group presented higher hydroxyproline concentration in relation to N and the transected tendons of T, FS and FS + ASC had a higher amount of collagen I and tenomodulin in comparison to N group. Although no marked differences were observed in the other biomechanical parameters, T group had higher value of maximum load compared to the groups ASC and FS + ASC. In conclusion, the FS kept constant the number of transplanted ASC in the transected region until the 14th day after injury. Our data suggest this FS to be a good scaffold for treatment during tendon repair because it was the

9.2 Capítulo de livro publicado

ARROTEIA, Kélen Fabíola; BRAGA, Daniela Paes de Almeida Ferreira; DA VEIGA, Fernanda Cristina; DA COSTA, Carolina Frandsen Pereira; PEREIRA, Luis Antônio Violin Dias. Fisiologia Reprodutiva. In: JÚNIOR, Edson Borges; BRAGA, Daniela Paes de Almeida Ferreira; SETTI, Amanda Souza (org.). Reprodução Humana Assistida. 2. Ed. Rio de Janeiro: Atheneu, 2020. p. 1-18. ISBN 978-65-5586-021-4.

9.3 Declaração de bioética e biossegurança



DECLARAÇÃO

Em observância ao **§5º do Artigo 1º da Informação CCPG-UNICAMP/001/15**, referente a Bioética e Biossegurança, declaro que o conteúdo de minha Tese de Doutorado, intitulada ***“Análise da regulação gênica e expressão de Dact1 no desenvolvimento da musculatura esquelética de amniotos”***, desenvolvida no Programa de Pós-Graduação em Biologia Celular e Estrutural do Instituto de Biologia da Unicamp, não versa sobre pesquisa envolvendo seres humanos, animais ou temas afetos a Biossegurança.

Assinatura: 
Nome do(a) aluno(a): Fernanda Cristina da Veiga

Assinatura: 
Nome do(a) orientador(a): Lucia Elvira Alvares

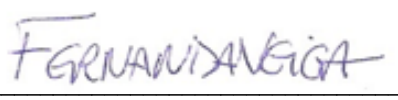
Data: 27/11/2020

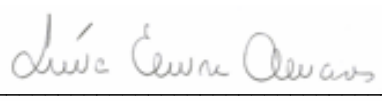
9.4 Declaração de direitos autorais

Declaração

As cópias de artigos de minha autoria ou de minha co-autoria, já publicados ou submetidos para publicação em revistas científicas ou anais de congressos sujeitos a arbitragem, que constam da minha Dissertação/Tese de Mestrado/Doutorado, intitulada **Análise da regulação gênica e expressão de Dact1 no desenvolvimento da musculatura esquelética de amniotos**, não infringem os dispositivos da Lei n.º 9.610/98, nem o direito autoral de qualquer editora.

Campinas, 27 de novembro de 2020.

Assinatura : 
Nome do(a) autor(a): **Fernanda Cristina da Veiga**
RG n.º 43.682.978-2

Assinatura : 
Nome do(a) orientador(a): **Lucia Elvira Alvares**
RG n.º 17.371.210-1