

INSTITUTO CARLOS CHAGAS
Doutorado em Biociências e Biotecnologia

**CARACTERIZAÇÃO MOLECULAR DA REGULAÇÃO PÓS-TRANSCRICIONAL
DURANTE A DIFERENCIAÇÃO CARDIOMIOGÊNICA**

ISABELA TIEMY PEREIRA

CURITIBA/PR
2019



Ministério da Saúde
FIOCRUZ
Fundação Oswaldo Cruz
Instituto Carlos Chagas
Pós-Graduação em Biociências e Biotecnologia

ISABELA TIEMY PEREIRA

**CARACTERIZAÇÃO MOLECULAR DA REGULAÇÃO PÓS-TRANSCRICIONAL
DURANTE A DIFERENCIAÇÃO CARDIOMIOGÊNICA**

Tese apresentada ao Instituto Carlos Chagas como
parte dos requisitos para a obtenção do título de
Doutora em Biociências e Biotecnologia

Orientador: Dr. Bruno Dallagiovanna Muñiz

**CURITIBA/PR
2019**

Pereira, Isabela Tiemy.

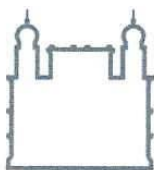
Caracterização molecular da regulação pós-transcricional durante a diferenciação cardiomiogênica / Isabela Tiemy Pereira. - Curitiba, 2019.
173 f.; il.

(Doutorado) - Instituto Carlos Chagas, Pós-Graduação em Biociências e Biotecnologia, 2019.

Orientador: Bruno Dallagiovanna.

Bibliografia: Inclui Bibliografias.

1. células-tronco pluripotentes humanas. 2. diferenciação cardiomiogênica. 3. RNA-seq. 4. *polysome profiling* . 5. lncRNA. I. Título.



Ministério da Saúde

FIOCRUZ - PARANÁ
Instituto Carlos Chagas

Ata da Sessão Pública de exame de tese para obtenção do grau de Doutora em
Biociências e Biotecnologia.

Aos 22 dias do mês de fevereiro de dois mil e dezenove, às 14:00 horas, nas dependências do Instituto Carlos Chagas/Fiocruz Paraná, reuniu-se a Banca Examinadora designada pelo Colegiado do Programa de Pós-Graduação em Biociências e Biotecnologia, composta pelos Professores: Dr. Samuel Goldenberg, Dr. Bruno Solano de Freitas Souza e Dra. Daniela Fiori Gradia, com a finalidade de julgar a tese da candidata **Isabela Tiemy Pereira**, intitulada: “**Caracterização molecular da regulação pós - transcricional durante a diferenciação cardiomiogênica**”, sob a orientação de orientação de Dr. Bruno Dallagiovanna, para obtenção do grau de **Doutora** em Biociências e Biotecnologia. A candidata teve até 45 (quarenta e cinco) minutos para a apresentação, e cada examinador teve um tempo máximo de arguição de 30 (trinta) minutos, seguido de 30 (trinta) minutos para resposta do (a) candidato(a) ou de 60 (sessenta) minutos quando houve diálogo na arguição. O desenvolvimento dos trabalhos seguiu o roteiro de sessão de defesa, estabelecido pela Coordenação do Programa, com abertura, condução e encerramento da sessão solene de defesa feito pelo(a) Presidente **Dr. Samuel Goldenberg**. Após haver analisado o referido trabalho e arguido a candidata, os membros da banca examinadora deliberaram pela:


Aprovação da tese por unanimidade

Aprovação somente após satisfazer as exigências de modificações no prazo fixado pela banca (não superior a noventa dias).

Reprovação da tese

Na forma regulamentar foi lavrada a presente ata que é abaixo assinada pelos membros da banca, na ordem acima determinada, e pela candidata.

Curitiba, 22 de fevereiro de 2019


Dr. Samuel Goldenberg (Presidente) ICC/Fiocruz-PR


Dr. Bruno Solano de Freitas Souza – IGM/Fiocruz-BA


Dra. Daniela Fiori Gradia - UFPR

Candidato(a): Isabela Tiemy Pereira

+55(41)3316-3230

Rua Prof. Algacyr Munhoz Mader, 3775 - CIC 81350-010 Curitiba/PR, Brasil



*“Pés, pra que te quero, se tenho asas para voar?”
Frida Kahlo*

AGRADECIMENTOS

Aos meus pais, France e Ozias, e irmã, Beatriz, pelo amor e suporte incondicional.

Aos amigos, que sempre deixam um pedacinho de si para me fortalecer.

Ao Dr. Bruno Dallagiovanna, por ter aceitado me orientar e acreditar no meu potencial em desenvolver este trabalho. Por todo conhecimento transmitido, discussões científicas e pela amizade.

Aos amigos e (ex) membros do Laboratório de Biologia Básica de Células-tronco (LABCET): Alejandro Correa Dominguez, Marco Augusto Stimamiglio, Patrícia Shigunov, Alessandra Melo de Aguiar, Crisciele Kuligovski, Anny Robert, Ana Paula Abud, Axel Cofré, Bruna Marcon, Addeli Angulski, Amanda Leitolis, Thamile Reus, Maria Júlia Barison, Ariane Paschoal, Cíntia Horinouchi e Bernardo Bonilauri. Por compartilhar um ambiente de trabalho maravilhoso, pelos ensinamentos, pelos jogos de vôlei, pelas comilanças e pela amizade.

À Lucía Spangenberg e ao Hugo Naya, da Unidade de Bioinformática do Instituto Pasteur de Montevideo, pela intensa contribuição com o trabalho.

Ao Dr. Michael Kyba, pela orientação no Doutorado Sanduíche, por proporcionar o engrandecimento deste trabalho e por todo conhecimento compartilhado. Estendo este agradecimento a todos os membros do Kyba lab na Universidade de Minnesota (EUA), em especial ao Dr. Sunny Chan, pelos conselhos e ajuda imensurável.

Ao pessoal técnico, administrativo e de manutenção do Instituto Carlos Chagas por propiciarem excelentes condições de trabalho.

Aos membros da banca examinadora: Dr. Samuel Goldenberg, Dr. Bruno Solano e Dra. Daniela Gradia, pelo tempo dedicado à leitura deste documento e pela contribuição na minha formação.

À Pós-graduação em Biociências e Biotecnologia pela oportunidade de completar mais esta etapa da minha formação.

Às agências financiadoras FIOCRUZ, CNPq, CAPES e FUNDAÇÃO ARAUCÁRIA pelo auxílio financeiro.

RESUMO

A diferenciação de células-tronco pluripotentes é um processo coordenado que envolve o gatilho, a manutenção e o controle dos padrões de expressão gênica. O presente trabalho teve como objetivo analisar as frações de RNAs livre e polissomal como uma estratégia para investigar mecanismos de regulação pós-transcricional envolvidos na diferenciação cardiomiogênica de células-tronco. A linhagem de células embrionárias humanas NKX2-5^{eGFP/w} foi cultivada e diferenciada à cardiomiócitos e submetida à análise em cinco diferentes pontos ao longo da diferenciação (dias D0, D1, D4, D9 e D15) usando *polysome profiling* e sequenciamento de RNA em larga escala (RNA-seq). Os resultados mostraram que os RNAs polissomais refletem o fenótipo das células em diferenciação e que um grande número de RNAs está sob efeito de regulação pós-transcricional. Vias relacionadas à tradução apresentaram-se reguladas negativamente após a diferenciação, e constatou-se que cardiomiócitos diferenciados apresentaram níveis de síntese proteica menores que as células pluripotentes. Além disso, os genes *SHISA3*, *RXRG* e *CSDC2* foram escolhidos para serem avaliados funcionalmente na diferenciação cardíaca usando linhagens de expressão induzida. As novas linhagens foram geradas a partir da linhagem hESC H1 e mostraram manutenção da pluripotência após indução da expressão dos genes com doxiciclina, indicando que os genes em questão não parecem estar envolvidos com a manutenção ou escape da pluripotência. Por outro lado, a super expressão de *CSDC2* entre os dias D7-D10 da diferenciação cardíaca mostrou um aumento significativo na expressão de genes cardíacos, como *TNNT2* e *TNNI3*, sugerindo uma possível contribuição na diferenciação e/ou maturação de cardiomiócitos. Os dados de RNA-seq também possibilitaram a análise de RNAs não-codificantes longos (lncRNAs) expressos durante a diferenciação cardiomiogênica. Tais lncRNAs mostraram um padrão de expressão temporal ao longo da diferenciação, e um grande número mostrou-se diferencialmente associado aos polissomos. O lncRNA *LINC00890* se apresentou expresso em progenitores cardíacos e foi escolhido para a avaliação funcional. Linhagens de expressão induzida e com o *LINC00890* silenciado foram construídas e submetidas à diferenciação cardiomiogênica. A super expressão de *LINC00890* entre os dias D5-D10 mostrou indícios do favorecimento na maturação de cardiomiócitos e na determinação de propriedades eletrofisiológicas de células atriais. O silenciamento de *LINC00890* usando a técnica de CRISPR/Cas9 foi feito nas linhagens hESC H1 e iPSC PLZ, e as células silenciadas foram capazes de se diferenciar até o estágio de mesoderme, mas não em cardiomiócitos, indicando um possível papel essencial desse lncRNA na diferenciação cardiomiogênica.

Palavras-chave: células-tronco pluripotentes humanas, diferenciação cardiomiogênica, RNA-seq, *polysome profiling*, lncRNA

ABSTRACT

Differentiation of pluripotent stem cells is a highly coordinated process involving trigger, maintenance and coordination of gene expression patterns. We aimed to analyze the ribosome-free and polysome-associated RNA fractions as a strategy to investigate post-transcriptional mechanisms involved in cardiomyogenic differentiation. The human embryonic cell line NKX2-5^{eGFP/w} was differentiated to cardiomyocytes and submitted to a time course analysis (D0, D1, D4, D9 and D15 days of differentiation) using polysome profiling and high-throughput RNA sequencing (RNA-seq). Our results showed that polysome-bound RNAs reflect differentiating cells phenotype and that the majority of mRNAs was under post-transcriptional regulation. Translation-related pathways were downregulated after differentiation and final cardiomyocytes showed decreased protein synthesis when compared to pluripotent cells. Furthermore, *SHISA3*, *RXRG* and *CSDC2* were chosen for functional analysis during cardiomyogenesis using inducible expression cell lines. The new cell lines were derived from H1 hESC and maintained their pluripotency after expression induction of the mentioned genes using doxycycline, suggesting that these genes are not involved with pluripotency maintenance or escape. On the other hand, inducible expression of *CSDC2* during days 7 and 10 of cardiomyogenic differentiation increased the expression of cardiac genes, such as *TNNT2* and *TNNI3*, suggesting its contribution to cardiomyocytes differentiation or maturation. Additionally, RNA-seq data also allowed the analysis of long non-coding RNAs (lncRNAs) expression along cardiomyogenesis. lncRNAs showed temporal expression patterns and also differential association with polysomes during cardiac differentiation. The lncRNA *LINC00890* is expressed in the cardiac progenitor stage and was chosen for functional analysis. Inducible expression and knock-out cell lines were derived and submitted to cardiomyogenic differentiation. Cells in which *LINC00890* was expressed during D5-D10 of differentiation showed improvement in maturation and more atrial physiological properties. CRISPR/Cas9 technology was used to knock-out *LINC00890* in H1 hESC and PLZ iPSCs, which were able to differentiate to mesoderm stage, but not to cardiomyocytes, suggesting an essential role of this lncRNA in cardiac development.

Key-words: human pluripotent stem cells, cardiomyogenic differentiation, RNA-seq, polysome profiling, lncRNA

LISTA DE FIGURAS

Figura 1.1	Desenvolvimento cardíaco em mamíferos.....	21
Figura 1.2	Vias de sinalização e marcadores das diferentes etapas da diferenciação cardíaca.....	23
Figura 1.3	Características eletrofisiológicas e marcadores de células atriais e ventriculares.....	26
Figura 1.4	Diferenciação cardiomiogênica <i>in vitro</i> a partir de CTP humanas.....	28
Figura 4.1	Mapa do vetor AAVS1-TRE-FLAG-rtTA.....	77
Figura 4.2	Estratégia de deleção dos motivos de ligação ao RNA da proteína codificada pelo gene <i>CSDC2</i>	80
Figura 5.1	Expressão de marcadores ao longo da diferenciação cardiomiogênica em monocamada.....	82
Figura 5.2	Expressão de marcadores cardíacos durante a diferenciação em monocamada.....	83
Figura 5.3	Expressão dos genes candidatos ao longo da diferenciação cardíaca em CE e monocamada.....	84
Figura 5.4	Sistema de super expressão induzível com doxiciclina.....	86
Figura 5.5	Marcação com imunofluorescência mostrando a expressão de <i>CSDC2</i> e <i>RXRG</i>	87
Figura 5.6	Linhagens celulares H1 <i>iRXRG</i> , <i>iSHISA3</i> e <i>iCSDC2</i> mantiveram características pluripotentes após algumas passagens em cultivo (-DOX) e após indução com doxiciclina (+DOX).....	88/ 89
Figura 5.7	Ensaio funcional de diferenciação cardiomiogênica com a linhagem H1 <i>iRXRG</i>	90
Figura 5.8	Ensaio funcional de diferenciação cardiomiogênica com a linhagem H1 <i>iSHISA3</i>	92
Figura 5.9	Ensaio funcional de diferenciação cardíaca com a linhagem H1 <i>iCSDC2</i>	94
<i>Figure 1.</i>	LncRNAs along human cardiomyogenesis.....	106
<i>Figure 2.</i>	Differentially expressed lncRNAs during cardiomyogenic commitment.....	107
<i>Figure 3.</i>	Gene expression pattern clusters of ribosome-free and polysome-bound lncRNAs during cardiomyogenesis.....	108
<i>Figure 4.</i>	LncRNAs association with polysomes during cardiomyogenesis.	109
<i>Sup Fig 1.</i>	Gene expression profiles of the 7 clusters for polysome-bound lncRNAs.....	110
<i>Sup Fig 2.</i>	Gene expression profiles of the 7 clusters for ribosome-free lncRNAs.....	110
<i>Sup Fig 3.</i>	Venn diagrams showing specific and common lncRNAs from Figure 3.....	110
Figura 6.1	Mapa do vetor AAVS1-TRE-GW-rtTA.....	116

Figura 6.2	Estratégia de mutação na CDS predita para o gene <i>LINC00890</i>	118
Figura 6.3	Estratégia de silenciamento do gene <i>LINC00890</i> através de deleção utilizando um par de gRNAs (dentre os gRNAs A, B, C e D).....	119
Figura 7.1	Expressão do lncRNA <i>LINC00890</i> durante a diferenciação cardiomiogênica em CE e monocamada.....	123
Figura 7.2	Sistema de super expressão induzível com doxiciclina.....	124
Figura 7.3	Linhagem H1 iLINC manteve características pluripotentes após algumas passagens em cultivo (-DOX) e após indução com doxiciclina (+DOX).....	125
Figura 7.4	Ensaio funcionais de triagem com a linhagem H1 iLINC.....	127
Figura 7.5	Expressão dos marcadores cardíacos nas linhagens H1 iLINC e PLZ iLINC após indução com doxiciclina nos dias 5-10 (DOX 5-10).....	129
Figura 7.6	Expressão de marcadores de subtipos e de maturação de cardiomiócitos avaliada por qPCR no dia 30 da diferenciação.....	130
Figura 7.7	Expressão gênica nas linhagens H1 iLINC e PLZ iLINC nos dias 10 e 30 da diferenciação.....	131
Figura 7.8	Quantificação da frequência de batimentos nos cardiomiócitos derivados da linhagem H1 iLINC.....	131
Figura 7.9	Análise eletrofisiológica de cardiomiócitos por Patch clamp.....	133
Figura 7.10	Construção de linhagens com silenciamento de <i>LINC00890</i>	135
Figura 7.11	Seleção de clones com silenciamento de <i>LINC00890</i> por deleção.....	136
Figura 7.12	Clones silenciados para <i>LINC00890</i> mantêm o estado pluripotente.....	137
Figura 7.13	Expressão de <i>LINC00890</i> nas linhagens H1 LINC KO e PLZ LINC KO ao longo da diferenciação cardiomiogênica.....	138
Figura 7.14	Diferenciação cardiomiogênica com a linhagem H1 LINC KO.....	139
Figura 7.15	Expressão gênica de marcadores de pluripotência e mesoderme durante a diferenciação cardíaca das linhagens H1 LINC KO e PLZ LINC KO.....	140
Figura 7.16	Expressão gênica de marcadores cardíacos durante a diferenciação cardíaca das linhagens H1 LINC KO e PLZ LINC KO.....	141
Figura 7.17	Mapeio dos dados de RNA-seq na região do gene <i>LINC00890/SERTM2</i>	142
Figura 7.18	Árvore filogenética relacionando os organismos de acordo com a sequência de RNA de <i>SERTM2</i>	143
Figura 7.19	Árvore filogenética e alinhamento das sequências proteicas de <i>SERTM2</i> em diferentes mamíferos.....	145

LISTA DE TABELAS

Tabela 4.1	Lista de genes e suas respectivas Taqman probes utilizadas....	76
Tabela 4.2	Lista de genes e seus respectivos oligos forward (F) e reverso (R) utilizados.....	76
<i>Sup Table 1.</i>	Pathway analysis of lncRNAs in each RNA fraction.....	109
Tabela 6.1	Oligos usados na PCR de avaliação de deleção do gene <i>LINC00890</i>	121
Tabela 7.1	Análise comparativa de sequências nucleotídicas de diferentes organismos quanto à similaridade com o gene <i>LINC00890/SERTM2</i> humano.....	144
Tabela 7.2	Número de ortólogos de <i>SERTM2</i> , em nível proteico.....	146

LISTA DE ABREVIATURAS E SIGLAS

ACTA2	- <i>Alpha-smooth muscle actin</i>
AP	- Potencial de ação
APC	- <i>Adenomatous Polyposis Coli</i>
aprox.	- Aproximadamente
α -SMA	- <i>Alpha-smooth muscle actin</i>
BMP	- <i>Bone Morphogenetic Protein</i>
BMP4	- <i>Bone Morphogenetic Protein 4</i>
BP	- <i>Biological process</i>
BSA	- Albumina de soro bovino
Ca ²⁺	- Íon de cálcio
CARMEN	- <i>(CAR)diac (M)esoderm (E)nhancer-associated (N)oncoding RNA</i>
cDNA	- DNA complementar
CDS	- Sequência codificante
CE(s)	- Corpos embrióides
CHIR99021	- <i>6-[[2-[[4-(2,4-Dichlorophenyl)-5-(5-methyl-1H-imidazol-2-yl)-2-pyrimidinyl]amino]ethyl]amino]-3-pyridinecarbonitrile</i>
cMyc	- Proto-oncogene c-Myc
CO ₂	- Gás carbônico
COA	- <i>Correspondence analysis</i>
CPM	- <i>Counts per million mapped reads</i>
CRISPR	- <i>Clustered Regularly Interspaced Short Palindromic Repeats</i>
CSD	- <i>Cold shock domains</i>
CSDC2	- <i>Cold Shock Domain Containing C2</i>
CT(s)	- Células-tronco
CTE(s)	- Células-tronco embrionárias
cTnI	- Troponina I cardíaca
cTnT	- Troponina T cardíaca
CTP(s)	- Células-tronco pluripotentes
Cx43	- <i>Connexin 43</i>
D0	- Dia 0 (inicial) da diferenciação cardiomiogênica
D1	- Dia 1 da diferenciação cardiomiogênica
D4	- Dia 4 da diferenciação cardiomiogênica
D9	- Dia 9 da diferenciação cardiomiogênica
D15	- Dia 15 da diferenciação cardiomiogênica
DEG	- <i>Differentially expressed genes</i>
DKK1	- <i>Dickkopf-1</i>
DMEM	- Meio <i>Dulbecco's modified eagle</i>
DMSO	- Dimetilsulfóxido
DNA	- Ácido desoxirribonucleico
DOX	- Doxiciclina
EB(s)	- <i>Embryoid body</i>
eGFP	- Proteína fluorescente verde

EOMES	- Eomesodermina
FGF	- <i>Fibroblast growth factor</i>
FZD	- Receptor <i>Frizzled</i>
GAPDH	- <i>Glyceraldehyde-3-Phosphate Dehydrogenase</i>
GATA	- <i>Gata binding protein</i>
GATA4	- <i>Gata binding protein 4</i>
GO	- <i>Gene Ontology</i>
gRNA	- RNA guia
GSK3	- <i>Glycogen Synthase Kinase 3</i>
HAND	- <i>Heart And Neural Crest Derivatives Expressed</i>
hASC	- <i>Human adipose stem cells</i>
HBL1	- <i>Heart Brake LncRNA 1</i>
HCN4	- <i>Hyperpolarization Activated Cyclic Nucleotide Gated Potassium Channel 4</i>
hESC	- <i>Human embryonic stem-cell</i>
HLA	- Antígeno leucocitário humano
iPSCs	- Células-tronco de pluripotência induzida
IRES	- <i>Internal Ribosome Entry Sites</i>
IRX	- <i>Iroquois Homeobox</i>
IWP2	- <i>Inhibitor of WNT Production-2</i>
JUN	- <i>Jun-N-terminal kinase</i>
KCNA5	- <i>Potassium Voltage-Gated Channel Subfamily A Member 5</i>
KCNJ3	- <i>Potassium Voltage-Gated Channel Subfamily J Member 3</i>
KH	- <i>K-homology domain</i>
KLF-4	- <i>Kruppel Like Factor 4</i>
KO	- <i>Knock-out, silenciado</i>
KSR	- <i>Knock-out serum</i>
LA	- <i>Left atrium</i>
LB	- Meio Luria-Bertani
LCAD	- <i>Acyl-CoA Dehydrogenase Long Chain</i>
LINC00890	- lncRNA <i>LINC00890</i>
lincRNA	- RNA não-codificante longo intergênico
lncRNA	- RNAs não-codificantes longos
logFC	- <i>Log Fold change</i>
LV	- <i>Left ventricle</i>
MEF	- <i>Mouse Embryonic Fibroblast</i>
MEF2	- <i>Myocyte Enhancer Factor 2</i>
MESP1	- <i>Mesoderm Posterior BHLH Transcription Factor 1</i>
miRNA	- microRNAs
mRNA	- RNA mensageiro
mRNP	- complexos de ribonucleoproteínas mensageiras
MTG	- Monotioglicerol
MYH	- <i>Myosin Heavy Chain</i>
MYL	- <i>Myosin Light Chain</i>
NANOG	- <i>Homeobox transcription factor nanog</i>
ncRNA	- RNA não-codificante

NKX2	- <i>NK2 Homeobox</i>
NKX2-5	- <i>NK2 Homeobox 5</i>
NPPA	- <i>Natriuretic Peptide A</i>
NR2F2	- <i>Nuclear Receptor Subfamily 2 Group F Member 2</i>
O ₂	- Oxigênio
Oct4	- <i>Octamer-binding protein 4</i>
OFT	- <i>Outflow tract</i>
OPP	- <i>O-propargyl-puromycin</i>
ORF	- <i>Open reading frames</i>
PAX6	- <i>Paired box 6</i>
PBS	- Salina tamponada com fosfato
PCR	- Reação em cadeia da polimerase
PFA	- Paraformaldeído
piRNA	- RNAs associados a PIWI
PRC	- complexo repressivo Polycomb
qPCR	- PCR quantitativa
RA	- <i>Right atrium</i>
RBP	- <i>RNA-binding protein</i>
RFP	- proteína fluorescente vermelha
RISC	- complexo de silenciamento induzido por RNA
RNA	- Ácido ribonucleico
RNA-seq	- sequenciamento de RNAs em larga escala
RPKM	- <i>Reads Per Kilobase per Million mapped reads</i>
RPMI	- Meio Roswell Park Memorial Institute
RRM	- <i>RNA recognition motif</i>
rRNA	- RNAs ribossomais
RT-PCR	- PCR acoplada à reação de transcrição reversa
rtTA	- <i>Reverse tetracycline-controlled transactivator</i>
RV	- <i>Right ventricle</i>
RXRG	- <i>Retinoid X Receptor Gamma</i>
SERCA2	- <i>ATPase Sarcoplasmic/Endoplasmic Reticulum Ca²⁺ Transporting 2</i>
SERTM2	- <i>Serine Rich And Transmembrane Domain Containing 2 (LINC00890)</i>
SFB	- Soro fetal bovino
SHISA3	- <i>Shisa Family Member 3</i>
siRNA	- pequenos RNAs de interferência
snoRNA	- RNAs nucleolares pequenos
snRNA	- RNAs nucleares pequenos
SOC	- Meio <i>Super Optimal broth with Catabolite repression</i>
SOX2	- <i>Sex determining region y-box 2</i>
SRF	- Fator de resposta ao soro
SV	- <i>Sinus venosus</i>
TBX	- <i>T-box transcription factor</i>
TER	- <i>Tetracyclin responsive element</i>
TNNT2	- <i>Cardiac Troponin T 2</i>
tRNA	- RNAs transportadores

uORF	- <i>Upstream open reading frames</i>
UTR	- <i>Untranslated region</i>
VEGF	- <i>Vascular Endothelial Growth Factor</i>
VEGFR2	- <i>Vascular Endothelial Growth Factor Receptor 2</i>
VWF	- <i>Von willebrand factor</i>
Wnt	- <i>Wingless</i>
WT	- <i>Wildtype, selvagem</i>
XAV939	- <i>3,5,7,8-Tetrahydro-2-[4-(trifluoromethyl)phenyl]-4H-thiopyrano[4,3-d]pyrimidin-4-one</i>
XIST	- <i>X-inactive specific transcript</i>
ZFN	- <i>Zinc fingers nucleases</i>

SUMÁRIO

1	INTRODUÇÃO	18
1.1	Células-tronco	18
1.1.1	Células-tronco pluripotentes	19
1.2	Desenvolvimento cardíaco	20
1.2.1	Sinalização molecular durante o desenvolvimento cardíaco	21
1.2.2	Desenvolvimento das câmaras cardíacas	23
1.3	Diferenciação cardiomiogênica a partir de células-tronco pluripotentes.....	26
1.4	Regulação da expressão gênica	29
1.4.1	Regulação pós-transcricional da expressão gênica.....	30
1.5	RNAs não-codificantes longos (lncRNAs)	34
1.5.1	lncRNAs no desenvolvimento cardíaco	38
2	JUSTIFICATIVA	40
3	OBJETIVOS	42
3.1	Objetivo Geral	42
3.2	Objetivos específicos	42
	CAPÍTULO I	43
	<i>Polysome profiling followed by RNA-seq of cardiac differentiation stages in hESCs</i>	43
	CAPÍTULO II	55
	<i>Cardiomyogenic differentiation is fine-tuned by differential mRNA association with polysomes</i>	55
	CAPÍTULO III	72
4	METODOLOGIA	72
4.1	Cultura de células-tronco pluripotentes humanas (CTPs)	73
4.2	Protocolo de diferenciação cardiomiogênica em monocamada	73
4.3	Citometria analítica.....	74
4.4	Imunofluorescência	74
4.5	Extração de RNA e síntese de cDNA.....	75

4.6	Quantificação da expressão gênica por PCR quantitativo (qPCR)	75
4.7	Construção de linhagens celulares com super expressão induzida	77
4.7.1	Clonagem dos genes candidatos no vetor AAVS1-TRE-FLAG-rtTA	77
4.7.2	Transformação de bactérias e extração de DNA plasmidial	78
4.7.3	Clonagem do gene <i>CSDC2</i> com deleção no vetor AAVS1-TRE-FLAG-rtTA	79
4.7.4	Transfecção dos vetores AAVS1 em células-tronco pluripotentes humanas	80
4.7.5	Seleção de clones com super expressão induzida	81
5	RESULTADOS	81
5.1	Diferenciação cardiomiogênica de CTPs	81
5.2	Genes candidatos escolhidos para avaliação funcional	83
5.3	Construção de linhagens celulares de super expressão induzida	85
5.4	Ensaio funcionais com a linhagem H1 iRXRG	89
5.5	Ensaio funcionais com a linhagem H1 iSHISA3	90
5.6	Ensaio funcionais com a linhagem H1 iCSDC2	92
5.6.1	Perspectivas para a análise do gene <i>CSDC2</i>	95
CAPÍTULO IV		96
	<i>Polysome-associated lncRNAs along cardiomyogenesis of hESC</i>	96
CAPÍTULO V		114
6	METODOLOGIA	115
6.1	Construção de linhagens celulares com super expressão induzida de <i>LINC00890</i>	115
6.1.1	Clonagem do gene <i>LINC00890</i> no vetor AAVS1-GW-TRE-rtTA	115
6.1.2	Clonagem do gene <i>LINC00890</i> com mutação no vetor AAVS1-GW-TRE-rtTA	117
6.1.3	Transfecção dos vetores AAVS1 em células-tronco pluripotentes humanas e seleção de clones	118
6.2	Silenciamento do gene <i>LINC00890</i> usando CRISPR/Cas9	118

6.2.1	Montagem do sistema Cas9/gRNA no vetor PX458	118
6.2.2	Transfecção dos vetores PX458 em células-tronco pluripotentes humanas	120
6.2.3	Avaliação e validação da deleção por PCR	120
6.2.4	Seleção de clones com silenciamento gênico	121
6.3	Ensaio de fisiologia de cardiomiócitos derivados de CTP	121
6.3.1	Quantificação da frequência de batimento de cardiomiócitos.....	121
6.3.2	Eletrofisiologia de cardiomiócitos usando Patch clamp	122
7	RESULTADOS	122
7.1	Escolha de <i>LINC00890</i> para avaliação funcional	122
7.2	Construção de linhagens celulares de super expressão induzida de <i>LINC00890</i>	124
7.3	Ensaio funcional com as linhagens H1 iLINC e PLZ iLINC.....	125
7.4	Construção de linhagens com silenciamento do gene <i>LINC00890</i>	133
7.5	Ensaio funcional com as linhagens H1 LINC KO e PLZ LINC KO	137
7.6	Análise <i>in silico</i> do gene <i>LINC00890</i>	142
7.7	Perspectivas para a análise do gene <i>LINC00890</i>	146
8	DISCUSSÃO	147
9	CONCLUSÕES	155
10	PERSPECTIVAS GERAIS	157
11	REFERÊNCIAS	158
	ANEXO I.....	169
	ANEXO II.....	170
	ANEXO III.....	171
	ANEXO IV	172
	ANEXO V	173

1 INTRODUÇÃO

1.1 Células-tronco

As células-tronco (CT) caracterizam-se por possuírem a capacidade de autorrenovação e de originar uma ou mais linhagens celulares diferenciadas (FUCHS & SEGRE, 2000; WEISSMAN, 2000). Essas células podem ser divididas de acordo com seu potencial de diferenciação em totipotentes, pluripotentes, multipotentes ou unipotentes (WAGERS & WEISSMAN, 2004). Durante a embriogênese, um organismo multicelular é derivado de apenas uma célula – o ócito fecundado. À medida que o organismo vai se desenvolvendo, as células e tecidos adquirem características específicas que os possibilitam realizar diferentes funções. Após sua formação, os tecidos se mantêm através de um processo chamado de homeostase, o qual repõe células mortas naturalmente ou por injúria, garantindo sua funcionalidade (WAGERS & WEISSMAN, 2004). Essa progressão durante o desenvolvimento ilustra os diferentes potenciais das CTs na formação e manutenção dos organismos. Células toti e pluripotentes, encontradas nos embriões, permitem a formação dos mais variados tecidos. Por sua vez, CTs adultas, presentes nos tecidos já desenvolvidos, possuem capacidade de diferenciação mais restrita (multi ou unipotentes), geralmente associada ao seu tecido de residência (WAGERS & WEISSMAN, 2004).

As habilidades de autorrenovação e de responder a diferentes condições fisiológicas nutriram grandes expectativas no uso das CTs na medicina regenerativa. Algumas terapias efetivas utilizando CTs já beneficiam milhões de pessoas, como o transplante de medula óssea no tratamento de leucemias, por exemplo (TROUNSON & MCDONALD, 2015). Entretanto, importantes aspectos devem ser considerados e desafios superados antes que a terapia baseada em CTs se torne uma realidade mais concreta. A pesquisa básica e pré-clínica trabalha continuamente para avançar sobre os desafios da diferenciação direta de CTs, buscando maior eficiência de diferenciação e maturação das células derivadas; da rejeição imune e do risco de desenvolvimento de tumores (revisado por BHAGAVATI, 2015). Avanços consideráveis têm sido alcançados no entendimento do comprometimento das CTs a tecidos específicos, porém a identificação de sinais que regulam esses processos ainda requer uma minuciosa exploração (WAGERS & WEISSMAN, 2004).

1.1.1 Células-tronco pluripotentes

Células-tronco pluripotentes são aquelas capazes de gerar todos os tecidos de um organismo adulto, exceto os anexos embrionários (YOUNG, 2011). As primeiras CTs embrionárias humanas (CTEs) foram isoladas por Thomson e colaboradores em 1998, a partir da massa interna de blastocistos. Essas células podem ser mantidas em cultura sob condições que as permite propagar indefinidamente e manter sua capacidade pluripotente (THOMSON, 1998). Um pequeno grupo de fatores de transcrição, composto por Oct4, Nanog e Sox2, é responsável pela regulação da expressão de genes envolvidos nos mecanismos que controlam a pluripotência e a autorrenovação (BOYER *et al.*, 2005). Eles agem regulando positivamente seus próprios promotores, ativam a expressão de genes responsáveis pela manutenção do estado pluripotente e contribuem para a repressão de genes linhagem-específicos que compõem vias de sinalização de desenvolvimento (BOYER *et al.*, 2005; YOUNG, 2011).

O amplo potencial de diferenciação dessas células trouxe expectativas quanto à sua utilização na terapia celular, onde poderiam ser diferenciadas em laboratório no tipo celular desejado e recuperar a função de um órgão após o implante, ou ainda poderiam ser utilizadas no estudo de mecanismos de desenvolvimento, de doenças e na seleção de drogas (THOMSON, 1998). Entretanto, a utilização de CTEs com fins terapêuticos e de pesquisa ainda enfrenta muitas questões éticas, uma vez que o isolamento dessas células não é possível sem a destruição do embrião (DE MIGUEL-BERIAIN, 2015). Tendo em vista driblar essas questões e ainda possibilitar a obtenção de células paciente-específicas e/ou doença-específicas, Takahashi e colaboradores em 2007 mostraram a bem-sucedida indução do estado pluripotente através da reprogramação de células adultas (TAKAHASHI *et al.*, 2007). Eles promoveram a expressão dos fatores de transcrição Oct3/4, Sox2, Klf-4 e c-Myc através de vetores retrovirais em fibroblastos derivados da derme, e obtiveram células com características similares a células embrionárias pluripotentes quanto à morfologia, proliferação, expressão de marcadores de superfície e expressão gênica. Além disso, essas células foram capazes de se diferenciar nos três folhetos embrionários e formar teratomas (TAKAHASHI *et al.*, 2007). Tais células passaram a ser chamadas de células pluripotentes induzidas (iPSCs – do inglês, *inducible pluripotent stem cells*). Outros grupos de pesquisa também começaram a derivá-las, com variações na fonte

de obtenção das células adultas, no método de entrega dos fatores de reprogramação e nos próprios fatores de reprogramação (revisado por SINGH & KUMAR, 2015).

A possibilidade de geração de células paciente-específicas ampliou o potencial uso das iPSCs na medicina regenerativa, uma vez que reduziria enormemente o risco de rejeição tecidual. Entretanto, a produção de células para cada indivíduo resultaria em um custo muito elevado, tornando-se inviável na prática. Para contornar esse limitante, têm se proposto a criação de um banco de iPSCs derivadas de doadores com diferentes HLA (do inglês, *human leukocyte antigen*), o qual daria cobertura às necessidades de compatibilidade entre doadores e recipientes (INOUE *et al.*, 2014). As iPSCs também são uma importante ferramenta na modelagem de doenças, onde a reprogramação de células provenientes de um paciente permite o estudo dos mecanismos associados àquela determinada doença e possibilita a pesquisa de tratamentos alternativos, como a terapia gênica e a triagem de novos fármacos. Outro desafio é que ainda há muita heterogeneidade nas populações reprogramadas à iPSCs, o que leva à variação na eficiência de diferenciação e maturação das células, provavelmente decorrente de reprogramação incompleta, variabilidade de background genético e memória epigenética. Por este motivo, o controle de qualidade na produção dessas células ainda se faz essencial para o estabelecimento delas como modelo ou fonte terapêutica (INOUE *et al.*, 2014).

1.2 Desenvolvimento cardíaco

Na embriogênese, o coração é o primeiro órgão a ser formado e é essencialmente derivado da mesoderme (BRADE *et al.*, 2013; SYLVA; VAN DEN HOFF; MOORMAN, 2014). Os precursores cardíacos residem em duas áreas simétricas da mesoderme anterior, na divisa com a endoderme. Durante a gastrulação, essas células são as primeiras a migrar através da linha primitiva formando o chamado crescente cardíaco, que por sua vez se funde e forma o tubo cardíaco (BOOGERD; MOORMAN; BARNETT, 2009; BRADE *et al.*, 2013; BRUNEAU, 2013). As células que compõem o tubo cardíaco formam o chamado primeiro campo cardíaco. O crescimento subsequente do tubo cardíaco é decorrente do recrutamento de células da região mesodérmica periférica, chamada de segundo campo cardíaco. As fronteiras entre essas duas populações distintas – primeiro e segundo campo cardíaco – ainda não são bem definidas, assim como os marcadores moleculares para

identificá-las (BARNETT; DEN BOOGAARD; CHRISTOFFELS, 2012; SYLVA; VAN DEN HOFF; MOORMAN, 2014). As câmaras cardíacas se desenvolvem através da expansão de áreas do tubo cardíaco primário, e suas células se tornam mais especializadas a fim de proporcionar contrações fortes e rápidas (SYLVA; VAN DEN HOFF; MOORMAN, 2014). O tubo cardíaco, derivado diretamente do primeiro campo cardíaco, dá origem ao ventrículo esquerdo do coração; o ventrículo direito e o tronco arterial são formados pela migração posterior dos precursores do segundo campo cardíaco, e os átrios têm origem mista (OLSON, 2006; PAIGE *et al.*, 2015) (Figura 1.1).

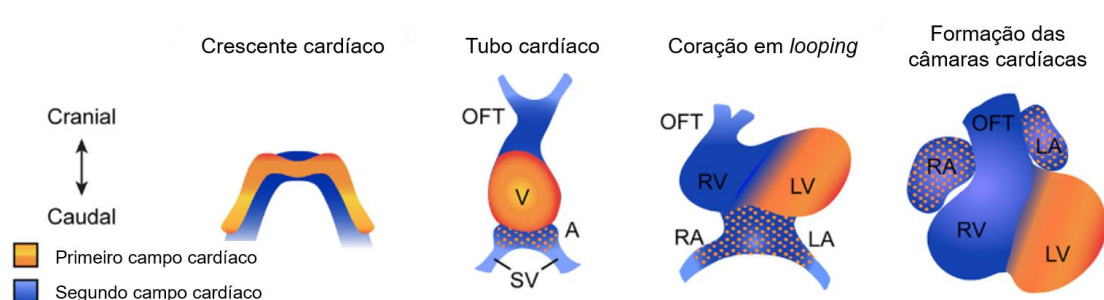


Figura 1.1 Desenvolvimento cardíaco em mamíferos. Durante a gastrulação, células da mesoderme anterior migram através da linha primitiva formando o crescente cardíaco. Progenitores do primeiro (laranja) e segundo (azul) campos cardíacos proliferam e se diferenciam, formando o tubo cardíaco. O ventrículo esquerdo (LV, *left ventricle*) é derivado do primeiro campo cardíaco. O ventrículo direito (RV, *right ventricle*) e o tronco arterial (OFT, *outflow tract*) são derivados do segundo campo cardíaco. Os átrios (LA, *left atrium* e RA, *right atrium*) têm origem mista. SV, *sinus venosus*. Retirado e adaptado de PAIGE *et al.*, 2015.

1.2.1 Sinalização molecular durante o desenvolvimento cardíaco

Sinais providos da endoderme adjacente participam na ativação e na repressão das vias de sinalização de BMP (do inglês, *bone morphogenetic protein*), FGF (do inglês, *fibroblast growth factor*) e WNT (do inglês, *wingless type*), contribuindo para a diferenciação cardíaca (SYLVA; VAN DEN HOFF; MOORMAN, 2014). A indução da mesoderme é primordial no desenvolvimento do coração e é marcada pela expressão do fator de transcrição T-box Brachyury/T, um alvo da sinalização Wnt/b-catenina (BRADE *et al.*, 2013). As vias de sinalização Wnt são classificadas em canônica e não-canônica, e participam de maneira ativa e complexa durante a diferenciação cardíaca. A via canônica envolve a participação da proteína β -catenina, que na ausência de sinalização Wnt interage com um complexo de degradação

composto pelas proteínas axina, APC e GSK3. Na presença de um ligante de Wnt, o complexo é desfeito e a β -catenina fica livre para entrar no núcleo e ativar a transcrição gênica. A ativação desta via é essencial para a formação da mesoderme nos estágios iniciais do embrião (revisado por GESSERT; KÜHL, 2010). Durante a especificação dos precursores cardíacos, em uma etapa mais posterior, a regulação negativa da via Wnt/ β -catenina é essencial. Nesse momento, as vias não-canônicas contribuem para uma correta diferenciação celular. Tais vias envolvem diversos mediadores intracelulares, como a sinalização por JNK (do inglês, *jun-N-terminal kinase*) ou Ca^{2+} (revisado por GESSERT; KÜHL, 2010).

Células Brachyury/T positivas passam, então, a expressar o receptor de fator de crescimento vascular endotelial 2 (VEGFR2, do inglês *vascular endothelial growth factor receptor 2* - ou também conhecido como Flk-1 ou KDR) quando se desenvolvem à uma população mesodermal intermediária, a mesoderme cardíaca (YANG *et al.*, 2008). Essa população têm a expressão de Brachyury/T regulada negativamente com a concomitante ativação da expressão de MESP1 (do inglês, *mesoderm posterior 1*), através da ação do fator de transcrição T-box Eomesodermina (EOMES) (BRADE *et al.*, 2013; COSTELLO *et al.*, 2015). MESP1 atua ativando a rede transcricional cardiogênica e define o comprometimento das células à linhagem cardiovascular (BONDUE *et al.*, 2008).

Os fatores de transcrição NKX2, MEF2, GATA, TBX e HAND são conservados evolutivamente e representam o núcleo que controla a diferenciação cardíaca, ativando a expressão de proteínas contráteis e contribuindo na coordenação da morfogênese das estruturas cardíacas (BARNETT; DEN BOOGAARD; CHRISTOFFELS, 2012; BRUNEAU, 2013; OLSON, 2006). Os genes alvo destes fatores de transcrição vêm sendo identificados e suas contribuições no desenvolvimento cardíaco determinadas. Através da técnica de ChIP-chip em cardiomiócitos murinos, Schlesinger e colaboradores (2011) identificaram uma série de alvos de Gata4, Mef2a, Nkx2-5 e Srf, e a análise de ontologia gênica daqueles mostrou grande associação com desenvolvimento e função do coração (SCHLESINGER *et al.*, 2011). Também foram observados vários genes alvo em comum, incluindo genes de proteínas estruturais como Actn2, Tnnt2 e Myh6, fator de crescimento Igf1 e de transcrição Tbx20. Ainda, há regulação entre os próprios fatores, onde por exemplo, Gata4 e Nkx2-5 aparecem regulados por Mef2a (SCHLESINGER *et al.*, 2011). Esses dados são um exemplo da regulação complexa

e cooperativa protagonizada durante a diferenciação cardíaca. A Figura 1.2 resume alguns dos reguladores chave nas diferentes etapas do desenvolvimento cardíaco.

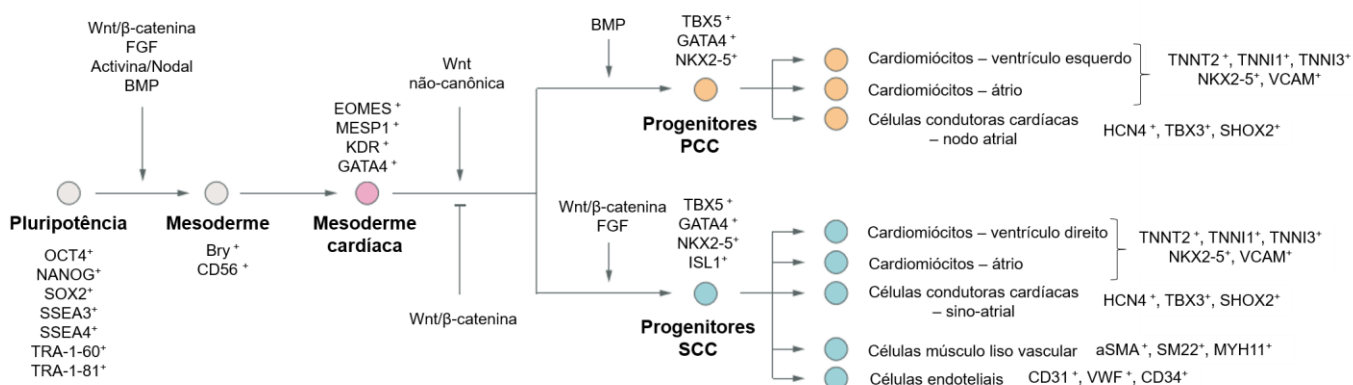


Figura 1.2 Vias de sinalização e marcadores das diferentes etapas da diferenciação cardíaca. PCC, primeiro campo cardíaco. SCC, segundo campo cardíaco. Retirado e adaptado de SPATER *et al.*, 2014, com informações adicionais de BRADÉ *et al.*, 2013 e SKELTON *et al.*, 2017.

1.2.2 Desenvolvimento das câmaras cardíacas

O coração é formado por subestruturas que possuem forma e função específica: septos que garantem a compartimentalização dos sangues arterial e venoso, válvulas que permitem o fluxo unidirecional, câmaras contráteis para a propulsão do sangue e o sino atrial de onde derivam os impulsos elétricos que ativam a contração muscular (BARNETT; DEN BOOGAARD; CHRISTOFFELS, 2012). As câmaras do coração adulto humano são compostas por um átrio e um ventrículo esquerdo e um átrio e um ventrículo direito (BOOGERD; MOORMAN; BARNETT, 2009; SYLVA; VAN DEN HOFF; MOORMAN, 2014).

Cardiomiócitos que formam o tubo cardíaco primário se caracterizam pela baixa taxa de proliferação, automaticidade limitada, velocidade de condução baixa e contração fraca. Conforme essas células se proliferam e seletivamente se diferenciam em atriais e ventriculares, o tubo cardíaco se expande, num processo chamado de balonamento, para a formação das câmaras cardíacas (PAIGE *et al.*, 2015; SYLVA; VAN DEN HOFF; MOORMAN, 2014). As células atriais e ventriculares obtêm a identidade das câmaras do miocárdio através do aumento da contratilidade e automaticidade, velocidade de condução rápida e maior organização dos sarcômeros (PAIGE *et al.*, 2015).

Recentemente foi mostrado que cardiomiócitos atriais e ventriculares derivam de populações distintas da mesoderme, e que, portanto, a determinação de sua identidade acontece precocemente no desenvolvimento (LEE *et al.*, 2017). Fatores determinantes no estabelecimento dessas células ainda não são bem esclarecidos, entretanto alguns candidatos vêm sendo propostos (BRUNEAU, 2013). Genes da família NKX se mostraram capazes de induzir a expressão de genes alvo que participam da determinação de cardiomiócitos ventriculares, além de participarem da funcionalidade do tecido adulto (GEORGE; COLOMBO; TARGOFF, 2015). O fator natriurético atrial, transcrito pelo gene NPPA, é expresso nas células miocárdicas e também está relacionado com a determinação da identidade destas. Sua expressão é regulada pelos fatores de transcrição Nkx2-5, Gata4 e Tbx5, indicando que estes também estão envolvidos na determinação atrioventricular (BARNETT; DEN BOOGAARD; CHRISTOFFELS, 2012).

Os fatores de transcrição T-box também participam na regulação da identidade dos cardiomiócitos. Tbx5 e Tbx20 são importantes ativadores do programa gênico das câmaras cardíacas, composto pelos genes das conexinas 40 e 43, fatores de transcrição Hand1, Irx4, Irx5 e Nppa (BOOGERD; MOORMAN; BARNETT, 2009; SYLVA; VAN DEN HOFF; MOORMAN, 2014). Por outro lado, Tbx2 e Tbx3 têm papel de repressores do miocárdio, mantendo as células com características primárias e que irão compor o polo e o trato venoso, o canal atrioventricular e o trato arterial (BOOGERD; MOORMAN; BARNETT, 2009; SYLVA; VAN DEN HOFF; MOORMAN, 2014). Tbx3 permite, então, a expressão do gene *HCN4*, que transcreve para um canal de potássio, característico das células sinoatriais e que participa da geração do potencial de ação determinante para o batimento cardíaco (PAIGE *et al.*, 2015; SYLVA; VAN DEN HOFF; MOORMAN, 2014).

O ácido retinóico também possui um importante papel na sinalização do comprometimento cardíaco. Xavier-Neto e colaboradores (1999) mostraram que a exposição de embriões murinos a altos níveis de ácido retinóico levou a má-formações cardíacas severas, caracterizadas por átrios alargados e ventrículos reduzidos (XAVIER-NETO *et al.*, 1999). Também já foi demonstrado a possibilidade de gerar *in vitro* células com características atriais, a partir de células pluripotentes humanas, através da modulação da sinalização por ácido retinóico (DEVALLA *et al.*, 2015; LEE *et al.*, 2017; ZHANG *et al.*, 2011). Portanto, essas e outras evidências levam a se acreditar que o ácido retinóico participa da determinação de células atriais.

As células dos átrios e ventrículos são distintas em sua morfologia, funcionalidade e características moleculares. Propriedades eletrofisiológicas e de contração muscular diferenciam essas células e, dentre os genes que determinam essas diferenças, apenas um grupo seletivo já foi identificado (NG; WONG; TSANG, 2010; WU *et al.*, 2013).

O potencial de ação das células musculares cardíacas é responsável pela propagação do impulso nervoso e contração muscular. A despolarização da membrana é mantida por alguns milissegundos antes da repolarização, fase que constitui o chamado platô. Nesta etapa, o tempo de duração do platô é a principal característica que distingue células atriais e ventriculares. Células ventriculares apresentam uma fase de platô mais longa que as células atriais, fazendo com que contraíam e relaxem de maneira mais lenta (BRANDENBURG *et al.*, 2016; NG; WONG; TSANG, 2010) (Figura 1.3).

A expressão diferencial de fatores de transcrição e proteínas estruturais também distingue células atriais e ventriculares. Os fatores de transcrição *Irx4* e *Hey2* são expressos exclusivamente nos ventrículos e conferem sua identidade ativando genes ventrículo-específicos e reprimindo genes átrio-específicos (EVANS *et al.*, 2010; PAIGE *et al.*, 2015). O receptor nuclear COUP-TFII, transcrito pelo gene *NR2F2*, se mostrou essencial na identidade atrial através da ligação direta e modulação da expressão de genes característicos das células atriais (WU *et al.*, 2013). COUP-TFII atua regulando genes de canais de íons, como *KCNA5* e *KCNJ3*, responsáveis pelas características eletrofisiológicas únicas das células atriais (DEVALLA *et al.*, 2015). Proteínas envolvidas na contração muscular, como as miosinas, também apresentam padrão de expressão diferencial entre as células atriais e ventriculares. Nos átrios há o predomínio da expressão das isoformas α -MHC (transcrita pelo gene *MYH6*) e MLC-2a (transcrita pelo gene *MYL7*), a qual inicialmente é expressa em todo tubo cardíaco primário e posteriormente sua expressão é exclusivamente atrial. Por sua vez, os ventrículos se caracterizam pela expressão das miosinas β -MHC (transcrita pelo gene *MYH7*) e MLC-2v (transcrita pelo gene *MYL2*) (DEVALLA *et al.*, 2015; NG; WONG; TSANG, 2010). A Figura 1.3 mostra alguns genes que são expressos diferencialmente entre átrios e ventrículos e características eletrofisiológicas típicas.

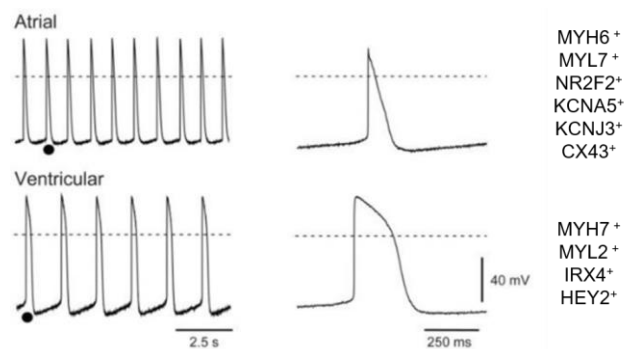


Figura 1.3 Características eletrofisiológicas e marcadores de células atriais e ventriculares. Retirado e adaptado de MUMMERY *et al.*, 2012, com informações complementares de DEVALLA *et al.*, 2015 e SKELTON *et al.*, 2017.

1.3 Diferenciação cardiomiogênica a partir de células-tronco pluripotentes

O isolamento de CTEs humanas (THOMSON, 1998) e, mais recentemente, a reprogramação de células pluripotentes induzidas (TAKAHASHI *et al.*, 2007) motivou os pesquisadores a guiarem seus esforços no desenvolvimento de novas estratégias para direcionar a diferenciação de linhagens cardíacas mais eficientemente e de maneira reprodutível. Frente à possibilidade de gerar cardiomiócitos a partir dessas células para reconstruir um coração, por exemplo, avanços são necessários com respeito a produção de células em uma escala clinicamente relevante (BURRIDGE *et al.*, 2012).

Cardiomiócitos gerados a partir de CT pluripotentes (CTP) representam uma fonte ilimitada para reposição ou terapia celular. Muitas doenças cardíacas têm como último recurso o transplante do órgão, entretanto, há um número muito restrito de doadores. A produção de cardiomiócitos *in vitro* aparece como uma alternativa, por isso a necessidade de estabelecer uma metodologia eficiente de produção e transplante dessas células. O descobrimento de drogas para tratamento de doenças cardíacas e os testes de desenvolvimento e toxicidade cardíaca de novas drogas são processos caros e árduos que podem ser facilitados com a utilização de cardiomiócitos em laboratório. Além disso, o modelamento de doenças cardíacas, de desenvolvimento embrionário e até medicina personalizada são ramos de aplicação de cardiomiócitos derivados de CTP. A fim de concretizar a potencial aplicação das CTPs, é essencial sermos capazes de controlar a diferenciação e direcionar o desenvolvimento dessas células por vias específicas (BOLLINI; SMART; RILEY, 2011;

BURRIDGE *et al.*, 2012; MURRY & KELLER, 2008; VIDARSSON; HYLLNER; SARTIPY, 2010).

A diferenciação cardiomiogênica a partir de CTPs recapitula muitos aspectos do desenvolvimento embrionário (BURRIDGE *et al.*, 2012; MUMMERY *et al.*, 2012). As principais vias de sinalização reguladas durante o desenvolvimento cardíaco têm sido utilizadas na indução de CTP à cardiomiócitos *in vitro*. Diferentes metodologias e protocolos foram desenvolvidos e vêm sendo aprimorados, os quais utilizam fatores de crescimento e pequenas moléculas a fim de modular as vias de BMP, TGF- β /activina/NODAL, WNT e FGF em janelas específicas durante a diferenciação (BURRIDGE *et al.*, 2012; MUMMERY *et al.*, 2012; SPATER *et al.*, 2014). Dois desses métodos são amplamente utilizados e, inclusive, foram utilizados no presente trabalho: (1) formação de corpos embrióides (CE) (KATTMAN *et al.*, 2011) e (2) diferenciação em monocamada (LIAN *et al.*, 2012) (Figura 1.4).

No protocolo de diferenciação por formação de CE, as células são levemente desagregadas e cultivadas em suspensão a fim de que formem um agregado estável, chamado então de CE. São utilizados os fatores BMP4, FGF e activina A no início da diferenciação a fim de se induzir a formação de mesoderme (entre os dias 1 e 4). Posteriormente, a inibição da via Wnt é feita através do fator DKK1 ou de um composto chamado XAV939 (WANG; HAO; HONG, 2011) (adaptação feita pelos nossos colaboradores), que juntamente com VEGF induz o comprometimento a progenitores cardíacos (entre os dias 4 a 8). Em torno do dia 10 da diferenciação é possível observar CE pulsantes (KATTMAN *et al.*, 2011; YANG *et al.*, 2008).

Por sua vez, o protocolo em monocamada permite a diferenciação das células aderidas, o que possibilita uma melhor difusão dos indutores entre as células comparado com os CE (MUMMERY *et al.*, 2012). Nesse método, a modulação da via Wnt é o principal componente que conduz a diferenciação cardiomiogênica (LIAN *et al.*, 2012). No dia inicial, o composto CHIR99021 é utilizado a fim de inibir a proteína GSK3, a qual participa do complexo de degradação de β -catenina, permitindo então que esta fique atuante na célula e a via Wnt se ative. A ativação da via Wnt leva à indução a mesoderme. Em uma etapa posterior, a via Wnt deve ser inibida para uma maior eficiência da diferenciação, o que é feito através da utilização do composto IWP-2 (LIAN *et al.*, 2012) ou do XAV939 (adaptação feita pelo autor).

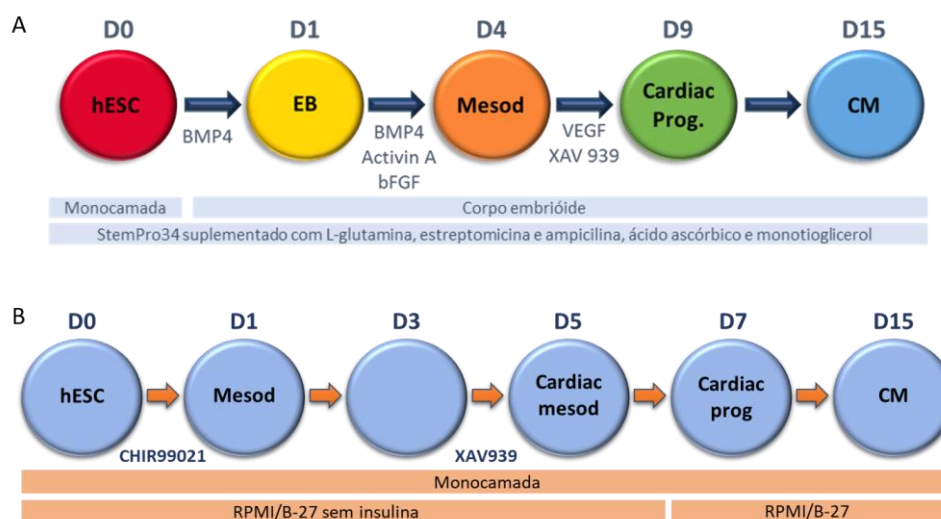


Figura 1.4 Diferenciação cardiomiogênica *in vitro* a partir de CTP humanas. Protocolos de diferenciação (A) por corpos embrioides (EB, *embryoid body*) (KATTMAN *et al.*, 2011) e (B) em monocamada (LIAN *et al.*, 2012). Dias de diferenciação, meio base de cultivo, fatores e pequenas moléculas usadas para indução estágio-específica da diferenciação até cardiomiócitos estão indicados. hESC, *human embryonic stem cells*. Mesod, mesoderme. Cardiac Prog, progenitores cardíacos. CM, cardiomiócitos. Fonte: o autor.

O sucesso da diferenciação geralmente é avaliado através do número de CE pulsantes ou pela área pulsante em monocamada, e pela expressão da troponina cardíaca T (cTnT, transcrita pelo gene TNNT2) (BURRIDGE *et al.*, 2012; MUMMERY *et al.*, 2012). Entretanto, outros marcadores podem ser usados a fim de monitorar as etapas progressivas da diferenciação (BURRIDGE *et al.*, 2012). As mudanças sequenciais na expressão gênica que acontecem durante o desenvolvimento cardíaco também são observadas durante a diferenciação *in vitro*, e podem ser usadas como marcadores (MUMMERY *et al.*, 2012; SKELTON *et al.*, 2017). Alguns destes marcadores já foram mencionados no tópico anterior (Figura 1.2) e podem ser avaliados através de diferentes metodologias, como RT-PCR, imunofluorescência e citometria de fluxo.

As células diferenciadas *in vitro* podem apresentar diferentes níveis de maturação e, geralmente, quanto maior o seu tempo em cultura, mais se assemelham a células adultas (LUNDY *et al.*, 2013; SARTIANI *et al.*, 2007). Essas características de maturação podem ser, então, avaliadas por microscopia eletrônica, uma vez que permite a visualização das ultraestruturas típicas das células musculares do coração, como os sarcômeros e o grande número de mitocôndrias (LUNDY *et al.*, 2013; MUMMERY *et al.*, 2012). Ensaios de eletrofisiologia e medição do potencial de ação

das células também são técnicas que permitem inferir sobre a maturação e o subtipo das células cardíacas, além de fornecerem evidências sobre a presença de cardiomiócitos funcionais (MUMMERY *et al.*, 2012; ROBERTSON; TRAN; GEORGE, 2013). Desta maneira, tanto técnicas moleculares quanto funcionais podem ser associadas a fim de se determinar a eficiência da diferenciação *in vitro*.

1.4 Regulação da expressão gênica

Os níveis finais de qualquer proteína dependem dos diversos passos do processo de expressão gênica, onde muitos mecanismos regulatórios estão envolvidos dentro de uma rede de sinalização que reflete condições celulares específicas (MATOULKOVA *et al.*, 2012). A expressão gênica envolve alguns processos celulares principais: transcrição, processamento do mRNA, sua exportação, tradução e degradação (KOMILI; SILVER, 2008; SCHWANHÄUSSER *et al.*, 2011). Todas estas etapas são extensivamente reguladas e coordenadas entre si (revisado por KOMILI; SILVER, 2008).

O estado transcricional de um gene é dependente, entre outros, da estrutura da cromatina ao seu redor, da ação de elementos regulatórios em *cis*, como promotores e acentuadores (do inglês, *enhancers*), e de fatores de transcrição (KOMILI; SILVER, 2008; SPITZ; FURLONG, 2012). O acesso da maquinaria de transcrição ao DNA genômico pode ser modulado através das modificações pós-traducionais de histonas e/ou substituição por variantes, as quais compõem os nucleossomos, a fim de converter a cromatina de um estado mais fechado a um mais aberto (revisado por IGLESIAS-PLATAS; MONK, 2016). Por sua vez, fatores de transcrição são capazes de se ligar a regiões regulatórias e ativar ou inibir a transcrição, promovendo ou inibindo o recrutamento da RNA polimerase (KOMILI; SILVER, 2008; SAINSBURY; BERNECKY; CRAMER, 2015).

Três diferentes RNA polimerases são responsáveis pela transcrição nas células eucariotas. A RNA polimerase II transcreve genes codificantes de proteínas, gerando mRNAs, e também alguns genes não-codificantes (BUNCH, 2018; SAINSBURY; BERNECKY; CRAMER, 2015). A regulação da ação da RNA polimerase II influencia diretamente importantes processos celulares como a diferenciação celular, a manutenção da identidade das células e as respostas às mudanças no ambiente. Essa regulação acontece em diferentes etapas da transcrição, entretanto, o

mecanismo chave de controle é durante a iniciação (revisado por SAINSBURY; BERNECKY; CRAMER, 2015).

A transcrição pervasiva e promíscua da RNA polimerase II vêm sendo descrita (DJEBALI *et al.*, 2012; PORRUA; LIBRI, 2015; SPITZ; FURLONG, 2012) e está envolvida na expressão gênica estocástica, caracterizada pela flutuação aleatória nos níveis de mRNA (RAJ *et al.*, 2006). Esses escapes de transcrição seriam também decorrentes da atividade de modeladores da cromatina, levando a uma lenta conversão entre os estados ativado e desativado de um gene, e, ainda, pela abundância de determinados fatores de transcrição e número de sítios de ligação dos mesmos (KOMILI; SILVER, 2008; RAJ *et al.*, 2006; RASER; O'SHEA, 2005). Tais observações sugerem que a regulação transcricional pode ser um processo bastante permeável (SPITZ; FURLONG, 2012), e que a expressão gênica estocástica poderia ser tamponada através da regulação pós-transcricional, atribuindo ao controle da síntese proteica o ajuste da expressão gênica final (RAJ *et al.*, 2006).

1.4.1 Regulação pós-transcricional da expressão gênica

Tradicionalmente, a regulação da transcrição era considerada o principal nível de regulação da expressão gênica, por razões históricas e técnicas. O controle transcricional é o mais básico e intuitivo passo da expressão gênica, estudado por técnicas bem estabelecidas (MATA; MARGUERAT; BÄHLER, 2005). Contudo, o progressivo aprofundamento no conhecimento sobre a molécula de RNA tem permitido o reconhecimento de outros importantes mecanismos de regulação (MATOULKOVA *et al.*, 2012).

A regulação pós-transcricional da expressão gênica possibilita uma resposta mais rápida que a regulação da transcrição, por exemplo, a hormônios, fatores de crescimento e nutrientes (KELEN *et al.*, 2009) ou a condições de estresse e dano ao DNA (SUNNERHAGEN, 2007). A regulação pós-transcricional compreende diversas etapas desde a síntese do transcrito primário até a degradação da molécula de mRNA, incluindo o capeamento 5', *splicing*, clivagem 3' e poliadenilação, edição de RNA, exportação nuclear, localização citoplasmática, início e alongação da tradução e processos de decaimento de mRNA (KELEN *et al.*, 2009; SUNNERHAGEN, 2007).

Considerando que o mRNA é um intermediário nas mudanças fenotípicas em uma célula, agindo como mensageiro para a síntese de proteína, os mecanismos que regulam a tradução do mRNA são altamente controlados. Entre eles está o

acoplamento do mRNA aos ribossomos e o início da tradução. No citoplasma, o cap 5' e a cauda poli-A dos mRNAs maduros são reconhecidos pelos fatores de iniciação eucarióticos (eIF), formando um complexo inicial juntamente com o RNA transportador acoplado ao aminoácido metionina (Met-tRNA) e a subunidade ribossomal 40S, o qual é chamado de complexo de pré iniciação 43S. Este complexo percorre a região 5'-UTR (UTR – do inglês, *untranslated region*) até encontrar o códon AUG, recrutando a subunidade ribossomal 60S e formando o complexo 80S, dando início ao processo de alongação da tradução. Além disso, a subunidade 40S pode ser recrutada diretamente através de elementos da região 5'-UTR chamados de sítios de iniciação interna (IRES – do inglês, *Internal Ribosome Entry Sites*), não dependendo do reconhecimento do cap 5' (revisado por HELLEN & SARNOW, 2001). Alguns mecanismos conhecidos por regularem a tradução envolvem a fosforilação do fator de iniciação eIF2a, a ligação da 4E-BP ao fator de iniciação eIF4E, a presença de quadros de leitura na região 5' (uORFs - do inglês *upstream open reading frames*) antes da sequência codificante (KELEN *et al.*, 2009; SUNNERHAGEN, 2007) e o ajuste do número de ribossomos disponíveis (MATA; MARGUERAT; BÄHLER, 2005).

Durante a alongação da tradução, os mRNAs podem se associar a vários ribossomos em uma estrutura chamada de polissomo. Esses complexos são abundantes e possuem alto peso molecular, podendo ser facilmente separados de outras frações celulares, como as de RNA livre de ribossomos (SUNNERHAGEN, 2007). Mudanças nas taxas de início de tradução resultam na alteração do carregamento de mRNA nos ribossomos. Nessa perspectiva, a ativação traducional pode ser indicada pela redistribuição de moléculas de RNA da fração livre de ribossomos para a fração associada a polissomos. A mobilização diferencial de mRNA para os polissomos pode ser usada para identificar mRNAs sujeitos à regulação traducional em determinado contexto celular (MIKULITS *et al.*, 2000).

Características intrínsecas da molécula de mRNA, como as regiões não codificantes (UTRs), têm papel fundamental na regulação da expressão gênica. Essas regiões representam elementos regulatórios importantes na regulação pós-transcricional, uma vez que apresentam sítios de ligação de fatores capazes de recrutar ou excluir maquinarias de degradação, reduzindo ou aumentando a estabilidade da molécula de mRNA, além de poder determinar sua localização e eficiência de tradução influenciando o nível total de proteína sintetizada (MATOULKOVA *et al.*, 2012). Existem dois grupos principais de elementos de ligação

ao mRNA: as proteínas de ligação ao RNA (RBPs - do inglês *RNA-binding proteins*) e os RNAs não codificantes (revisado por WU & BREWER, 2012).

1.4.1.1 Proteínas de ligação ao RNA

As RBPs interagem com o RNA e podem tanto atuar diretamente sobre ele ou recrutar a ação de outras proteínas ou complexos proteicos (GEHRING; WAHLE; FISCHER, 2017; WU; BREWER, 2012). Elas podem reconhecer tanto uma sequência nucleotídica, quanto estruturas secundárias do RNA alvo. Entre os domínios característicos de ligação ao mRNA estão o motivo de reconhecimento de RNA (RRM – do inglês *RNA recognition motif*), domínio de homologia K (KH, do inglês *K-homology domain*), dedos de zinco (do inglês, *zinc fingers*), DEAD-box, motivo de reconhecimento de RNA dupla fita (GEHRING; WAHLE; FISCHER, 2017) e domínios de choque térmico (CSD – do inglês *cold shock domain*) (MIHAILOVICH *et al.*, 2010).

De maneira geral, as RBPs são classificadas em três grupos, de acordo com a diversidade dos seus alvos: proteínas core comuns, que fazem parte de mecanismos gerais de processamento dos mRNAs; proteínas específicas, que reconhecem determinadas sequências ou estruturas de um grupo restrito de RNAs, ligando-se ao transcrito com alta afinidade; e as proteínas não-específicas, que possuem uma especificidade baixa, reconhecendo uma maior variedade de transcritos e sua associação pode ser direcionada através da interação com outras proteínas ou complexos proteicos a fim de proporcionar maior especificidade (GEHRING; WAHLE; FISCHER, 2017).

Os complexos formados por RBPs e mRNAs são chamados de complexos de ribonucleoproteínas mensageiras (mRNP) e atuam controlando muitos aspectos do metabolismo e função dos mRNAs, como o processamento, exportação, localização, estabilidade e tradução (KEENE, 2007; MATA; MARGUERAT; BÄHLER, 2005). Tais complexos de manutenção e degradação de RNA podem se concentrar em regiões do citoplasma, onde formam os chamados grânulos de mRNP. Esses grânulos geralmente possuem mRNAs traducionalmente silenciados (BUCHAN, 2014), e estão em constante interação com os polissomos, permitindo que os mRNAs transitem entre essas estruturas e participando da regulação traducional dos mesmos (BRENGUES; TEIXEIRA; PARKER, 2005). Os grânulos podem, ainda, ser classificados como de estresse e de processamento (*P bodies*). Estudos sugerem que estas duas estruturas atuem na manutenção do mRNA, sendo que os grânulos de estresse teriam como papel principal a triagem, seleção e remodelamento de mRNP, enquanto que os *P*

bodies seriam os principais responsáveis pela degradação do mRNA (KEDERSHA *et al.*, 2005).

De acordo com a teoria do “RNA-operon”, mRNAs funcionalmente relacionados são coordenadamente regulados por fatores de ação *trans* – primariamente as RBPs, mas também RNAs não-codificantes e metabólitos (GEHRING; WAHLE; FISCHER, 2017; KEENE, 2007). Esses fatores interagem com múltiplos elementos regulatórios dos mRNAs e podem regular uma mesma molécula de mRNA cooperando ou competindo por um sítio. O modelo prediz uma alta organização e dinâmica de regulação pós-transcricional de mRNAs cuja função seja relacionada, por exemplo, que participam de uma mesma via de sinalização. Essa coordenação garantiria um uso eficiente e flexível da informação genética (KEENE, 2007), e adicionaria mais um nível de complexidade na regulação da expressão gênica.

1.4.1.2 RNAs não-codificantes

O termo RNA não-codificante (ncRNA) é comumente empregado a RNAs que não codificam proteínas, entretanto, não significando que não possuem informação ou função (MATTICK; MAKUNIN, 2006). Com o advento das novas tecnologias de isolamento e identificação de RNAs, o catálogo de tipos e funções biológicas dessas moléculas vêm crescendo consideravelmente (DJEBALI *et al.*, 2012). O genoma humano comporta pouco mais de 20.000 genes codificantes de proteínas anotados, os quais ocupam menos de 2% da sequência genômica. Entretanto, há evidências de que cerca de 85% do genoma seja transcrito (DUNHAM *et al.*, 2012; HANGAUER; VAUGHN; MCMANUS, 2013).

Os ncRNAs possuem funções estruturais, como os RNAs ribossomais (rRNA) e RNAs transportadores (tRNA) – os quais estão envolvidos na tradução de proteínas; RNAs nucleares pequenos (snRNA) – envolvidos no *splicing*; e os RNAs nucleolares pequenos (snoRNA) – responsáveis pelo processamento e modificação dos rRNAs (revisado por CECH; STEITZ, 2014). Outros ncRNAs estão envolvidos com a regulação da expressão gênica e são classificados de acordo com seu tamanho em pequenos e longos. Os pequenos ncRNAs incluem os microRNAs (miRNAs), RNAs associados a PIWI (piRNAs) e pequenos RNAs de interferência (siRNA) (CARTHEW; SONTHEIMER, 2009; CECH; STEITZ, 2014).

Os miRNAs são moléculas de aproximadamente 22 nucleotídeos derivadas de precursores em dupla fita ou que formam uma estrutura de “grampo de cabelo” (do inglês *hairpin*) (MATTICK; MAKUNIN, 2006). Transcritos primários (pri-miRNAs)

longos são clivados no núcleo pelo complexo microprocessador, que inclui a enzima RNase Drosha, gerando o precursor pre-miRNA. Por sua vez, este é deslocado ao citoplasma onde sofre uma nova clivagem pela enzima Dicer, para gerar duas moléculas de miRNAs maduras (CARTHEW; SONTHEIMER, 2009; CECH; STEITZ, 2014). Os processos de biogênese dos miRNAs também podem ser regulados e contribuir no controle da expressão gênica (CECH; STEITZ, 2014).

A especificidade da ação dos miRNAs está nas interações por pareamento de bases. Moléculas maduras de miRNAs se associam ao complexo de silenciamento induzido por RNA (RISC) e proporcionam a regulação pós-transcricional dos genes alvo, os quais possuem sítios de ligação geralmente na região 3'UTR. Os miRNAs reprimem a tradução através do pareamento imperfeito com o mRNA alvo ou provocam a degradação deste no caso do pareamento ser perfeito (CARTHEW; SONTHEIMER, 2009; CECH; STEITZ, 2014; MATTICK; MAKUNIN, 2006). Os detalhes de como esses mecanismos funcionam foram revisados por CARTHEW & SONTHEIMER, 2009 e JONAS & IZAURRALDE, 2015.

A regulação da expressão gênica por miRNAs está envolvida em uma série de processos biológicos, inclusive no desenvolvimento embrionário, diferenciação celular e processos patológicos (CECH; STEITZ, 2014; MATTICK; MAKUNIN, 2006). É estimado que cerca de um terço dos genes codificantes de proteínas sejam controlados por miRNAs (MATTICK; MAKUNIN, 2006). Ainda, um único mRNA pode conter sítios de ligação para vários miRNAs diferentes, sugerindo uma complexa rede regulatória (GUIL; ESTELLER, 2015).

Por sua vez, os siRNAs promovem a degradação dos mRNAs alvo pelo mesmo mecanismo dos miRNAs, entretanto são derivados de moléculas de RNA dupla fita tanto de origem endógena quanto exógena. Os piRNAs participam da repressão da transcrição através do recrutamento de proteínas modificadoras da cromatina (revisado por CECH; STEITZ, 2014). Os ncRNAs longos serão discutidos no tópico a seguir.

1.5 RNAs não-codificantes longos (lncRNAs)

Durante a última década, os RNAs não codificantes longos (lncRNAs) têm sido amplamente estudados e sua importância biológica consolidada (BUNCH, 2018). Os lncRNAs são caracterizados por serem maiores que 200 nucleotídeos (PIRCHER;

GEBETSBERGER; POLACEK, 2014), são tipicamente transcritos pela RNA polimerase II e compartilham muitas características similares aos mRNAs, como por exemplo o capeamento 5', a poliadenilação e podem sofrer *splicing* (PIRCHER; GEBETSBERGER; POLACEK, 2014; ULITSKY; BARTEL, 2013). Dentre os critérios que tentam distinguir genes não-codificantes dos codificantes estão a procura por padrões de sequência codificante conservada entre espécies, domínios proteicos conhecidos, similaridades a proteínas conhecidas e potenciais fases abertas de leitura (ORFs - do inglês *open reading frames*) (KARAPETYAN *et al.*, 2013; ULITSKY; BARTEL, 2013).

Os lncRNAs podem ser classificados de acordo com sua proximidade a genes codificantes. São eles: *sense* – quando compartilham um ou mais éxons com um gene codificantes; *antisense* – quando compartilham um ou mais éxons de um gene codificante, mas na fita oposta; bidirecionais ou divergentes – quando têm a expressão iniciada na região próxima e oposta do gene vizinho codificante; intrônico – localizado no íntron de outro transcrito; e intergênico (lincRNA) – transcrito de uma região independente e entre genes (ANGRAND *et al.*, 2015).

Os lncRNAs divergentes e os *antisense* são as duas maiores classes de lncRNAs no genoma humano, cada um compreendendo cerca de 20% do total de lncRNAs (LUO *et al.*, 2016). Suas funções ainda não são bem compreendidas, mas acredita-se que eles atuem regulando seus genes vizinhos codificantes (JOHNSSON *et al.*, 2014; LUO *et al.*, 2016). Inúmeros pares de lncRNAs divergentes e genes codificantes já foram descritos como especificamente co-expressos, e a regulação entre eles comprovada em alguns exemplos (LUO *et al.*, 2016). Ainda, estima-se que 20% a 40% dos genes codificantes de proteínas tenham a transcrição de um RNA *antisense*. Esses ncRNAs compartilham similaridades na sequência com seus transcritos senso, conferindo a possibilidade de regulação pós-transcricional (revisado por JOHNSSON *et al.*, 2014).

A identificação massiva de lncRNAs nos últimos anos expandiu o conhecimento sobre suas possíveis funções na célula. Essas moléculas participam de processos essenciais, como na embriogênese, ciclo celular, manutenção de pluripotência, apoptose e diferenciação (PIRCHER; GEBETSBERGER; POLACEK, 2014). Os lncRNAs são expressos de maneira mais tecido-específica que os genes codificantes, entretanto, com níveis de expressão cerca de 10 vezes mais baixa (CABILI *et al.*, 2011). Acredita-se que os lncRNAs participam da regulação da expressão gênica em

diferentes níveis, incluindo a organização da cromatina, a regulação transcricional e também em processos pós-transcricionais (ANGRAND *et al.*, 2015).

A regulação da transcrição em *cis* atua em genes vizinhos ao que transcreve para o lncRNA. Já a atuação em *trans* visa a regulação de genes distantes na sequência gênica (BUNCH, 2018; KARAPETYAN *et al.*, 2013). Os primeiros lncRNAs descritos funcionalmente eram localizados no núcleo e participavam do recrutamento de proteínas modificadoras da cromatina (ANGRAND *et al.*, 2015). O exemplo melhor conhecido é do gene XIST (do inglês, *X-inactive specific transcript*), que é responsável por inativar em *cis* o cromossomo X em células femininas (PENNY *et al.*, 1996) através do recrutamento do complexo repressivo Polycomb (PRC) (revisado por BUNCH, 2018).

Muitos trabalhos, inclusive do nosso grupo, têm demonstrado a capacidade dos lncRNAs de associação com polissomos (CARLEVARO-FITA *et al.*, 2016; DALLAGIOVANNA *et al.*, 2017; INGOLIA; LAREAU; WEISSMAN, 2011; VAN HEESCH *et al.*, 2014), entretanto esta característica não garante seu potencial codificante. Guttman e colaboradores (2013) demonstraram que os lncRNAs que ocupavam ribossomos seguiam o mesmo padrão de outros transcritos sabidamente não codificantes e que, portanto, eles não teriam função na codificação de proteínas (GUTTMAN *et al.*, 2013). Por outro lado, estudos têm demonstrado que muitos lncRNAs são escaneados por ribossomos e tem potencial de serem traduzidos em pequenos peptídeos. Muitos lncRNAs têm ORFs maiores que 24 aminoácidos, que poderiam corresponder a proteínas funcionais. Entretanto, pequenos peptídeos são mais difíceis de detectar e podem estar sendo subestimados (RUIZ-ORERA *et al.*, 2014).

Independente do potencial codificante, a associação de lncRNAs com ribossomos pode ter funções além do produto da tradução (ULITSKY; BARTEL, 2013). A possibilidade de regulação fina da velocidade e especificidade da maquinaria de tradução é uma das suas possíveis atribuições (PIRCHER; GEBETSBERGER; POLACEK, 2014). Sugere-se que lncRNAs associados a ribossomos poderiam estar os sequestrando e regulando seu uso. Ainda, lncRNAs poderiam parear especificamente com mRNAs e promover ou inibir sua tradução (revisado por PIRCHER; GEBETSBERGER; POLACEK, 2014 e KARAPETYAN *et al.*, 2013). A habilidade dos lncRNAs em parearem com sequências complementares possibilita a regulação de mRNAs em diferentes níveis além da tradução, como também o *splicing*, o transporte e a degradação (ANGRAND *et al.*, 2015).

Além disso, os lncRNAs têm a capacidade de formar estruturas secundárias que podem participar ativamente da sua função. O recrutamento de proteínas parceiras formando complexos ribonucleoproteicos amplia enormemente sua atuação (KARAPETYAN *et al.*, 2013). Alguns lncRNAs podem agir recrutando fatores de transcrição para promotores gênicos ou atuar como co-fatores. Também podem controlar negativamente a transcrição através do sequestro desses fatores (ANGRAND *et al.*, 2015). Os lncRNAs podem também ser precursores de miRNAs ou então funcionarem como inibidores, uma vez que podem conter inúmeros sítios de ligação de miRNAs e sequestram estes da atuação em seus mRNA alvo, sendo chamados de esponjas de miRNA (GUIL; ESTELLER, 2015; KARAPETYAN *et al.*, 2013).

Estudos têm demonstrado que a conservação de estrutura e sequência gênica é rara entre ortólogos e loci equivalentes em diferentes organismos (revisado por JOHNSON *et al.*, 2014). Apenas 12% dos lncRNAs intergênicos (lincRNAs) foram achados expressos em outras espécies de vertebrados (CABILI *et al.*, 2011). Um estudo com 16 espécies de vertebrados, separadas por mais de 50 milhões de anos de divergência evolutiva, mostrou que mais de 70% dos lncRNAs não possuem ortólogos com similaridades de sequência (HEZRONI *et al.*, 2015). Além disso, comparações entre lncRNAs de humanos e camundongos mostraram menos de 70% de identidade, níveis de conservação parecidos com regiões intrônicas (PANG; FRITH; MATTICK, 2006).

Sequências dos éxons de lncRNAs evoluem muito mais rápido que sequências de genes codificantes ou UTRs de mRNAs, sugerindo que muitos lncRNAs não sejam funcionais ou que suas funções não exigem sequências específicas (HEZRONI *et al.*, 2015). A equivalência funcional de lncRNAs entre espécies pode ser possível, uma vez que as estruturas secundárias e as afinidades de interação entre proteínas e outros RNAs não dependem unicamente da sequência dos transcritos (JOHNSON *et al.*, 2014). Essas características suportam o modelo de que haja conservação funcional de lncRNAs entre mamíferos, e que tais funções exigem apenas pequenas sequências específicas e tolerem grandes mudanças na arquitetura gênica (HEZRONI *et al.*, 2015).

Apesar da descoberta de algumas funções específicas de lncRNAs individuais, ainda é um desafio generalizar essas descobertas aos mais de 80.000 lncRNAs preditos (alta confiança) (VOLDERS *et al.*, 2015) no genoma humano (CABILI *et al.*, 2011). Informações detalhadas de uma pequena porção de lncRNAs foram descritas

até então, e o papel biológico e os mecanismos de regulação de inúmeros lncRNAs já anotados ainda são desconhecidos (BUNCH, 2018). Por isso, a caracterização funcional de lncRNAs e a inferência dos processos biológicos em que estão envolvidos são de imensurável importância no entendimento da complexidade do genoma e na regulação mediada por RNA (BUNCH, 2018; LUO *et al.*, 2016).

1.5.1 lncRNAs no desenvolvimento cardíaco

A diferenciação de células cardíacas requer a ativação de programas genéticos específicos e altamente regulados (OLSON, 2006), conforme mencionado anteriormente. Estudos recentes têm demonstrado que a expressão de lncRNAs contribui para a especificação da identidade celular através da regulação da expressão gênica em diferentes níveis (revisado por LOPEZ-PAJARES, 2016). A identificação de novos lncRNAs e sua especificidade tecidual podem fazer com que representem uma nova classe de moléculas reguladoras durante o desenvolvimento cardíaco, e sejam peças chave na determinação das estruturas mais complexas do coração de mamíferos (SCHEUERMANN; BOYER, 2013).

O transcriptoma de células pluripotentes murinas (WAMSTAD *et al.*, 2012) e humanas (KURIAN *et al.*, 2015; LI; LIN; YANG, 2015) durante a diferenciação cardiovascular revelou um grande número de lncRNAs diferencialmente expressos entre os diferentes estágios da diferenciação. Estes estudos mostraram que os lncRNAs seguem padrões específicos de expressão temporal, fornecendo uma lista preliminar de lncRNAs com potencial participação na regulação dos processos de desenvolvimento cardíaco (KURIAN *et al.*, 2015; LI; LIN; YANG, 2015; WAMSTAD *et al.*, 2012; revisado por RIZKI; BOYER, 2015).

Dois lncRNAs murinos – *Braveheart* e *Fendrr* - já foram investigados quanto ao seu mecanismo de ação e se mostraram essenciais no desenvolvimento cardíaco, dando suporte ao importante papel dos lncRNAs nesse processo. *Braveheart* e *Fendrr* interagem com o complexo de silenciamento epigenético Polycomb 2, afetando a expressão de genes essenciais durante o desenvolvimento cardíaco inicial (GROTE *et al.*, 2013; KLATTENHOFF *et al.*, 2013; revisado por RIZKI; BOYER, 2015 e SCHMITZ; GROTE; HERRMANN, 2016). *Braveheart* não é conservado em primatas, entretanto, um lncRNA transcrito na fita *antisense* da sequência ortóloga de *Braveheart* - CARMEN - se mostrou essencial na determinação da identidade cardíaca em humanos (OUNZAIN *et al.*, 2015).

CARMEN, (*CAR*)*diac (M)esoderm (E)nhancer-associated (N)oncoding RNA*, foi identificado durante a diferenciação de cardiomiócitos e sua expressão se correlaciona com a atividade de *enhancers*. CARMEN pode agir tanto em *cis*, promovendo a atividade transcricional de genes, quanto em *trans*, uma vez que pode se ligar ao PRC. Esse lncRNA deriva de um *enhancer* cardíaco ligado pelos fatores de transcrição SRF/NKX2-5, essenciais na especificação cardíaca (OUNZAIN *et al.*, 2015).

Outro lncRNA humano envolvido na diferenciação cardíaca é o HBL1, *Heart Brake LncRNA 1*, que é conservado entre os primatas não humanos, mas não entre outros vertebrados. HBL1 é altamente expresso em células pluripotentes indiferenciadas e sua expressão diminui durante a diferenciação cardiovascular. Ensaios de ganho e perda de função demonstraram que HBL1 é um regulador negativo da diferenciação cardíaca, uma vez que forma um complexo silenciador com a proteína AGO2 e interage com o miRNA MIR1, reprimindo a função de MIR1 no início do desenvolvimento cardíaco (LIU *et al.*, 2017a).

2 JUSTIFICATIVA

As doenças cardiovasculares são a principal causa de mortalidade no mundo atualmente (LUI *et al.*, 2014; SIEGEL *et al.*, 2014). Segundo a Organização Mundial da Saúde (OMS, 2016), cerca de 17 milhões de pessoas morrem em decorrência dessas patologias por ano. No Brasil, as doenças cardiovasculares são responsáveis por cerca de 20% de todas as mortes em indivíduos acima de 30 anos (MANSUR & FAVARATO, 2012). Dentre elas, o infarto do miocárdio, por exemplo, pode resultar na perda de um bilhão de cardiomiócitos em um único evento, causando um declínio da função cardíaca (DEUTSCH *et al.*, 2013; ZAMILPA *et al.*, 2014). Além disso, a má-formação do coração ocorre em 1% a 2% da população, resultando em uma série de doenças cardíacas congênitas (BRUNEAU, 2013).

Em condições de homeostase, os cardiomiócitos proliferam em uma taxa extremamente baixa, estimando-se ser menor que 1% na vida adulta (BERGMANN *et al.*, 2015). Essa baixa proliferação é incapaz de compensar a perda de células após injúria, por exemplo (SENYO; LEE; KÜHN, 2014). Nesse contexto, a introdução de cardiomiócitos derivados de células-tronco oferece uma alternativa de terapia, apesar dos muitos desafios que ainda precisam ser superados, como o direcionamento, a entrega, a integração, a rejeição e a maturação das células (MURRY & KELLER, 2008).

CTEs são células pluripotentes e importantes ferramentas para o estudo da diferenciação celular cardíaca (GAN *et al.*, 2014). Entretanto, muitas questões ainda perduram sobre os mecanismos envolvidos no processo de diferenciação e precisam ser esclarecidas. O avanço no conhecimento sobre a biologia básica do desenvolvimento é de essencial importância e traz essa possibilidade, permitindo a otimização do uso terapêutico de células-tronco (THOMSON, 1998), e a utilização de outras estratégias, como a reprogramação celular direta e a reativação de mecanismos regenerativos endógenos (DEUTSCH *et al.*, 2013; LUI *et al.*, 2014; SENYO *et al.*, 2014).

A resposta aos estímulos biológicos de diferenciação resulta na ativação transcricional de genes específicos que são posteriormente controlados por mecanismos de regulação pós-transcricionais. Portanto, há um crescente aumento no interesse do entendimento dos padrões de expressão e regulação gênica, bem como do gatilho de iniciação, manutenção e coordenação com os processos de

diferenciação (GAN *et al.*, 2014). Neste contexto, o sequenciamento de RNAs em larga escala (RNA-seq) possibilita o amplo levantamento dos perfis de expressão gênica e representa uma importante ferramenta de análise dos mecanismos moleculares envolvidos na diferenciação específica a cardiomiócitos funcionais.

Perfis transcricionais têm colaborado com a investigação das mudanças na expressão gênica durante a diferenciação cardiomiogênica (BEQQALI *et al.*, 2006; DEN HARTOGH *et al.*, 2016; LI; LIN; YANG, 2015; LIU *et al.*, 2017b; TOMPKINS *et al.*, 2016; XU *et al.*, 2009), entretanto, já foi demonstrado que os níveis de mRNA não refletem diretamente os níveis proteicos da célula devido, principalmente, à extensiva regulação traducional a que estes mRNAs estão sujeitos (SCHWANHÄUSSER *et al.*, 2011; TEBALDI *et al.*, 2012). Por isso, a fim de melhor conhecer a tradução efetiva dos mRNAs que contribuem para a composição proteica celular e acompanhar as mudanças ao longo do tempo, é preciso estabelecer a relação entre os padrões de expressão de mRNA total e aqueles que efetivamente sintetizarão proteínas (SONENBERG & HINNEBUSCH, 2007; TIAN *et al.*, 2004).

Nesta perspectiva, uma técnica conhecida como *polysome profiling* foi desenvolvida e consiste no isolamento do mRNA que se encontra associado aos ribossomos, através de um gradiente de sacarose (MASEK *et al.*, 2011; SPANGENBERG *et al.*, 2013). O protocolo consiste na imobilização dos ribossomos ativamente em tradução através da utilização do antibiótico cicloheximida, o qual se liga na subunidade 60S e previne a liberação da molécula de tRNA deacetilada do ribossomo, bloqueando a elongação da tradução (OBRIG *et al.*, 1971). Os polissomos são, então, ultracentrifugados através de um gradiente de sacarose, cuja densidade diferencial contribui para a separação dos grandes complexos polissomais e moléculas de RNA livre em frações distintas (FAYE; GRABER; HOLCIK, 2014; PANDA; MARTINDALE; GOROSPE, 2017). Durante a separação das frações é gerado um perfil polissomal, que por sua vez, pode ser usado para diagnosticar diferentes aspectos do estado traducional da célula. Por exemplo, as frações polissomais refletem o nível de atividade de tradução da célula, e uma redução na taxa de iniciação da tradução resulta na redução dessas frações e no concomitante aumento dos ribossomos livres (80S), chamados de monossomos (MASEK; VALASEK; POSPISEK, 2011; SUNNERHAGEN, 2007).

Diante disso, este trabalho teve como objetivo a caracterização da regulação pós-transcricional durante a diferenciação de células-tronco a cardiomiócitos utilizando a técnica de *polysome profiling* associado ao RNA-Seq.

3 OBJETIVOS

3.1 Objetivo Geral

Caracterizar a regulação pós-transcricional durante o processo de diferenciação cardiomiogênica através da análise da expressão diferencial de RNAs livres e associados aos polissomos, assim como identificar genes envolvidos nesse processo.

3.2 Objetivos específicos

- 1) Indução da diferenciação cardiomiogênica de células-tronco embrionárias humanas;
- 2) Isolamento e identificação dos RNAs livres e polissomais das células em diferenciação cardiomiogênica nos dias inicial (D0), 1º (D1), 4º (D4), 9º (D9) e 15º (D15) através de *polysome profiling* seguido de sequenciamento em larga escala (RNA-seq);
- 3) Análise dos genes diferencialmente expressos entre os diferentes tempos de diferenciação e entre as frações livres e polissomais;
- 4) Identificação de vias de sinalização reguladas durante a diferenciação e, especificamente, reguladas pós-transcricionalmente;
- 5) Análise de RNAs não-codificantes longos (lncRNAs) com expressão e associação diferencial aos polissomos durante a diferenciação;
- 6) Avaliação funcional de genes selecionados segundo seu perfil de expressão e potencial participação na diferenciação cardiomiogênica.

CAPÍTULO I

Artigo

Polysome profiling followed by RNA-seq of cardiac differentiation stages in hESCs

Situação – Publicado na revista Scientific Data (dezembro de 2018)

Pereira, I.T.; Spangenberg, L.; Robert, A.W.; Amorín, R.; Stimamiglio, M.A.; Naya, H.; Dallagiovanna, B. Polysome profiling followed by RNA-seq of cardiac differentiation stages in hESCs. *Sci. data* **2018**, 5, 1–11, doi:10.1038/sdata.2018.287.

Prefácio

Este capítulo apresenta o artigo publicado na revista Scientific Data, que descreve o protocolo de diferenciação cardíaca de CTE, a técnica de *polysome profiling* e a geração dos dados de RNA-seq que basearam esta Tese. É um artigo técnico, de descrição de metodologias, detalhamento das amostras e controle de qualidade dos dados. Os dados descritos neste artigo foram usados para as demais análises descritas ao longo da Tese.

SCIENTIFIC DATA

OPEN

Data Descriptor: Polysome profiling followed by RNA-seq of cardiac differentiation stages in hESCs

Isabela Tiemy Pereira¹, Lucia Spangenberg², Anny Waloski Robert¹, Rocío Amorín², Marco Augusto Stimamiglio¹, Hugo Naya² & Bruno Dallagiovanna¹

Received: 2 July 2018

Accepted: 23 October 2018

Published: 4 December 2018

The regulation of gene expression acts at numerous complementary levels to control and refine protein abundance. The analysis of mRNAs associated with polysomes, called polysome profiling, has been used to investigate the post-transcriptional mechanisms that are involved in different biological processes. Pluripotent stem cells are able to differentiate into a variety of cell lineages, and the cell commitment progression is carefully orchestrated. Genome-wide expression profiling has provided the possibility to investigate transcriptional changes during cardiomyogenesis; however, a more accurate study regarding post-transcriptional regulation is required. In the present work, we isolated and high-throughput sequenced ribosome-free and polysome-bound RNAs from *NKX2-5^{eGFP/w}* HES3 undifferentiated pluripotent stem cells at the subsequent differentiation stages of cardiomyogenesis: embryoid body aggregation, mesoderm, cardiac progenitor and cardiomyocyte. The expression of developmental markers was followed by flow cytometry, and quality analyses were performed as technical controls to ensure high quality data. Our dataset provides valuable information about hESC cardiac differentiation and can be used to investigate genes potentially controlled by post-transcriptional mechanisms.

Design Type(s)	organism development design • transcription profiling design
Measurement Type(s)	cardiocyte differentiation • transcriptional profiling assay
Technology Type(s)	flow cytometry assay • RNA sequencing
Factor Type(s)	biological replicate • experimental condition • developmental stage
Sample Characteristic(s)	embryonic stem cell

¹Stem Cells Basic Biology Laboratory, Instituto Carlos Chagas - FIOCRUZ-PR, Rua Professor Algacyr Munhoz Mader, 3775, 81.350-010 Curitiba, Brazil. ²Bioinformatics Unit, Institut Pasteur de Montevideo, Mataojo 2020, CP 11400 Montevideo, Uruguay. Correspondence and requests for materials should be addressed to B.D. (email: bruno.dallagiovanna@fiocruz.br)

Background & Summary

Gene expression is controlled by a series of mechanisms, finally leading to protein formation. Continued findings regarding new regulatory mechanisms of gene expression are due to an increased understanding of RNA^{1,2} and more integrative analysis tools^{3,4}. The post-transcriptional level of regulation includes transcript synthesis, 5' capping, splicing, polyadenylation, nuclear export, translation and decay⁵. Translation variants have already been shown as crucial determinants of mammalian gene expression^{6,7}, but genome-wide expression profiling is not able to detect the fine adjustment provided by post-transcriptional mechanisms. To overcome this issue, the polysome profiling technique has been used to isolate and further independently analyze ribosome-free and polysome-bound RNAs. RNAs associated with many ribosomes, called polysomes, form large complexes of high molecular weight⁸ and can be easily segregated from ribosome-free RNAs through a sucrose gradient^{9–12}.

Protein synthesis control pathways and post-transcriptional mechanisms involved in cell fate commitment are still being established^{13–15}. Cardiac tissue formation occurs through precise activation of specific sequential genetic programs to drive cells to differentiation. During embryonic development, cardiomyocytes are derived from the cardiogenic mesoderm through modulation of many pathways and signaling molecules, including bone morphogenetic proteins (BMP), fibroblast growth factors (FGF), NODAL, and canonical and non-canonical Wnt (reviewed by¹⁶). Moreover, the functional interconnection between transcription factors, their gene targets and signaling pathways delineates cardiomyogenesis and is evolutionarily conserved¹⁷. The use of isolated human embryonic stem cells (hESC) and induced pluripotent stem cells (iPSC) to derive specific cell lineages *in vitro* raised the possibility of artificially reproducing and studying this differentiation process. Modifications of signaling pathways were used to differentiate hESC to cardiomyocytes^{18–21}, which are potential resources for cell therapy and can be used as tools for developmental studies, investigation of endogenous regenerative promotion and cardiac toxicity assays^{18,22,23}. However, there is still a lack of detailed information about the complex gene regulatory network that controls cardiac commitment. Unveiling key regulatory elements and molecular signatures of the intermediate differentiation stages can further our current understanding of human cardiac development and produce, select and identify suitable cells for a range of different applications²⁴.

Here, we describe the polysome profiling during the developmental steps of cardiomyogenic commitment. Ribosome-free and polysome-bound mRNAs were isolated and sequenced on D0, D1, D4, D9 and D15, which represents pluripotency, embryoid body (EB) aggregation, cardiac mesoderm, cardiac progenitor and cardiomyocyte stages, respectively (Fig. 1b). Three independent experiments were prepared using 2 to 6 million cells on each time-point mentioned, and technical controls for each analyzed sample and experimental stage were done to ensure high quality data. An overview of the study design is illustrated in Fig. 1a. Our dataset provides valuable information regarding hESC cardiac differentiation and can be used to investigate genes potentially controlled by post-transcriptional mechanisms. Moreover, these data are a powerful tool to explore new elements involved in cardiac cell fate commitment and contributes to the development of novel therapy and research approaches.

Methods

Human ESC culture

NKX2-5^{eGFP/w} HES3 cell lineage was donated by Monash University (Victoria, Australia)²⁵. hESCs were cultured on irradiated mouse embryonic fibroblasts (iMEFs) in specific medium composed of Dulbecco's modified Eagle's medium (DMEM)/F12 supplemented with 20% KnockOut™ serum replacement (KSR, Gibco™), 100 U/mL penicillin, 100 µg/mL streptomycin, 2 mM L-Glutamine, 1% non-essential amino acid, 55 µM β-mercaptoethanol and 10 ng/mL of human basic fibroblast growth factor (FGF2) (Sigma). They were maintained in a humidified incubator with 5% CO₂ at 37 °C, with daily medium change and passage every 3–4 days by enzymatic dissociation using 0.05% trypsin/EDTA.

Cardiomyogenic differentiation of hESCs

A cardiac differentiation protocol was adapted from a previously described source in¹⁸ and consists of 3 steps: embryoid body (EB) formation, mesoderm induction and cardiac progenitor induction. Initially, 7 × 10⁵ cells/well were plated on Growth Factor Reduced Matrigel Matrix (Corning) 6-well coated dishes and maintained for 72 h in a humidified incubator with 5% CO₂ at 37 °C. At day 0 (D0) of protocol, hESCs were incubated with collagenase I (1 mg/mL) for 20 min at 37 °C, followed by trypsin-EDTA (0.05%) for approximately 1 min. Immediately after, trypsin was carefully removed, and a medium containing 50% of fetal bovine serum (FBS) and DNase I (20 U/mL, Invitrogen) was added to the plate. Cells were detached with a cell scraper to avoid single-cell detachment and centrifuged at 230 × g for 5 min. After removal of the supernatant, a basal medium composed of StemPro34 (StemPro™-34 SFM, Gibco™), 100 U/mL penicillin, 100 µg/mL streptomycin, 2 mM L-Glutamine, 150 µg/mL transferrin, 50 µg/mL ascorbic acid and 0.45 mM monothioglycerol (MTG) was supplemented with 1 ng/mL BMP4 (R&D systems, cat. 314-BP) and added gently. The cell pellet was resuspended to form small clusters of 10–20 cells, which were seeded into ultra-low attachment 6-well culture plates (Corning Costar Ultra-Low Attachment plate) and kept in a humid incubator at 37°C, 5% CO₂ and 5% O₂ (hypoxia) for EB aggregation for 24 h. At day 1 (D1), EBs were collected and decanted in a round bottom plastic tube for 30 min. After this period, the supernatant was gently removed, and EBs were resuspended in basal medium supplemented with 10 ng/mL BMP4, 6 ng/mL Activin A (R&D systems, cat. 338-AC) and

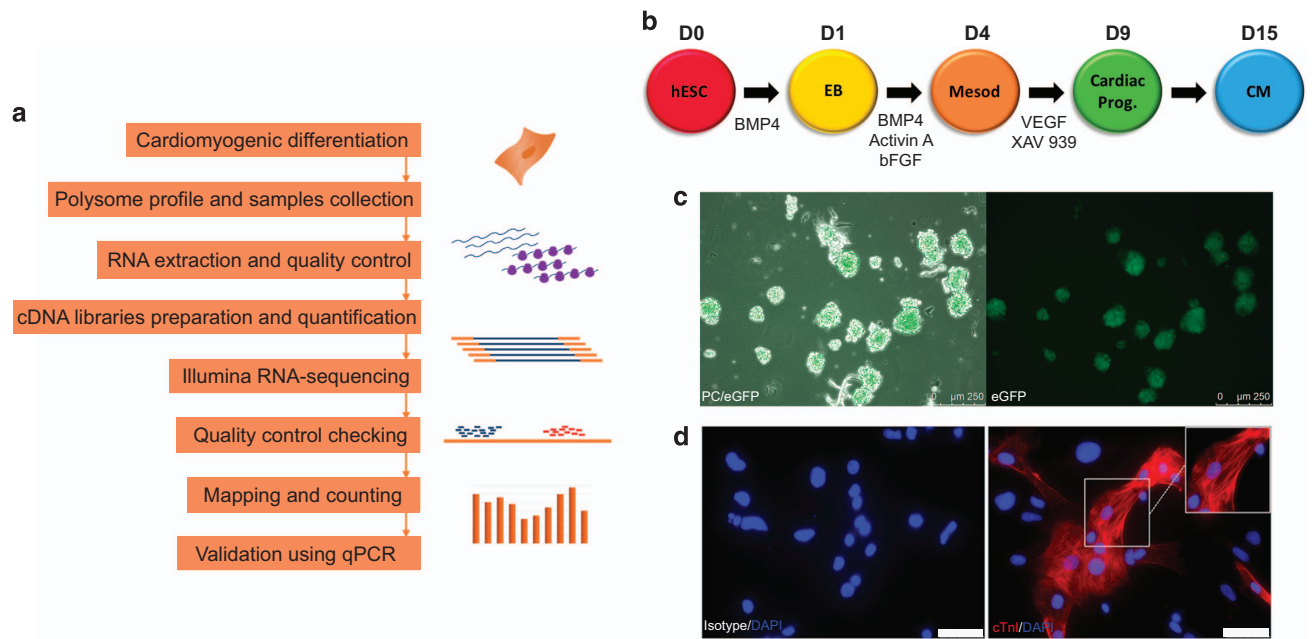


Figure 1. Cardiomyogenic differentiation of hESCs. (a) Schematic representation of the steps followed for RNA-seq data generation. (b) Schematic representation of the cardiomyogenic differentiation protocol, indicating days of differentiation and timing of specific induction. (c) Representative images of EBs during differentiation showing NKX2-5/eGFP expression on D15. Phase contrast (PC) and eGFP fluorescence (left image), eGFP fluorescence (right image). 250 μm scale. (d) Representative images of differentiated cardiomyocytes stained for cTnI on D20. Isotype control (left image), cTnI staining (right image). White rectangle as 50 μm scale.

5 ng/mL FGF2 (R&D systems, cat. 233-FB) to induce mesoderm specification. After 72 h, on day 4 (D4), the medium was replaced with basal medium supplemented with XAV939 (10 $\mu\text{M}/\text{mL}$) (Tocris, cat. 3748) and VEGF (10 ng/mL) (R&D systems, cat. 293-VE) to induce cells into cardiac progenitors. On days 8 and 11, the medium was replaced with basal medium supplemented with VEGF (10 ng/mL) and BMP4 (1 ng/mL). The cells were kept in a humid incubator at 37 $^{\circ}\text{C}$, 5% CO_2 and 5% O_2 during all the procedure. Three independent differentiation assays were used as experimental replicates. As a control of cardiomyogenic differentiation, hESC were submitted to the same processing without adding any induction factor (non induced differentiation).

Flow cytometry

EBs were dissociated on D4 using trypsin-EDTA (0.05%) and incubated for 20 min with the surface marker PE-conjugated anti-CD56 (BD cat. 347747, 1:25 in 0.5% PBS/BSA) and 1 $\mu\text{g}/\mu\text{L}$ DAPI. On D9, cells were disaggregated with trypsin-EDTA (0.05%) for 5 min and resuspended in PBS to evaluate eGFP expression. On D15, EBs were disaggregated using 1 mg/mL collagenase I for 16 h and trypsin-EDTA (0.05%) for 5 min, fixed with 4% paraformaldehyde for 20 min and permeabilized with 0.5% Triton X-100 for 30 min. Cells were incubated with anti-troponin T (1:100 in 0.5% PBS/BSA, cardiac isoform Ab-1, Thermo Scientific[™], cat. #MS-295-P0) for 30 min followed by Pacific Blue-conjugated anti-mouse (1:1000) for 30 more min. Analyses were carried out using a FACSCanto II flow cytometer (Bd Biosciences) and FlowJo software.

Immunofluorescence and fluorescent microscopy

EBs on D15 were visualized under a fluorescent microscope and cardiomyogenic committed cells showed NKX2-5/eGFP expression. For immunofluorescence staining, EBs on D15 were disaggregated using 1 mg/mL collagenase I for 16 h and trypsin-EDTA (0.05%) for 5 min and plated on Matrigel coated wells. After 2–5 days, cardiomyocytes were fixed with 4% paraformaldehyde, permeabilized with 0.5% Triton X-100 and blocked with 1% PBS-BSA. Cells were incubated overnight with anti-troponin I (Santa Cruz Biotechnology, cat.: sc-15368, 1:100 in 0.5% PBS/BSA) followed by Alexa 546-conjugated anti-rabbit IgG (1:800, Invitrogen) and 1 $\mu\text{g}/\mu\text{L}$ DAPI. EBs or fixed/stained cardiomyocytes were visualized using a Leica DMI6000B optical microscope, and images were acquired by LAS AF software.

	Sample	Fraction	Replicate	Deposit
D0	2.23E + 06 cells on day 0	Ribosome-free	Replicate 1	SAMN09405494
		Polysome-bound	Replicate 1	SAMN09405495
	4.02E + 06 cells on day 0	Ribosome-free	Replicate 2	SAMN09405504
		Polysome-bound	Replicate 2	SAMN09405505
	3.18E + 06 cells on day 0	Ribosome-free	Replicate 3	SAMN09405514
		Polysome-bound	Replicate 3	SAMN09405515
D1	4.65E + 05 cells on day 1	Ribosome-free	Replicate 1	SAMN09405496
		Polysome-bound	Replicate 1	SAMN09405497
	2.44E + 06 cells on day 1	Ribosome-free	Replicate 2	SAMN09405506
		Polysome-bound	Replicate 2	SAMN09405507
	2.38E + 06 cells on day 1	Ribosome-free	Replicate 3	SAMN09405516
		Polysome-bound	Replicate 3	SAMN09405517
D4	3.30E + 06 cells on day 4 (76.9% CD56)	Ribosome-free	Replicate 1	SAMN09405498
		Polysome-bound	Replicate 1	SAMN09405499
	4.38E + 06 cells on day 4 (89.2% CD56)	Ribosome-free	Replicate 2	SAMN09405508
		Polysome-bound	Replicate 2	SAMN09405509
	2.16E + 06 cells on day 4 (74.4% CD56)	Ribosome-free	Replicate 3	SAMN09405518
		Polysome-bound	Replicate 3	SAMN09405519
D9	5.04E + 06 cells on day 9 (59.5% eGFP)	Ribosome-free	Replicate 1	SAMN09405500
		Polysome-bound	Replicate 1	SAMN09405501
	6.38E + 06 cells on day 9 (50.7% eGFP)	Ribosome-free	Replicate 2	SAMN09405510
		Polysome-bound	Replicate 2	SAMN09405511
	2.96E + 06 cells on day 9 (64.9% eGFP)	Ribosome-free	Replicate 3	SAMN09405520
		Polysome-bound	Replicate 3	SAMN09405521
D15	4.58E + 06 cells on day 15 (61.5% cTnT/ 77.8% eGFP)	Ribosome-free	Replicate 1	SAMN09405502
		Polysome-bound	Replicate 1	SAMN09405503
	2.56E + 06 cells on day 15 (54% cTnT/ 71.2% eGFP)	Ribosome-free	Replicate 2	SAMN09405512
		Polysome-bound	Replicate 2	SAMN09405513
	3.45E + 06 cells on day 15 (53.2% cTnT/ 78% eGFP)	Ribosome-free	Replicate 3	SAMN09405522
		Polysome-bound	Replicate 3	SAMN09405523

Table 1. Description of the samples used to generate RNA-sequencing data of distinct days of differentiation. Day of differentiation, number of cells, gradient fraction and replicates information. Total of 30 samples were prepared (Data Citation 2).

Polysome profiling

Differentiating cells (2–6 million) at D0, D1, D4, D9 and D15 were treated with 0.1 mg/mL cycloheximide (Sigma-Aldrich) for 10 min at 37 °C, disaggregated with trypsin-EDTA (0.05%) for 10 min and washed twice with 0.1 mg/mL cycloheximide in 1X PBS. Polysome lysis buffer composed of 15 mM Tris HCl pH 7.4, 15 mM MgCl₂, 300 mM NaCl, 100 µg/mL cycloheximide, 1% Triton X-100, 40 U/µL RNase Out and 24 U/mL DNase was used to resuspend cells, followed by 10 min incubation on ice and 10 min centrifugation at 12000 × g at 4 °C. Sucrose gradients were prepared with BioComp model 108 Gradient Master using 10% and 50% sucrose solutions (sucrose diluted in polysomal buffer containing 15 mM Tris HCl pH 7.4, 15 mM MgCl₂ and 300 mM NaCl and prepared in RNase-free conditions). Clear supernatants from lysed cells were loaded into the 10 to 50% sucrose gradients and centrifuged at 150000 × g (SW40 rotor, HIMAC CP80WX HITACHI) for 160 min at 4 °C. Sucrose gradient fractions were separated using ISCO gradient fractionation system (ISCO Model 160 Gradient Former Foxy Jr. Fraction Collector), and the absorbance was monitored at 254 nm to record the polysome profile.

RNA isolation and quality control

Ribosome-free (fractions 1–3) and polysomal (fractions 10–22) fractions were pooled, and RNA was isolated using the Direct-zol RNA MiniPrep (Zymo Research), following the manufacturer's instructions. Quality and quantity of RNA were determined using RNA 6000 Pico (for ribosome-free) and Nano (for polysome-bound) kits and an Agilent 2100 Bioanalyzer (Agilent Technologies) and compared to reference samples. Ribosome-free RNA samples ranged from 0.2 to 3.4 µg and polysome-bound from 5 to 16 µg.

Gene	Sequence (5'-3')	Access number	Amplicon (pb)	Reference
POU5F1	F: ATGCATTCAAAGTGGAGTGCCTGC	NM_001173531	192 pb	(YE <i>et al.</i> , 2013)
	R: AACTTCACCTTCCCTCCAACCAAGT			
NANOG	F: ACCGAACTGTGTTCTCTCCACC	NM_024865	200 pb	(ZAEHRES <i>et al.</i> , 2005)
	R: CCATTGCTATTCTTCGGCCAGTTG			
T	F: AAAGAGATGATGGAGGAACCCGGA	NM_003181	108 pb	(YE <i>et al.</i> , 2013)
	R: AGGATGAGGATTGCAGGTGGACA			
EOMES	F: CAAATTCACCGCCACCAAACTGA	NM_001278182.1	108 pb	(OVCHINNIKOV <i>et al.</i> , 2014)
	R: TTGTAGTGGCAGTGGGATTGAGT			
GATA4	F: ACCTGGACTTGGAGGATAGCAA	NM_002052	169 pb	(YE <i>et al.</i> , 2013)
	R: TCCCATCAGCGTAAAGGCATCT			
NKX2.5	F: TTAAGTCACCGTCTGTCTCCCTCA	NM_001166175	124 pb	(YE <i>et al.</i> , 2013)
	R: ACCGACAGTCTCACTCAGCATT			
TNNT2	F: TGCAGGAGAAGTCAAGCAGCAGA	NM_000364	155 pb	(YE <i>et al.</i> , 2013)
	R: AGCGAGGAGCAGATCTTTGGTGAA			
TNNI3	F: GGGGGCCCGGGCTAAGGAGTC	NM_000363.4	183 pb	(SCHITTINI <i>et al.</i> , 2010)
	R: AGGGCAGGGCAGTAGGCAGGAAG			
GAPDH	F: GCGGATGCTGGCGCTGAGTAC	NM_002046.3	149 pb	PrimerBlast
	R: TGGTTCACACCCATGACGA			

Table 2. Primer sequences used for qPCR analysis of cardiac developmental genes.

High-Throughput sequencing and data analysis

A total of 30 samples were prepared for sequencing (Table 1). The cDNA libraries were prepared with 200–500 ng of ribosome-free or 2 µg of polysome-bound RNA using the TruSeq Stranded mRNA Sample Preparation kit (Illumina, Inc.). Quality and quantity of cDNA libraries were determined using the DNA 1000 kit, Agilent 2100 Bioanalyzer and KAPA Library Quantification qPCR (KAPA Biosystems). RNA-seq was carried out in an Illumina HiSeq 2500 platform, and raw data quality control was generated using FastQC Reports (<https://www.bioinformatics.babraham.ac.uk/projects/fastqc/>).

Mapping and counting of sequencing data were performed with Rsubread package²⁶ against the new version of the human genome GRCh38. Mapping was done with default parameters and set for unique mapping of the reads. Counting was performed using the annotation of Ensembl (GRCh38). For comparisons of gene expression within and between samples, RPKM values (reads per kilobase per million mapped reads, an expression measure) were determined. For quality check purposes, we performed a PCA analysis, a dimension reduction method of the matrix of counts, to explore associations between variables. Samples of the same condition should cluster together in order to ensure consistency and replicability of results.

Code availability

R code for data analysis is available upon request. The R version used for this study was 3.3.2. R-packages Rsubread and edgeR were used with versions 1.24.2 and 3.16.5, respectively.

cDNA preparation and quantitative PCR (qPCR)

Total RNA was extracted using TRIzol[®] Reagent (Invitrogen), and cDNA synthesis was performed with ImProm-II[™] Reverse Transcription System (Promega), following manufacturer's instructions. Quantitative analysis of transcripts was performed using SYBR[®] Green (Applied Biosystems) and LightCycler[®] 96 (Roche) equipment. The primers that were used are shown in Table 2, and for each reaction, 5 pmol of primer and 25 ng of cDNA were used. All samples were evaluated in triplicate.

Data Records

Flow cytometry data generated during this study were submitted to the FlowRepository (Data Citation 1). FCS files related to each replicate and cardiac differentiation time evaluated (day 3, day 4, day 9 and day 15) are available.

RNA-seq data related to this study were submitted to the NCBI repository SRA (Data Citation 2). Raw RNA-seq data (paired-end fastq files) as generated by Illumina HiSeq 2500 (Data Citation 2). This site serves as a landing page for the study: description of the project, metadata and raw sequencing files can be found there. Individual accession numbers for each biological sample are also provided in Table 1. Counts data and RPKM can be found in file table_genes_counts.xlsx (Data Citation 3). One tab corresponds to the read counts of each sample, and the other, to RPKM values. Each column of each file

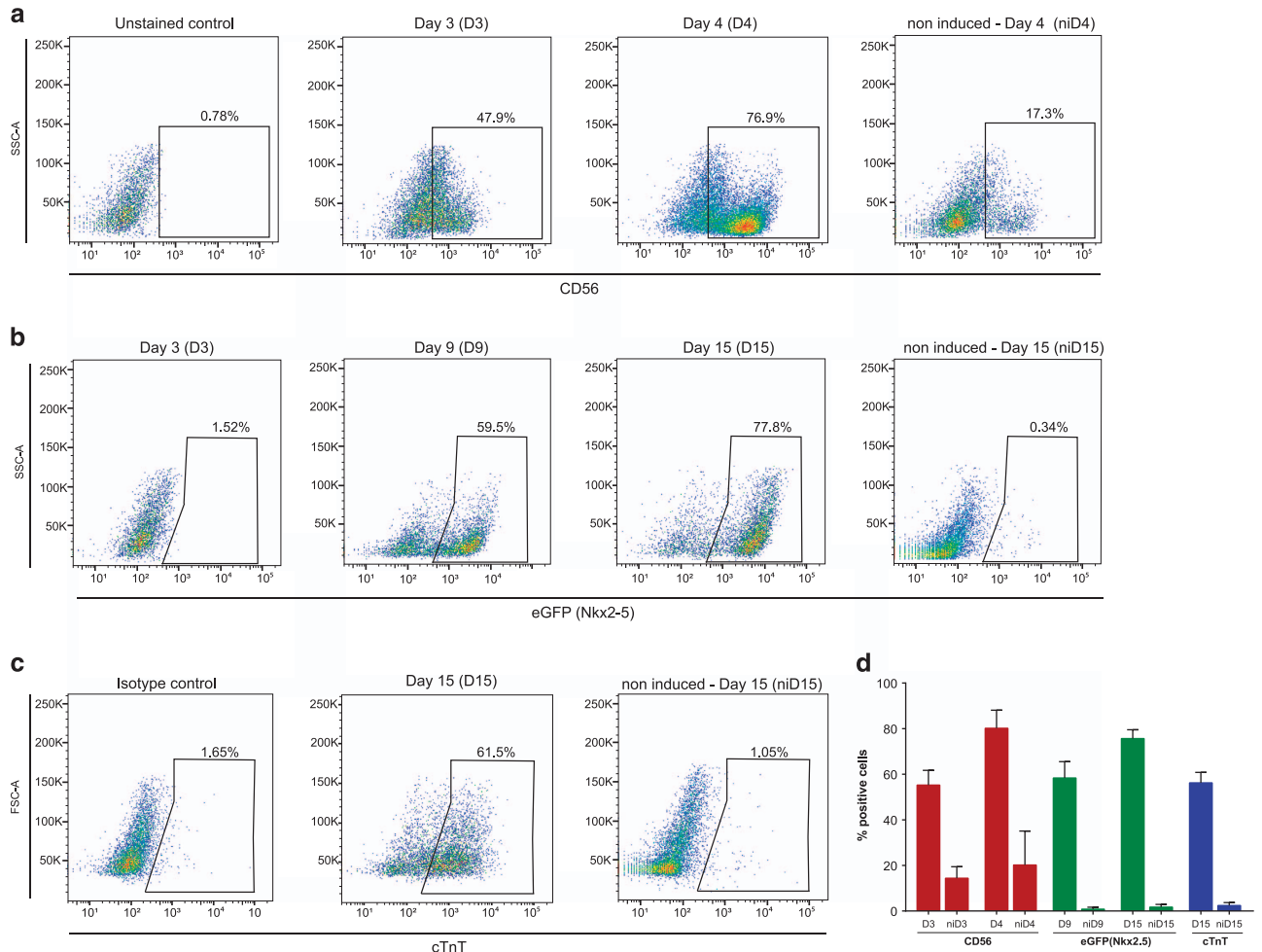


Figure 2. Expression of markers followed during cardiomyogenic differentiation. Flow cytometry analysis of (a) D3 and D4 (CD56), (b) D3, D9 and D15 (eGFP) and (c) D15 (cTnT) differentiating cells. Representative dot plots ($n = 3$). Non-induced (ni) cells were used as a control for differentiation. (d) Quantification of percentage of positive cells for the indicated markers ($n = 3$) (Data Citation 1).

is labeled with the sample condition, e.g., 1D4P corresponds to biological replicate 1, at day 4 and polysomal RNA fraction.

Beating cardiomyocytes were recorded on video (Online video I, Data Citation 3).

Technical Validation

Cardiomyogenic differentiation

The $NKX2-5^{eGFP/w}$ HES3 cell lineage is a reporter human embryonic stem cell (hESC) line that can be used to derive cardiomyocytes and follow the differentiation through eGFP expression²⁵. Here, we used a developmentally staged protocol^{18,27} to induce a cardiac mesoderm population on days 3 and 4 and a $NKX2-5^+/cTNT^+$ population by day 15 (Fig. 1b). Beating clusters were observed after 10 days of differentiation (Online video I, Data Citation 3), and cardiomyogenic cells were seen by $NKX2-5/eGFP$ expression (Fig. 1c). Immunostaining using cTNI showed the striations characteristic of sarcomere structures on day 20 of differentiation, as representative control of differentiation protocol (Fig. 1d). To follow the differentiation progress, we established two checkpoints during the cardiomyogenesis protocol: (1) presence of CD56+ cells on days 3–4, which corresponds to mesoderm specification²⁸, and (2) $NKX2-5/eGFP^+$ cells on day 9, meaning cardiac progenitor commitment. Moreover, cTnT expression was also determined on day 15 and considered proportional to the efficiency of differentiation. Those markers were followed by flow cytometry in all replicate experiments ($n = 3$) (Fig. 2 and Data Citation 1). Samples used for data acquisition yielded $80.17 \pm 7.9\%$ of CD56+ on D4, $58.37 \pm 7.1\%$ of eGFP+ on D9 and on D15 $75.67 \pm 3.8\%$ of eGFP+ and $56.23 \pm 4.5\%$ cTnT+ cells (Table 1 and Fig. 2).

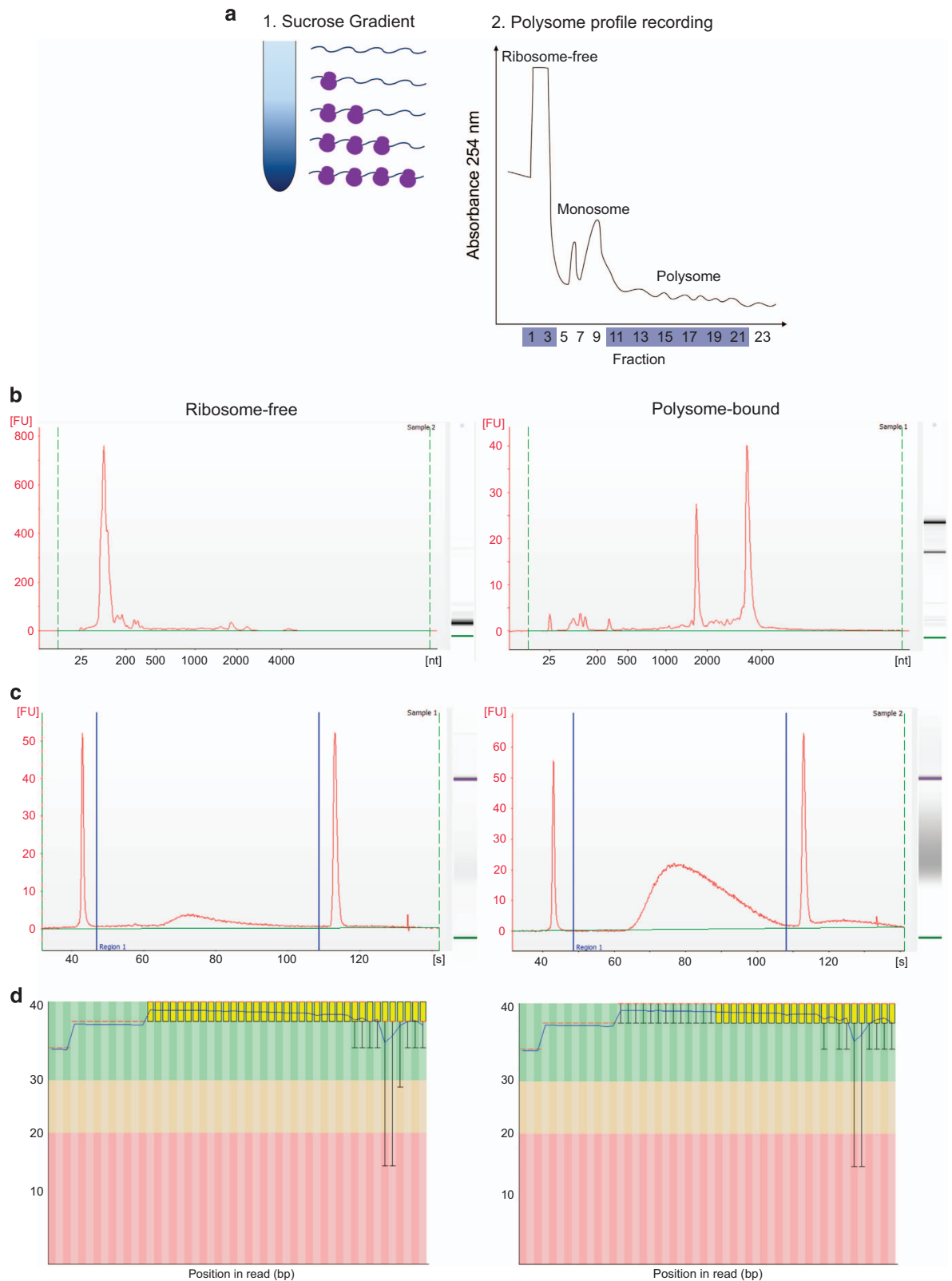


Figure 3. Polysome profiling followed by RNA-seq during cardiomyogenic differentiation. (a) Schematic representation of the sucrose gradient used to segregate ribosome-free and ribosome-bound RNAs and representative polysome profile (hESCs - D0 replicate 2) recorded at 254 nm. Ribosome-free and polysome fractions are indicated. (b-d) Representative quality analysis of ribosome-free and polysome-bound samples. (a) RNA quality analysis using Agilent 2100 Bioanalyzer. (b) cDNA library quality analysis using Agilent 2100 Bioanalyzer. (c) RNA-sequencing reads quality analysis using FastQC. All representative images correspond to D0 hESC sample, replicate 2.

	Fraction	Replicate	Processed reads	Mapped reads	% mapped reads	Genes detected
D0	Ribosome-free	Replicate 1	3.40E + 07	2.45E + 07	72.06	17,346
	Polysome-bound	Replicate 1	3.55E + 07	3.02E + 07	84.99	17,690
	Ribosome-free	Replicate 2	3.33E + 07	1.64E + 07	49.40	16,333
	Polysome-bound	Replicate 2	3.47E + 07	2.94E + 07	84.65	16,820
	Ribosome-free	Replicate 3	3.87E + 07	2.78E + 07	71.72	18,019
	Polysome-bound	Replicate 3	3.48E + 07	2.95E + 07	84.93	17,118
D1	Ribosome-free	Replicate 1	3.71E + 07	2.55E + 07	68.57	17,978
	Polysome-bound	Replicate 1	3.73E + 07	3.17E + 07	84.89	17,921
	Ribosome-free	Replicate 2	3.81E + 07	2.55E + 07	66.93	17,207
	Polysome-bound	Replicate 2	3.31E + 07	2.81E + 07	84.82	17,513
	Ribosome-free	Replicate 3	2.52E + 07	1.72E + 07	68.21	17,447
	Polysome-bound	Replicate 3	3.60E + 07	3.11E + 07	86.28	17,907
D4	Ribosome-free	Replicate 1	3.34E + 07	2.89E + 07	86.35	18,057
	Polysome-bound	Replicate 1	3.85E + 07	3.36E + 07	87.17	18,368
	Ribosome-free	Replicate 2	1.39E + 07	8.07E + 06	57.96	16,871
	Polysome-bound	Replicate 2	3.29E + 07	2.83E + 07	85.77	17,824
	Ribosome-free	Replicate 3	2.69E + 07	1.76E + 07	65.33	18,529
	Polysome-bound	Replicate 3	3.37E + 07	2.92E + 07	86.64	18,155
D9	Ribosome-free	Replicate 1	3.55E + 07	2.62E + 07	73.82	18,182
	Polysome-bound	Replicate 1	2.12E + 07	1.90E + 07	89.32	18,552
	Ribosome-free	Replicate 2	2.67E + 07	1.84E + 07	69.02	16,826
	Polysome-bound	Replicate 2	2.65E + 07	2.36E + 07	88.95	17,978
	Ribosome-free	Replicate 3	2.98E + 07	2.16E + 07	72.38	17,678
	Polysome-bound	Replicate 3	2.90E + 07	2.59E + 07	89.38	17,924
D15	Ribosome-free	Replicate 1	2.97E + 07	2.23E + 07	74.96	17,626
	Polysome-bound	Replicate 1	3.25E + 07	2.90E + 07	89.35	18,093
	Ribosome-free	Replicate 2	5.09E + 07	3.61E + 07	70.93	16,636
	Polysome-bound	Replicate 2	5.00E + 07	4.50E + 07	89.98	17,824
	Ribosome-free	Replicate 3	5.22E + 07	3.73E + 07	71.50	15,948
	Polysome-bound	Replicate 3	2.92E + 07	2.60E + 07	89.12	17,580

Table 3. Summary of RNA-seq data from ribosome-free and polysome-bound fractions of distinct cardiomyogenic differentiation time-points (n = 3).

Polysome profiling

In order to increase the accuracy of the cardiomyogenesis translome study, we chose to use the polysome profiling methodology to access the polysome-bound RNAs in distinct phases of cardiac differentiation. We performed polysome profiling on D0, D1, D4, D9 and D15 of the differentiation protocol, which represent pluripotency, EB aggregation, cardiac mesoderm, cardiac progenitor and cardiomyocyte stages, respectively (Fig. 1b). Differing densities within the sucrose gradient allowed for the isolation of ribosome-free and polysome-bound RNAs after ultracentrifugation (Fig. 3a). Several fractions were separated using the ISCO gradient fractionation system, while the polysome profiles were recorded using a UV detector (Fig. 3a). The polysome profile derived from each sample was used to determine the fractions that corresponded to ribosome-free (fractions 1–3) and polysome-bound RNAs (fraction 10–22), which were pooled and followed by RNA extraction. One representative image of polysome profile (D0 replicate 2) is shown on Fig. 3a.

RNA analysis, cDNA libraries and sequencing quality control

Isolated RNAs from ribosome-free and polysome-bound fractions were analyzed for quality and concentration to determine their suitability for RNA-sequencing using an Agilent 2100 BioAnalyzer. Figure 3b shows representative examples of quality results of ribosome-free and polysome-bound samples. Polysome-bound samples showed two distinct peaks, which represent 18S and 28S ribosomal RNAs. Those peaks were not shown in ribosome-free samples, which was expected, given the absence of

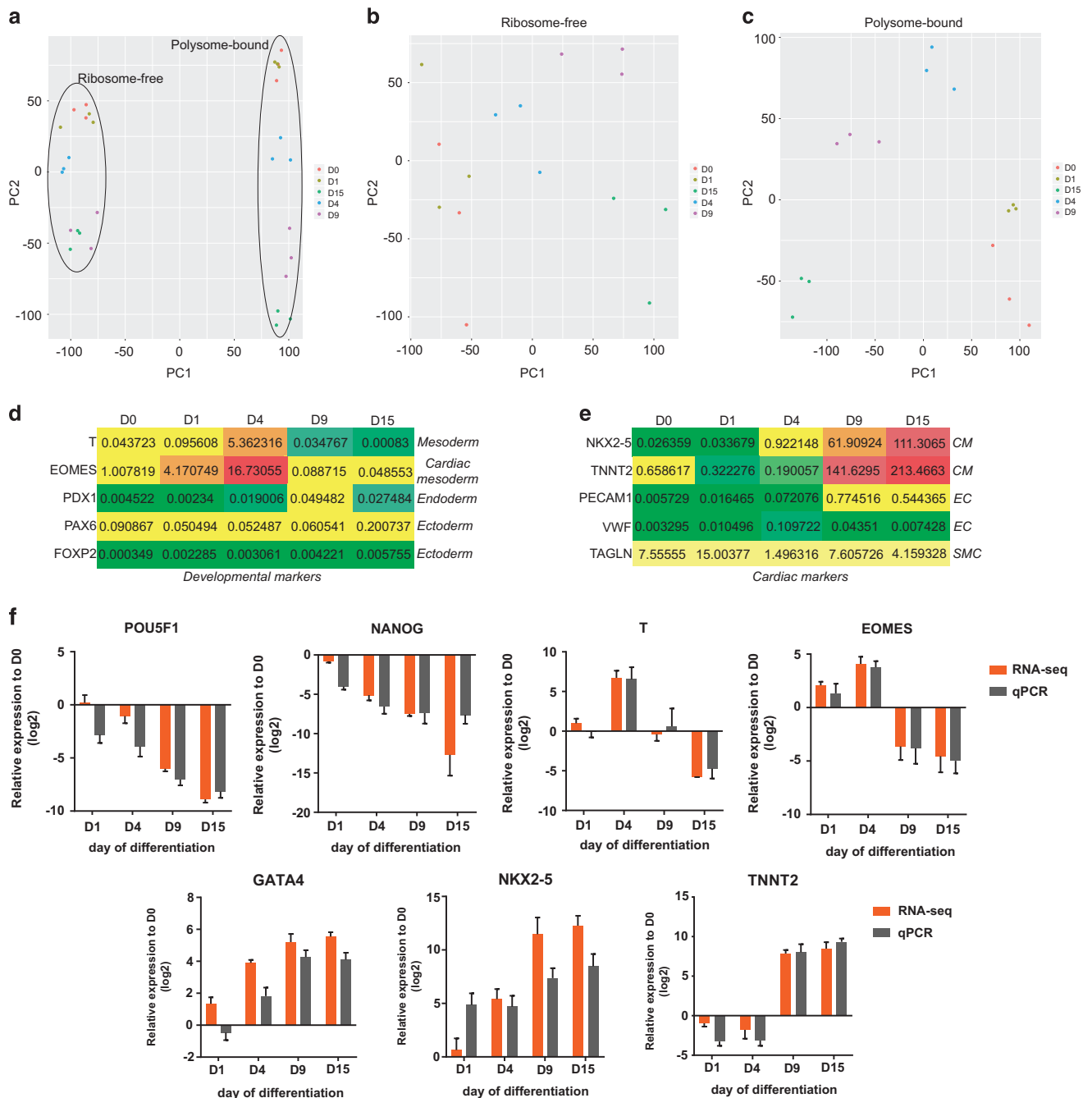


Figure 4. Data quality analysis of RNA-seq and validation. Principal component analysis (PCA) of (a) all sequenced samples (total 30 samples), (b) ribosome-free and (c) polysome-bound samples at D0, D1, D4, D9 and D15 ($n = 3$). RPKM values (polysome-bound) heatmap of (d) developmental markers showing mesoderm, cardiac mesoderm, endoderm and ectoderm genes expression; and (e) cardiac markers showing cardiomyocytes (CM), endothelial cells (EC) and smooth muscle cells (SMC) genes expression. (f) RNA-seq data validation of cardiac developmental marker gene expression using qPCR. Relative expression to hESC (D0) of pluripotency (POU5F1 and NANOG), mesoderm (T and EOMES) and cardiac markers (GATA4, NKX2-5 and TNNT2) on distinct days of cardiac differentiation using RNA-seq and q-PCR data. Values are expressed in \log_2 base.

ribosomes. On the other hand, a smaller peak corresponding to tRNAs was observed. All samples measured as high integrity and were considered of high quality to be used on RNA-sequencing.

The cDNA libraries prepared with the TruSeq Stranded mRNA Sample Preparation kit were analyzed to determine quality and quantity using an Agilent 2100 Bioanalyzer (Fig. 3c). As examples, representative images generated by this analysis are shown for ribosome-free and polysome-bound

samples. Moreover, cDNA libraries were also quantified by a KAPA Library Quantification kit (data not shown), and these values were used to calculate the sequencing input samples. Raw data derived from sequencing were analyzed using FastQC to determine the quality of the reads by comparing read signals to the probability of accurate base-reading. All samples showed suitable scores, and a ribosome-free and polysome-bound representative analysis is shown (Fig. 3d).

Biological RNA-seq data

A brief sample description is illustrated in Table 3. Three experimental replicates were done for each differentiation time-point and derived ribosome-free and polysome-bound RNA samples. All sequencing data were deposited at the SRA repository (NCBI) (Data Citation 2). Samples were grouped according to type of RNA fraction (ribosome-free vs. polysome-bound) using principal component analysis (PCA) (Fig. 4a) and according to day of differentiation (D0, D1, D4, D9 and D15) (Fig. 4b, c), indicating the reproducibility of biological replicates. Additionally, RPKM values of developmental markers were plotted on a heatmap (Fig. 4d) to show the specificity of mesoderm commitment among the three germ layers. RPKM values of cardiac markers were also plotted (Fig. 4e) to show the higher cardiomyocyte marker expression when compared to markers of endothelial (EC) and smooth muscle (SMC) cells, other cardiac progenitor derivatives.

qPCR validation

To identify if our RNA-seq data were compatible with cardiomyogenesis gene expression, we prepared total RNA samples from the same differentiation time-points for qPCR analysis of developmental and cardiac marker expression. Comparing the log₂ fold change of hESC on day 0 (D0), we demonstrated the similarity between RNA-seq and qPCR results. POU5F1 (OCT4) and NANOG are transcription factors expressed in pluripotency conditions which compose the pluripotency core regulatory circuitry²⁹. These genes represent markers for the pluripotent state and showed a gradual decrease in expression throughout differentiation in our RNA-seq and qPCR results (Fig. 4f). Developmental markers were also analyzed, as the T-box Brachyury/T, which has a conserved role in mesoderm differentiation³⁰, and Eomesodermin (EOMES), which expression marks the earliest cardiac mesoderm and promotes formation of cardiovascular progenitors³¹. The mesodermal markers T and EOMES showed increased expression on D4 (mesoderm stage) and were down-regulated on D9 and D15. Finally, the expression of cardiac-related genes such as GATA4, NKX2-5 and TNNT2 was increased during differentiation (Fig. 4f).

References

1. Matoulkova, E., Michalova, E., Vojtesek, B. & Hrstka, R. The role of the 3' untranslated region in post-transcriptional regulation of protein expression in mammalian cells. *RNA Biol.* **9**, 563–576 (2012).
2. Sonenberg, N. & Hinnebusch, A. G. New Modes of Translational Control in Development, Behavior, and Disease. *Mol. Cell* **28**, 721–729 (2007).
3. Tian, Q. *et al.* Integrated Genomic and Proteomic Analyses of Gene Expression in Mammalian Cells. *Mol. Cell. Proteomics* **3**, 960–969 (2004).
4. Kuersten, S., Radek, A., Vogel, C. & Penalva, L. O. F. Translation regulation gets its 'omics' moment. *Wiley Interdiscip. Rev. RNA* **4**, 617–630 (2013).
5. Sunnerhagen, P. Cytoplasmatic post-transcriptional regulation and intracellular signalling. *Mol. Genet. Genomics* **277**, 341–355 (2007).
6. Schwanhäusser, B. *et al.* Global quantification of mammalian gene expression control. *Nature* **473**, 337–342 (2011).
7. Tebaldi, T. *et al.* Widespread uncoupling between transcriptome and translome variations after a stimulus in mammalian cells. *BMC Genomics* **13**, 220 (2012).
8. Kelen, K. Van Der, Beyaert, R., Inzé, D. & Veylder, L. De. Translational control of eukaryotic gene expression. *Crit. Rev. Biochem. Mol. Biol.* **44**, 143–168 (2009).
9. Faye, M. D., Graber, T. E. & Holcik, M. Assessment of Selective mRNA Translation in Mammalian Cells by Polysome Profiling. *J. Vis. Exp* **92**, e52295 (2014).
10. Panda, A. C., Martindale, J. L. & Gorospe, M. Polysome Fractionation to Analyze mRNA Distribution Profiles. *Bio Protoc* **7**, e2126 (2017).
11. Spangenberg, L. *et al.* Polysome profiling shows extensive posttranscriptional regulation during human adipocyte stem cell differentiation into adipocytes. *Stem Cell Res* **11**, 902–912 (2013).
12. Mikulits, W. *et al.* Isolation of translationally controlled mRNAs by differential screening. *FASEB J.* **14**, 1641–1652 (2000).
13. King, H. A. & Gerber, A. P. Translatome profiling: Methods for genome-scale analysis of mRNA translation. *Brief. Funct. Genomics* **15**, 22–31 (2016).
14. Kolle, G. *et al.* Deep-transcriptome and ribonome sequencing redefines the molecular networks of pluripotency and the extra-cellular space in human embryonic stem cells. *Genome Res.* 2013–2025 (2014).
15. Sampath, P. *et al.* A Hierarchical Network Controls Protein Translation during Murine Embryonic Stem Cell Self-Renewal and Differentiation. *Cell Stem Cell* **2**, 448–460 (2008).
16. Brade, T., Pane, L. S., Moretti, A., Chien, K. R. & Laugwitz, K. Embryonic heart progenitors and cardiogenesis. *Cold Spring Harb Perspect Med* **3**, 1–18 (2013).
17. Olson, E. N. Gene Regulatory Networks in the Evolution and Development of the Heart. *Science.* **313**, 1922–1927 (2006).
18. Kattman, S. J. *et al.* Stage-specific optimization of activin/nodal and BMP signaling promotes cardiac differentiation of mouse and human pluripotent stem cell lines. *Cell Stem Cell* **8**, 228–240 (2011).
19. Kehat, I. *et al.* Human embryonic stem cells can differentiate into myocytes with structural and functional properties of cardiomyocytes. *J. Clin. Invest.* **108**, 407–414 (2001).
20. Laflamme, M., *a et al.* Cardiomyocytes derived from human embryonic stem cells in pro-survival factors enhance function of infarcted rat hearts. *Nat. Biotechnol* **25**, 1015–1024 (2007).

21. Xu, C., Police, S., Rao, N. & Carpenter, M. K. Characterization and enrichment of cardiomyocytes derived from human embryonic stem cells. *Circ. Res.* **91**, 501–508 (2002).
22. Bergmann, O. *et al.* Dynamics of Cell Generation and Turnover in the Human Heart. *Cell* **161**, 1566–1575 (2015).
23. Burridge, P. W., Keller, G., Gold, J. D. & Wu, J. C. Production of de novo cardiomyocytes: Human pluripotent stem cell differentiation and direct reprogramming. *Cell Stem Cell* **10**, 16–28 (2012).
24. Filipczyk, A. A., Passiera, R., Rochata, A. & Mummery, C. L. Regulation of cardiomyocyte differentiation of embryonic stem cells by extracellular signalling. *Cell. Mol. Life Sci.* **64**, 704–718 (2007).
25. Elliott, D. A. *et al.* NKX2-5eGFP/w hESCs for isolation of human cardiac progenitors and cardiomyocytes. *Nat. Methods* **8**, 1037–1040 (2011).
26. Liao, Y., Smyth, G. K. & Shi, W. The Subread aligner: Fast, accurate and scalable read mapping by seed-and-vote. *Nucleic Acids Res* **41**, e108 (2013).
27. Yang, L. *et al.* Human cardiovascular progenitor cells develop from a KDR + embryonic-stem-cell-derived population. *Nature* **453**, 524–528 (2008).
28. Evseenko, D. *et al.* Mapping the first stages of mesoderm commitment during differentiation of human embryonic stem cells. *PNAS* **107**, 13742–13747 (2010).
29. Boyer, L. A. *et al.* Core transcriptional regulatory circuitry in human embryonic stem cells. *Cell* **122**, 947–956 (2005).
30. Technau, U. Brachyury, the blastopore and the evolution of the mesoderm. *BioEssays* **23**, 788–794 (2001).
31. Costello, I. *et al.* The T-box transcription factor Eomesodermin acts upstream of *Mespl* to specify cardiac mesoderm during mouse gastrulation. *Nat Cell Biol* **13**, 1084–1091 (2015).

Data Citations

1. Robert, A. & Stimamiglio, M. *FlowRepository* FR-FCM-ZYX8 (2018).
2. *NCBI Sequence Read Archive* SRP150416 (2018).
3. Pereira, I. T. *et al.* *figshare* <https://doi.org/10.6084/m9.figshare.c.4208357> (2018).

Acknowledgements

The authors thank the Program for Technological Development in Tools for Health-RPT-FIOCRUZ for the use of flow cytometry and microscopy facilities at Carlos Chagas Institute – Fiocruz/PR, the National Cancer Institute (Rio de Janeiro, Brazil) and Dr. David Elliot from Monash University (Australia) for kindly providing the cells. We also thank the financial support from CNPq: INCT-REGENERA grant 465656/2014-5. ITP and AWR received fellowships from CAPES and BD from CNPq.

Author Contributions

I.T.P. performed the experiments and wrote the manuscript. A.W.R. performed the cardiac differentiation assays and wrote the manuscript. L.S. performed the bioinformatic analyses and wrote the manuscript. R.A.H. performed the data mapping. M.A.S. coordinated the cardiac differentiation experiments. H.N. coordinated the bioinformatic analyses. B.D. conceived the concept and coordinated the work.

Additional Information

Competing interests: The authors declare no competing interests.

How to cite this article: Pereira, I. T. *et al.* Polysome profiling followed by RNA-seq of cardiac differentiation stages in hESCs. *Sci. Data*. 5:180287 doi: 10.1038/sdata.2018.287 (2018).

Publisher's note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>

The Creative Commons Public Domain Dedication waiver <http://creativecommons.org/publicdomain/zero/1.0/> applies to the metadata files made available in this article.

© The Author(s) 2018

CAPÍTULO II

Artigo

Cardiomyogenic differentiation is fine-tuned by differential mRNA association with polysomes

Situação – Publicado na revista BMC Genomics (março de 2019)

Pereira, I.T.; Spangenberg, L.; Robert, A.W.; Amorín, R.; Stimamiglio, M.A.; Naya, H.; Dallagiovanna, B. Cardiomyogenic differentiation is fine-tuned by differential mRNA association with polysomes. *BMC Genomics* **2019**, *20*, 1–16, doi:10.1186/s12864-019-5550-3.

Prefácio

Este capítulo apresenta o artigo publicado na revista BMC Genomics, o qual descreve a análise dos dados de RNA-seq ao longo da diferenciação cardíaca de CTE. Esse manuscrito teve como enfoques a associação diferencial de RNAs mensageiros aos polissomos nas diferentes etapas da diferenciação cardiomiogênica e a regulação pós-transcricional que os genes expressos estariam submetidos. Tais abordagens compunham os objetivos específicos desta Tese.

RESEARCH ARTICLE

Open Access



Cardiomyogenic differentiation is fine-tuned by differential mRNA association with polysomes

Isabela Tiemy Pereira¹, Lucia Spangenberg², Anny Waloski Robert¹, Rocío Amorín², Marco Augusto Stimamiglio¹, Hugo Naya² and Bruno Dallagiovanna^{1*}

Abstract

Background: Cardiac cell fate specification occurs through progressive steps, and its gene expression regulation features are still being defined. There has been an increasing interest in understanding the coordination between transcription and post-transcriptional regulation during the differentiation processes. Here, we took advantage of the polysome profiling technique to isolate and high-throughput sequence ribosome-free and polysome-bound RNAs during cardiomyogenesis.

Results: We showed that polysome-bound RNAs exhibit the cardiomyogenic commitment gene expression and that mesoderm-to-cardiac progenitor stages are strongly regulated. Additionally, we compared ribosome-free and polysome-bound RNAs and found that the post-transcriptional regulation vastly contributes to cardiac phenotype determination, including RNA recruitment to and dissociation from ribosomes. Moreover, we found that protein synthesis is decreased in cardiomyocytes compared to human embryonic stem-cells (hESCs), possibly due to the down-regulation of translation-related genes.

Conclusions: Our data provided a powerful tool to investigate genes potentially controlled by post-transcriptional mechanisms during the cardiac differentiation of hESC. This work could prospect fundamental tools to develop new therapy and research approaches.

Keywords: Cardiomyogenesis, hESC-derived cardiomyocytes, Polysome profiling, Post-transcriptional regulation

Background

Recent approaches regarding hESCs differentiation made it possible to mimic features of developmental biology and address the key events that regulate early lineage commitment [1]. Cardiomyocytes have been derived from hESCs as an *in vitro* model to study cardiomyogenesis and as an attempt to produce clinically relevant cell populations [2–5]. Modeling congenital abnormalities of the heart or testing the cardiac toxicity of new drugs strengthen a particular interest in deriving cardiovascular lineages from pluripotent stem cells [2]. Additionally, *in vitro* cardiomyocyte generation and promotion of an endogenous regenerative capacity offers new therapeutic

strategies to replace heart tissue damaged by age or disease [6, 7]. Activation of a specific genetic program is essential to drive cells into the cell type of interest and contribute to these challenging goals. Cardiac tissue formation is controlled by sequential gene regulatory steps that define specialized cell fates [8], although understanding of molecular signatures of intermediate differentiation states of the cardiomyogenic lineage are still lacking.

Gene expression regulation acts at both transcriptional and post-transcriptional levels. There has been an increasing interest in understanding how these regulatory mechanisms interact and coordinate along the differentiation processes [9]. Large-scale studies help to uncover the complex and dynamic multi-layered regulation involving chromatin modifications, transcriptional, and post-transcriptional networks that control the reconfiguration of ESC gene expression program when

* Correspondence: bruno.dallagiovanna@fiocruz.br

¹Basic Stem-cell Biology Laboratory, Instituto Carlos Chagas - FIOCRUZ-PR, Rua Professor Algacyr Munhoz Mader, 3775, Curitiba, PR 81.350-010, Brazil
Full list of author information is available at the end of the article



cells are stimulated to differentiate [10, 11]. Genome-wide expression profiling, using microarrays or RNA-seq analysis, has provided the possibility to investigate transcriptional changes during cardiomyogenic differentiation [12–17]. However, most of these studies were centered at total transcriptome analysis. Since the cellular abundance of proteins is also controlled at the translation level [18], the most-reliable gene expression patterns cannot be based solely on total RNA analysis.

The role of translational control as a key regulatory node in gene expression during development and disease is still being established. Actively translated mRNAs are predicted to be associated with many ribosomes, which form large complexes called polysomes [19]. These complexes define the translome, the subpopulation of mRNAs that are engaged with the translation machinery. Methods to infer the translome have helped to expand our knowledge of protein synthesis control and revealed post-transcriptional mechanisms involved in cell fate commitment [20–22]. A classical technique called polysome profiling has been used to assess the mRNAs bound to ribosomes through ultracentrifugation in a sucrose gradient, and further identify them by high-throughput methods [23–25]. Previously, we used polysome profiling to study the fate of human adipose stem cells (hASCs) and their commitment to adipogenesis [25, 26]. We showed that 60% of the genes, which were differentially expressed after 72 h of differentiation induction, were controlled by post-transcriptional regulation.

Here, we used polysome profiling to assess the ribosome-free and polysome-bound mRNAs during the developmental steps of cardiomyogenic commitment and investigate genes potentially controlled by post-transcriptional mechanisms. Our results showed that 60–80% of differentially expressed genes (DEG) showed some degree of post-transcriptional regulation, suggesting its crucial role in shaping of the developmental pathways. Additionally, committed cells showed decreased protein synthesis rates compared to hESC.

Results

Polysome profiling during hESCs cardiomyogenic differentiation

The hES-NKX2-5^{eGFP/w} reporter human embryonic stem cell (hESC) line [27] was used to derive cardiomyocytes using a developmentally staged protocol [2, 28] that includes the induction of a cardiac mesoderm population on days 3 and 4 and a NKX2-5⁺/cTNT⁺ population by day 15 (Fig. 1a and b). Cardiomyogenesis progression was followed by flow cytometry using CD56 as a mesoderm marker [29] on day 4 (cut off < 40%) and NKX2-5/eGFP expression on day 9 as a cardiac progenitor marker (cut off < 50%) (Fig. 1b). Beating clusters

were observed after 10 days of differentiation (Fig. 1c and Online Additional file 1: Video S1), yielding a population of 50–60% cTNT⁺ cardiomyocytes on day 15 (D15) (Fig. 1b). At day 20, cTNI immunostaining showed the striations characteristic of sarcomere structures (Fig. 1c).

To investigate the differential association of mRNAs with polysomes and, therefore, post-transcriptional changes in gene expression during cardiac differentiation, we first performed polysome profiling on days D0, D1, D4, D9 and D15, which represent pluripotency, embryoid body (EB) aggregation, cardiac mesoderm, cardiac progenitor and cardiomyocyte stages, respectively (Fig. 1a). After 10 min of cycloheximide treatment, active ribosomes got arrested with associated RNAs and we analyzed cellular extracts by ultracentrifugation in a sucrose gradient (Fig. 1d). Differential density throughout the gradient allowed the isolation of ribosome-free (fractions 1–3) and polysome-bound (fractions 9–22) RNAs. Based on the polysome profile, pooled ribosome-free and polysome-bound RNA fractions were sequenced using the Illumina platform, yielding nearly 30 million reads for each sample. Approximately 70–80% of the reads were mapped onto the reference genome (GRCh38), and more than 17,000 genes were detected in each type of fraction (Additional file 2: Figure S1). As a control, D15 cells were also treated with puromycin to disassemble the polysomes and cardiomyocyte markers were evaluated by qPCR (Additional file 2: Figure S2).

Correspondence analysis (COA) showed that samples were grouped according to the type of RNA fraction (ribosome-free vs. polysome-bound) (Fig. 1e) and according to the day of differentiation (D0, D1, D4, D9 and D15) (Fig. 1f and g). Polysome-bound samples showed more distinct groups relative to the day of differentiation, indicating high similarity between translated genes in experimental replicates (Fig. 1g). On the other hand, ribosome-free samples showed more dispersion and less similarity between the replicates.

The translome delineates cardiomyogenic gene expression

During heart development, temporal gene expression changes occur to define each step of cardiogenic commitment [6, 30]. RPKM values (reads per kilobase per million mapped reads) for polysome-bound RNA fraction showed the expression levels of known lineage-specific genes throughout cardiac differentiation (Fig. 2a) which were confirmed by qPCR (Additional file 2: Figures S2 and S3). Pluripotency marker genes, including *SOX2*, *POU5F1* (OCT4) and *NANOG*, were expressed at higher levels on D0 and D1 and down-regulated at following time-points. The mesoderm marker genes *T* and *EOMES*, and early cardiac gene *MESPI* were

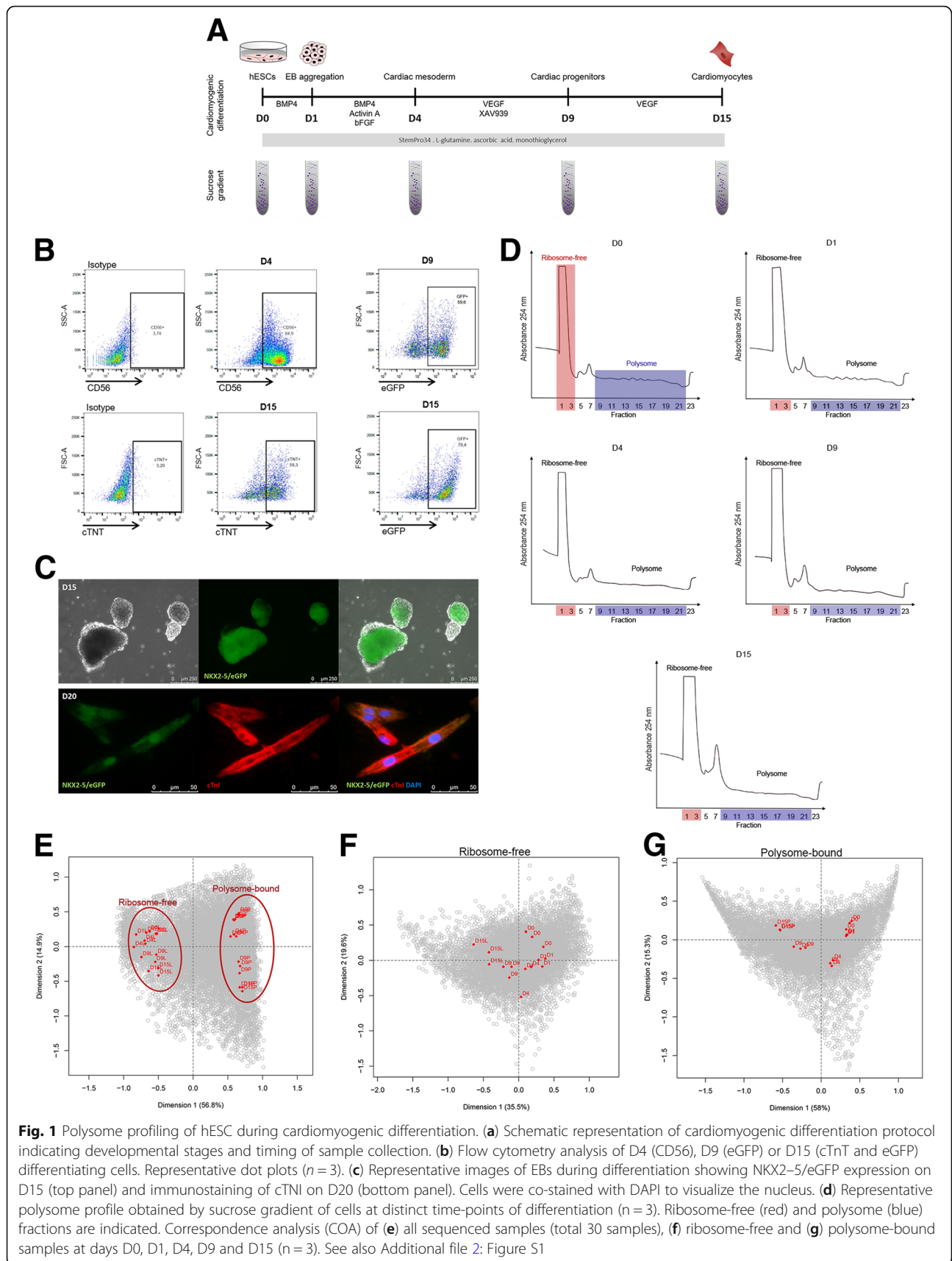


Fig. 1 Polysome profiling of hESC during cardiomyogenic differentiation. **(a)** Schematic representation of cardiomyogenic differentiation protocol indicating developmental stages and timing of sample collection. **(b)** Flow cytometry analysis of D4 (CD56), D9 (eGFP) or D15 (cTnT and eGFP) differentiating cells. Representative dot plots ($n = 3$). **(c)** Representative images of EBs during differentiation showing NKX2-5/eGFP expression on D15 (top panel) and immunostaining of cTnI on D20 (bottom panel). Cells were co-stained with DAPI to visualize the nucleus. **(d)** Representative polysome profile obtained by sucrose gradient of cells at distinct time-points of differentiation ($n = 3$). Ribosome-free (red) and polysome (blue) fractions are indicated. Correspondence analysis (COA) of **(e)** all sequenced samples (total 30 samples), **(f)** ribosome-free and **(g)** polysome-bound samples at days D0, D1, D4, D9 and D15 ($n = 3$). See also Additional file 2: Figure S1

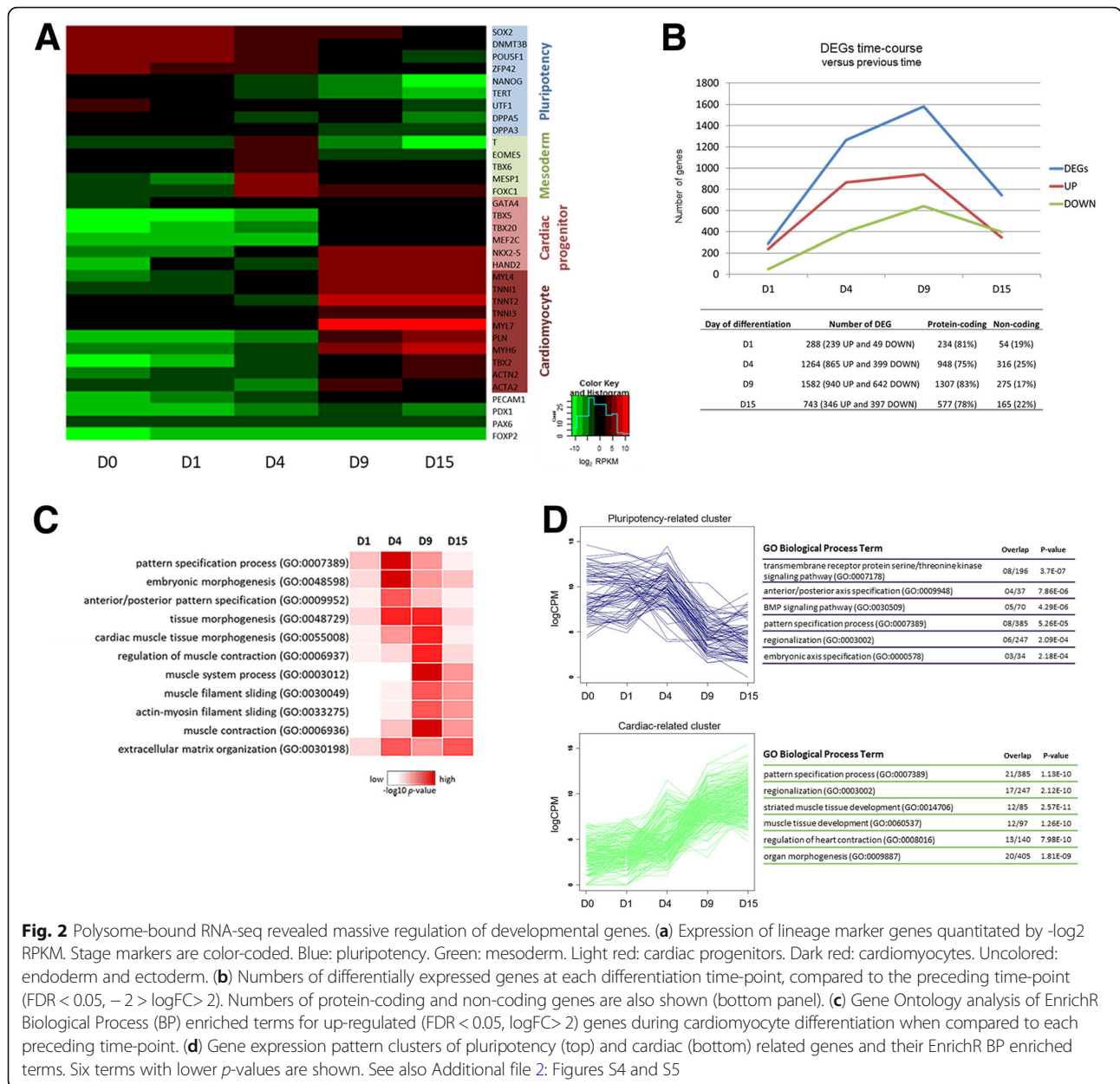


Fig. 2 Polysome-bound RNA-seq revealed massive regulation of developmental genes. **(a)** Expression of lineage marker genes quantitated by $-\log_2$ RPKM. Stage markers are color-coded. Blue: pluripotency. Green: mesoderm. Light red: cardiac progenitors. Dark red: cardiomyocytes. Uncolored: endoderm and ectoderm. **(b)** Numbers of differentially expressed genes at each differentiation time-point, compared to the preceding time-point ($FDR < 0.05$, $-2 > \log FC > 2$). Numbers of protein-coding and non-coding genes are also shown (bottom panel). **(c)** Gene Ontology analysis of EnrichR Biological Process (BP) enriched terms for up-regulated ($FDR < 0.05$, $\log FC > 2$) genes during cardiomyocyte differentiation when compared to each preceding time-point. **(d)** Gene expression pattern clusters of pluripotency (top) and cardiac (bottom) related genes and their EnrichR BP enriched terms. Six terms with lower p -values are shown. See also Additional file 2: Figures S4 and S5

highly specifically expressed on D4. Expression of cardiac-related genes such as *GATA4*, *NKX2-5* and *TBX5* as well as cardiomyocyte-specific sarcomeric genes *TNNI*, *TNNT*, *MYL7* and *MYH6* were observed from D9 to D15. In contrast, the endoderm (*PECAM1* and *PDX1*) and ectoderm (*PAX6* and *FOXP2*) marker genes did not change their polysome association during cardiomyogenic differentiation (Fig. 2a).

Comparisons between each differentiation time-point and the preceding time-point, considering an overall FDR of ≤ 0.05 and $-2 \geq \log FC \geq 2$, identified differentially expressed genes (DEGs) in polysome-bound RNA fractions (Fig. 2b, data available in Additional file 3).

Aggregation of embryoid bodies during the first 24 h of differentiation induced differential expression of 288 genes. Mesoderm commitment from D1 to D4 showed 1264 DEGs, and cardiac progenitor progression to D9 showed 1582 DEGs. The final step of differentiation analysis on D15 showed 743 DEGs compared to D9. In general, the majority of DEGs were up-regulated, except for D15 compared to D9, where the number of down-regulated genes was slightly higher than up-regulated genes (Fig. 2b). Similar numbers of genes and patterns of up- and down-regulated genes were shown in the ribosome-free samples (Additional file 2: Figure S4A and Additional file 4). Regarding the protein-coding and

non-coding genes in polysome-bound samples, approximately 20% of DEGs in each time-point analysis are annotated as non-coding RNAs, in which 39% correspond to “lincRNA”, 22% to “antisense” and 18% to “processed pseudogene” RNAs (Additional file 2: Figure S4B).

Gene Ontology (GO) analysis of polysome-bound up-regulated DEGs (FDR ≤ 0.05 and logFC ≥ 2) revealed developmental and cardiac-related “biological process” (BP) during cardiac differentiation (Fig. 2c). BP terms *p*-values (−log10) of DEGs for each time-point compared to the preceding one are represented in Fig. 2c heatmap. “Cardiac muscle tissue morphogenesis” and “regulation of muscle contraction” are highly enriched on D9 (compared to D4) and only slightly enriched on D15 (compared to D9), indicating that on D9, most cardiac characteristics are already committed. This pattern is similar for “muscle system process” and “muscle contraction”. “Extracellular matrix organization” seems to have important roles in two distinct phases: mesoderm commitment (D4) and cardiomyocyte final differentiation (D15) (Fig. 2c and Additional file 2: Figure S5).

To assess more information about variations in gene expression during cardiac differentiation, we performed gene clustering using logCPM (counts per million mapped reads). Distinct pattern expression groups were shown: genes with decreased expression during the differentiation, called pluripotency-related cluster and enriched in early developmental BP terms, such as “anterior/posterior axis

specification” and “BMP signaling pathway”; and genes with increased expression during cardiac differentiation, called cardiac-related cluster and enriched in lineage specific commitment terms, such as “muscle tissue development” and “regulation of heart contraction” (Fig. 2d).

Strong gene expression coordination is observed during mesoderm-to-cardiac progenitor commitment

When considering polysome-bound RNAs, the largest gene expression variation showing 1582 DEGs occurred during the D4 to D9 shift, which represents mesoderm-to-cardiac progenitor commitment (Fig. 2b). GO analysis revealed that some of D4 up-regulated BP terms were also enriched in the D9 down-regulated analysis (Fig. 3a and Additional file 2: Figure S5), such as “pattern specification process” and “embryonic morphogenesis”. These findings suggest a crucial gene expression regulation at this stage. Comparisons between D4 up-regulated and D9 down-regulated genes showed 217 in common, which are related to “mesoderm development” and “embryonic pattern specification” BP terms (Fig. 3b).

Moreover, an expression gene clustering pattern showed a distinct group of genes highly and specifically expressed on D4 (Fig. 3c). Those genes are called mesoderm-related genes and are enriched on developmental processes, such as “pattern specification process”, “regionalization” and “somitogenesis”. Altogether, these results indicate that the developmental progress of mesoderm-to-cardiac

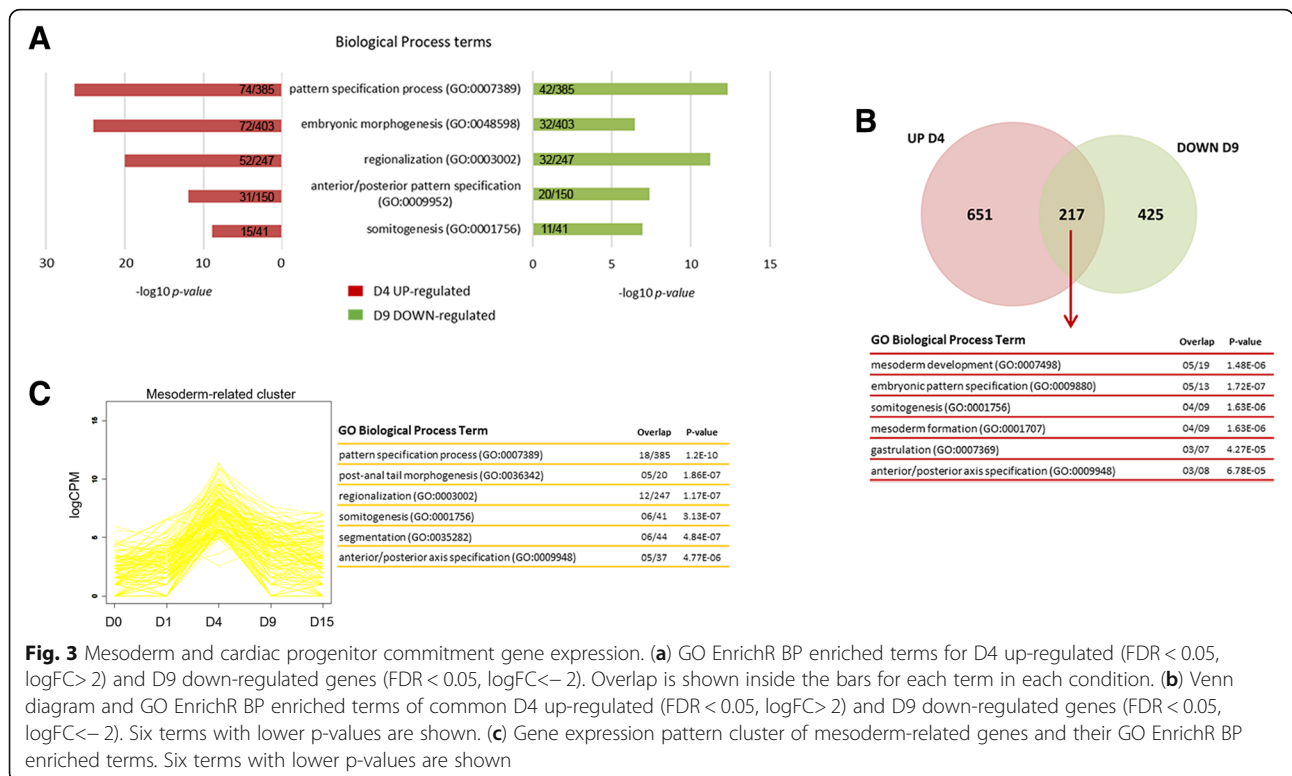
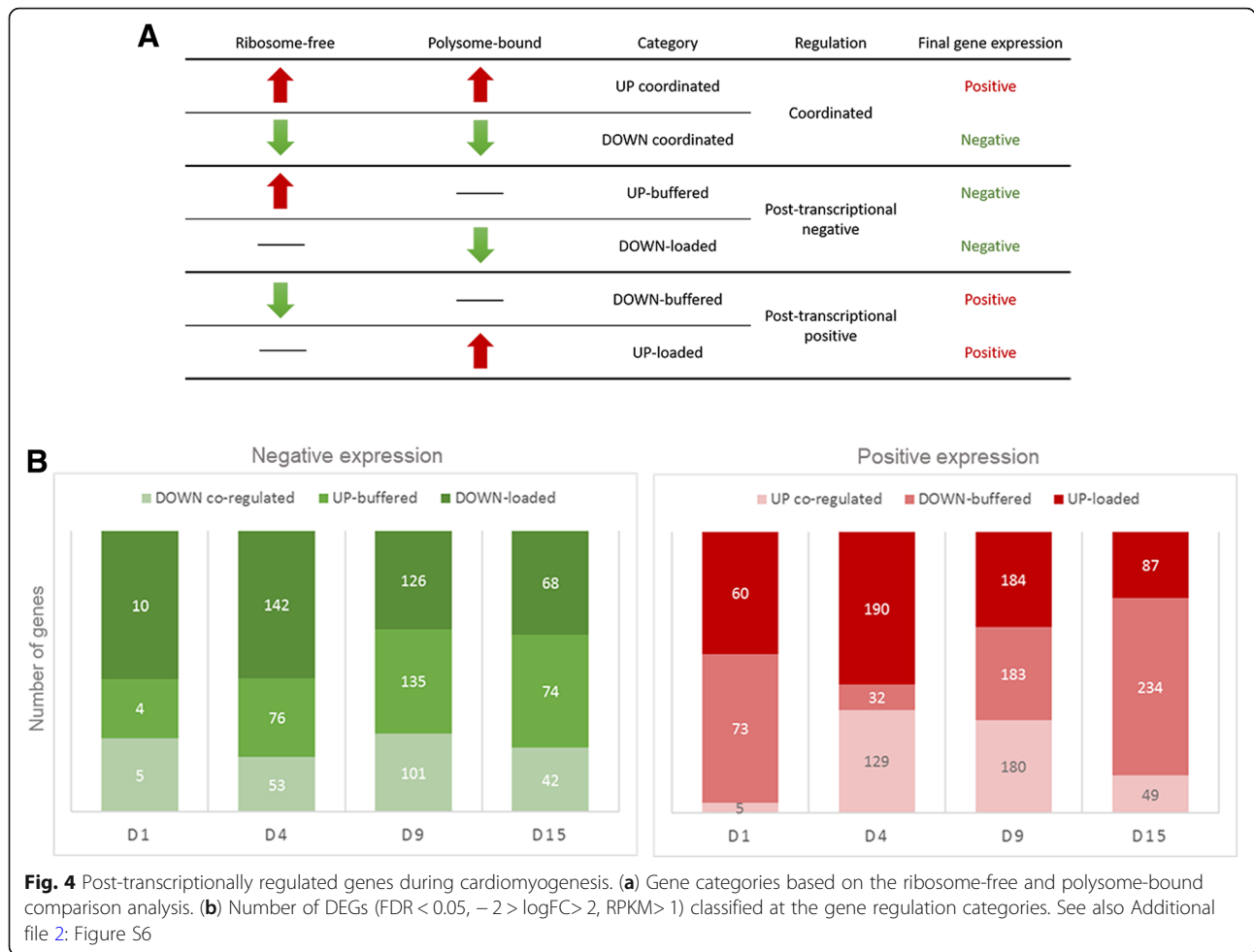


Fig. 3 Mesoderm and cardiac progenitor commitment gene expression. (a) GO EnrichR BP enriched terms for D4 up-regulated (FDR < 0.05, logFC > 2) and D9 down-regulated genes (FDR < 0.05, logFC < − 2). Overlap is shown inside the bars for each term in each condition. (b) Venn diagram and GO EnrichR BP enriched terms of common D4 up-regulated (FDR < 0.05, logFC > 2) and D9 down-regulated genes (FDR < 0.05, logFC < − 2). Six terms with lower p-values are shown. (c) Gene expression pattern cluster of mesoderm-related genes and their GO EnrichR BP enriched terms. Six terms with lower p-values are shown



progenitor is carefully regulated and can be assessed by polysome-bound RNA analysis.

Cardiac commitment is intensely tuned by differential mRNA association with polysomes

Gene expression fluctuations could be a consequence of coordination or lack thereof between transcription and translation rate changes. To assess the post-transcriptional regulation during the cardiomyogenic differentiation, we performed the two step-analysis: (1) DEGs were identified through the comparison between each differentiation time-point and the preceding time-point, considering an overall $FDR \leq 0.05$, $-2 \geq \log_{2}FC \geq 2$ and $RPKM > 1$ on ribosome-free or polysome-bound samples (Additional files 3 and 4, respectively); and (2) ribosome-free DEGs and polysome-bound DEGs were compared and classified according to the following categories (Additional file 5). Genes that were up- or down-regulated in both fractions were labeled “up-coordinated” or “down-coordinated”. Transcripts that were up-regulated in ribosome-free fraction but might be neutralized by post-transcriptional mechanisms were labeled “up-buffered”, or in the opposite case,

as “down-buffered”. Moreover, genes showing that their polysome association was increased or decreased, were labeled “up-loaded” or “down-loaded”, respectively (Fig. 4a).

We used the coordinated, buffered and loaded classification of DEGs and included one more label category considering the final gene expression as coordinated, post-transcriptional positive or negative regulation (Fig. 4a). Therefore, when compared to the preceding time-point, genes up- or down-coordinated were labeled being under coordinated regulation, genes up-buffered and down-loaded under post-transcriptional negative regulation and genes down-buffered and up-loaded under post-transcriptional positive regulation. Approximately 60–80% of DEGs showed one or another kind of post-transcriptional regulation, positive or negative, suggesting a crucial role of this level of gene expression control (Fig. 4b). Interestingly, during the initial steps of differentiation (D0-D1 and D1-D4), there was a prevalence of up- (10/19 and 142/271) or down-loaded (60/138 and 190/351) genes, suggesting a strong post-transcriptional regulation at these stages. During cardiac progenitor commitment on D9, numbers of coordinated, buffered and loaded genes

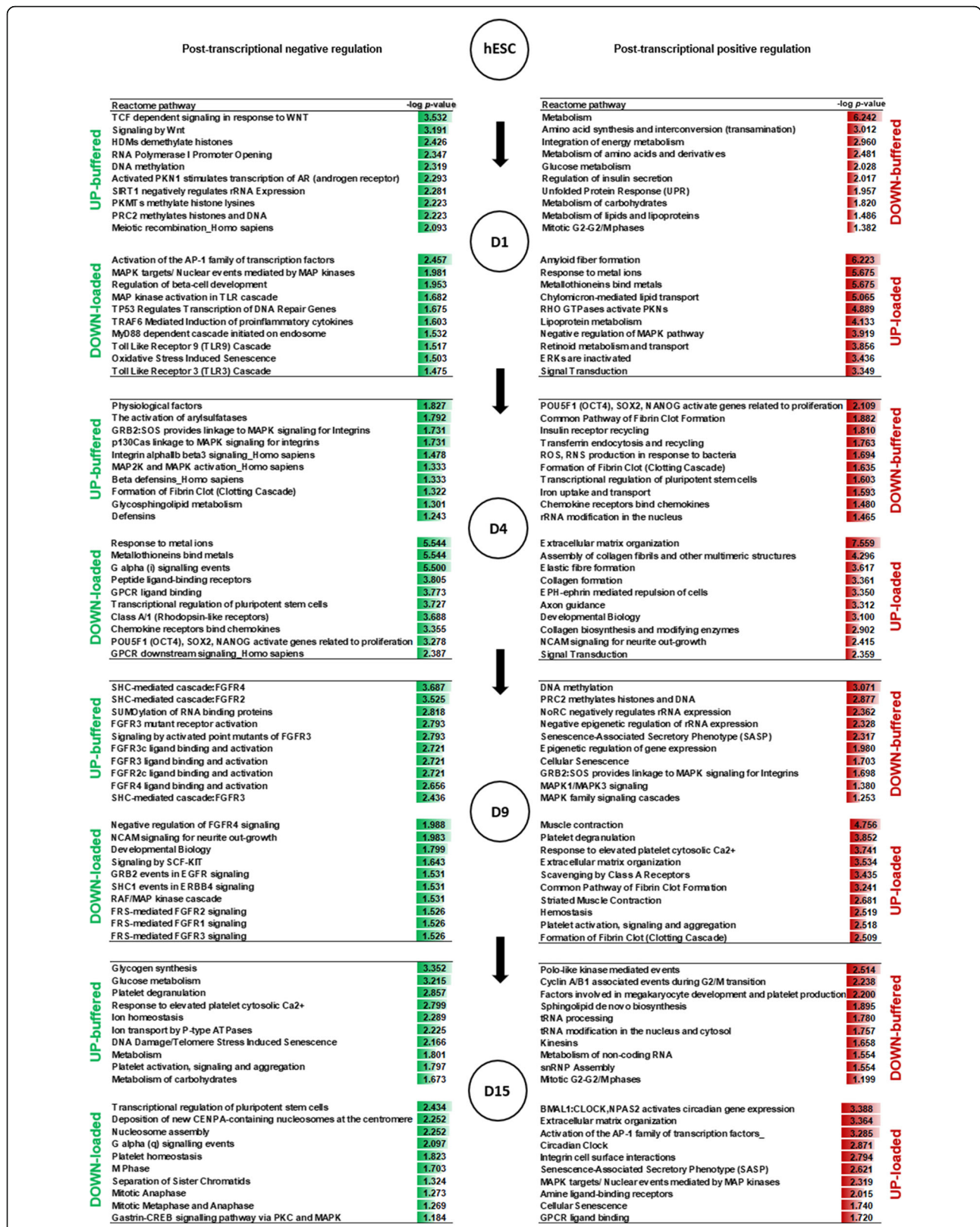
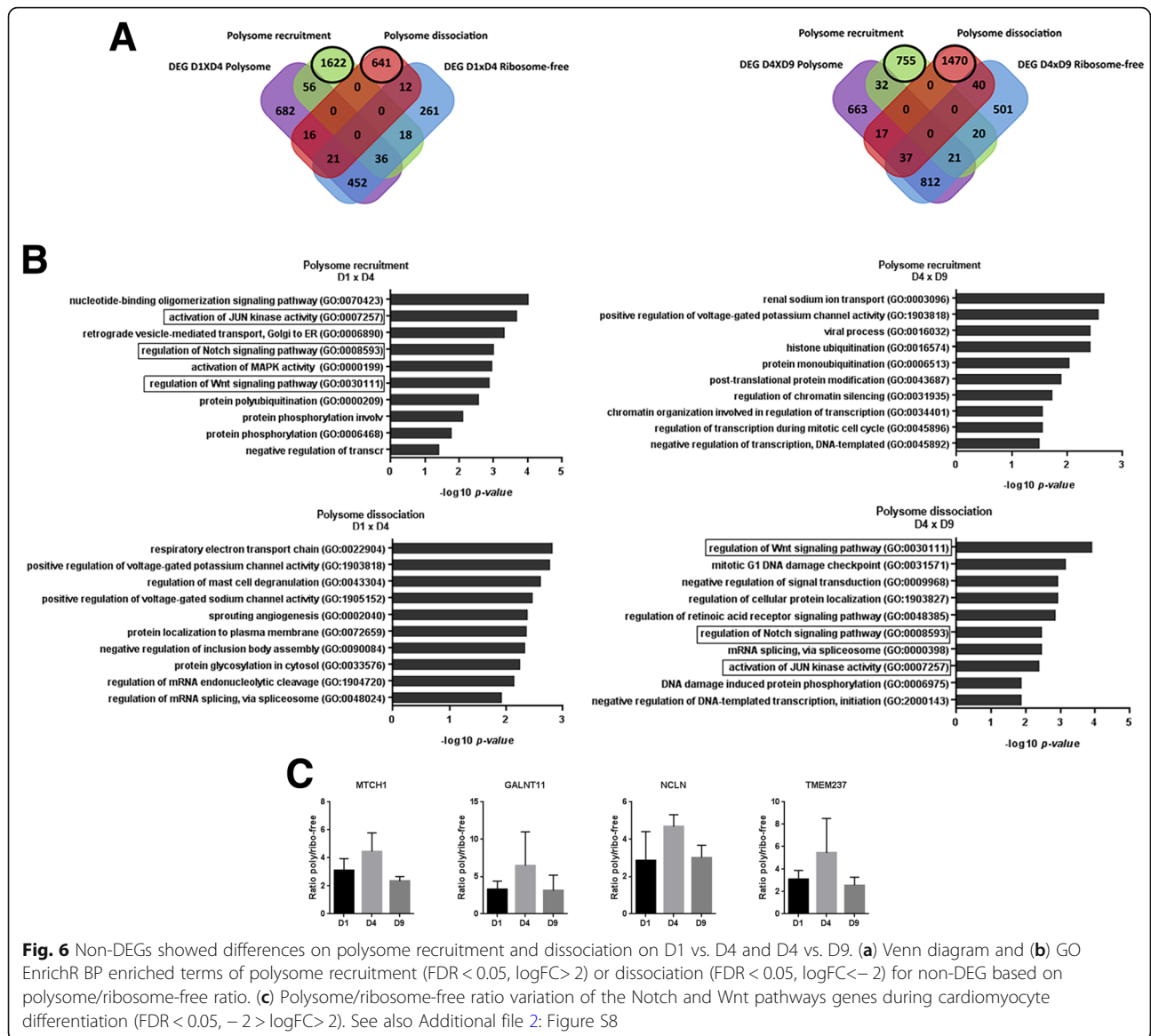


Fig. 5 Post-transcriptionally regulated genes during cardiomyogenic differentiation were involved in diverse biological processes. GO Reactome pathways enriched terms for up- (FDR < 0.05, logFC > 2, RPKM > 1) and down- (FDR < 0.05, logFC < -2, RPKM > 1) buffered and loaded genes. Ten terms with lower p-values are shown

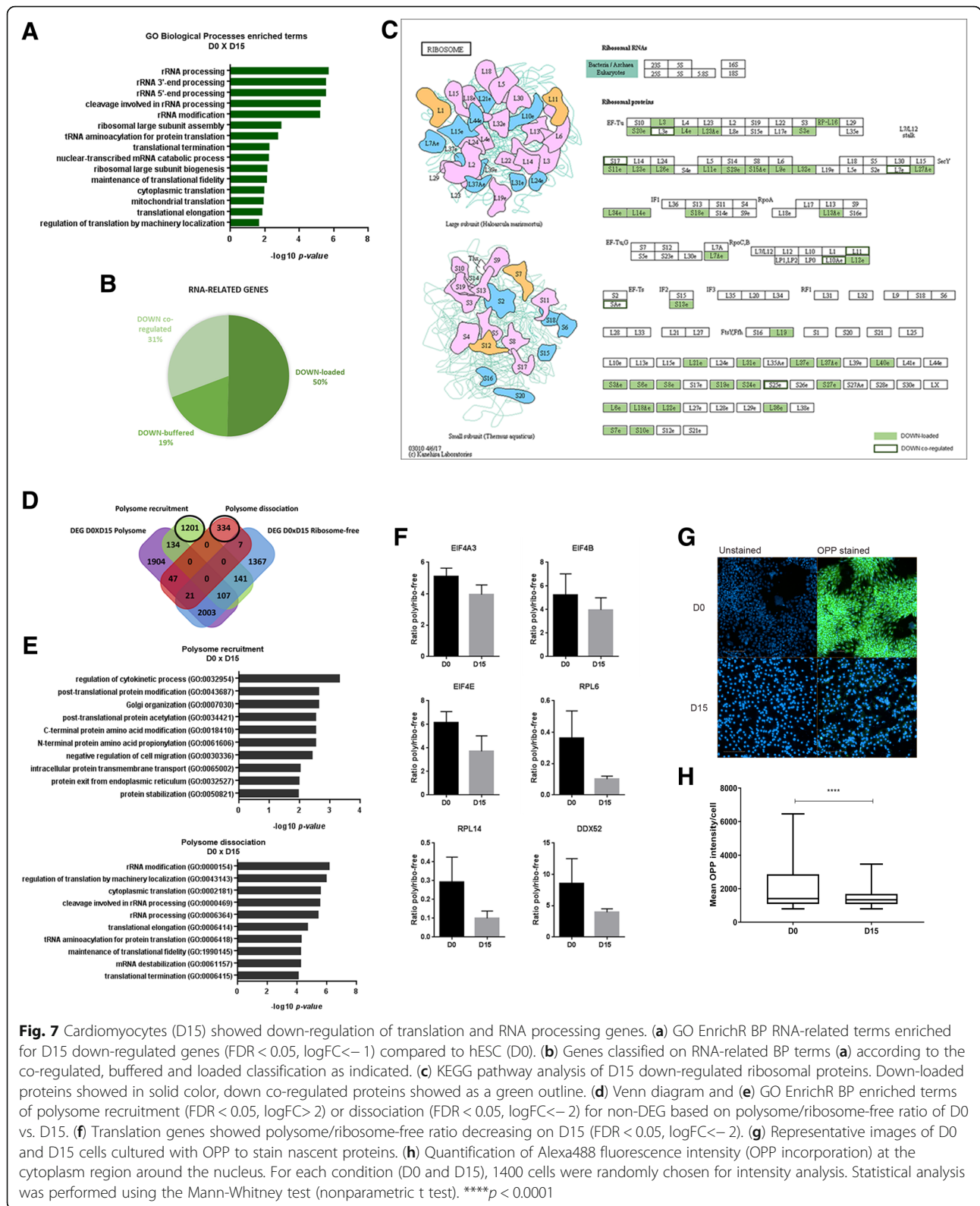


were similar, either on positive or negative regulation. On the other hand, between D9 and D15 most genes were classified as buffered, in particular, on positive regulation (234/370), indicating that transcriptional variations might be controlled by post-transcriptional mechanisms. These results corroborate our previous findings showing that most of cardiac characteristics are already committed on D9 (Fig. 2).

Considering that Gene Ontology-annotated genes are usually protein-coding, we filtered our data and performed GO analysis using only protein-coding genes. Analysis of coordinately regulated genes showed well-established pathways during hESC differentiation, such as the Reactome pathway terms “POU5F1 (OCT4), SOX2, NANOG repress genes related to differentiation”, “Transcriptional regulation of pluripotent stem cells” and “Developmental

Biology”, which were upregulated on D1 and D4 and down-regulated later (D9 and D15) (Additional file 2: Figure S6). Additionally, the cardiac-related pathway “Muscle contraction” and cardiac-specific pathway “Cardiac conduction” were coordinately up-regulated at D9 and D15 time-points (Additional file 2: Figure S6).

Genes classified as buffered or loaded showed enriched pathway terms with a diversity of biological processes (Fig. 5). The Reactome pathways “Developmental Biology” and “NCAM signaling for neurite out-growth” were up-loaded on D4 and down-loaded on D9, once more suggesting the critical regulation at this stage. For instance, developmental-related genes, such as *NOTUM*, *CER1* and *SOX17*, appeared as up-loaded on D4 and down-loaded on D9 (Additional file 5 and Additional file 2: Figure S7A), indicating the polysomal loading



regulation. The “M phase” and “Mitotic Metaphase and Anaphase” terms were shown as down-loaded on D15, while “Cyclin A/B1 associated events during G2/M transition” as down-buffered, indicating the fine adjustment of cell cycle during differentiation. The E2F Transcription Factor 1 plays a crucial role in the control of cell cycle [31] and was shown as involved in myoblast proliferation and differentiation through the auto-regulation loop with miR-20a-5p and miR-20b-5p [32]. E2F1 gene appeared less associated to polysomes on D9 (FDR < 0.05, logFC - 1.73, not included on down-loaded group), and down-buffered on D15, illustrating an initial polysomal dissociation followed by transcriptional down regulation of this gene (Additional files 3 and 4).

We have previously shown that “Extracellular matrix organization” seems to have an important role during differentiation (Fig. 2c), and it is probably also being post-transcriptionally regulated once it showed as enriched for up-loaded genes on D4, D9 and D15 (Fig. 5). Between the D15 up-loaded genes are *LUM*, *COL6A3* and *COL3A1* (Additional file 5). *COL3A1* was already shown as post-transcriptionally regulated by the interaction of the heterogeneous nuclear ribonucleoprotein (hnRNP) A1 mRNA-binding protein with its 3'-UTR [33]. Interestingly, the cardiac-specific *NKX2-5* gene was shown as up-buffered on D4, suggesting that its RNA was transcribed but not translated yet. On D9, this gene was shown as up-coordinated, suggesting, in this case, its transcription and translation (Additional file 2 and Additional file 5: Figure S7B). In addition, other crucial cardiac transcription factors, such as *MEF2A* and *TBX5* appeared as up-loaded on D9, indicating their polysome-binding increase on this stage (Additional file 5 and Additional file 2: Figure S7C).

To further characterize how recruitment to and dissociation from ribosomes influence gene expression, we performed polysome/ribosome-free ratio analysis using RPKM values. The results showed genes affected by polysome recruitment (FDR ≤ 0.05, logFC ≥ 2) or dissociation (FDR ≤ 0.05, - 2 ≥ logFC) during cardiomyocyte differentiation (Additional file 6). Some of them are also DEGs in polysome-bound or ribosome-free fractions, but most are not differentially expressed (Fig. 6a and Additional file 2: Figure S8). GO analysis for these non-differentially expressed and ratio-variated genes showed a variety of BP terms (Fig. 6b and Additional file 2: Figure S8). For instance, the development-related pathways JUN, Wnt and Notch were strongly regulated between D1, D4 and D9, were recruited from D1 to D4 and dissociated from D4 to D9 (Fig. 6b). The polysome/ribosome-free ratio of the JUN, Wnt and Notch pathway genes *MTCH1*, *GALNT11*, *NCLN* and *TMEM237* were plotted to visualize the variations on D1, D4 and D9 as an example (Fig. 6c).

mRNA loading into polysomes fine-tunes crucial processes during hESC cardiomyogenesis

To better understand the changes in translation between hESC and cardiomyocytes, we performed Gene Ontology (GO) analysis with DEGs of D0 (hESC) vs. D15 (cardiomyocytes), considering FDR ≤ 0.05, - 1 ≥ logFC ≥ 1 and ribosome-free and polysome-bound data combined (Additional file 7). Genes down-regulated on D15 were enriched in Biological Process (BP) terms such as “rRNA processing”, “tRNA aminoacylation for protein translation” and “cytoplasmic translation”, which were grouped as “RNA-related terms” (Fig. 7a). Combining the genes annotated in this group (143 genes down-regulated on D15) and analyzing their post-transcriptional regulation classification, 44 (30.7%) of them were down-coordinated and 72 (50.3%) were down-loaded (Fig. 7b and Additional file 7). In addition, many ribosomal proteins were shown down-regulated on D15, predominantly down-loaded (90.9%) (Fig. 7c). Other translation machinery proteins were also down-regulated after cardiomyocyte commitment, for instance, the initiation factors *EIF5A1* and *EIF4E1B*, and elongation factor *EEF1E1* (Additional file 7). On the other hand, the cardiac elongation factor *EEF1A2* [34] was up-regulated on D15 (Additional file 7).

Moreover, to explore if non-differentially expressed genes could have been affected by the variation in polysome occupancy (recruitment vs. dissociation), we also compared the polysome/ribosome-free RPKM ratio between hESCs (D0) and cardiomyocytes (D15) (Additional file 6). Interestingly, genes related to translation processes also showed polysomal dissociation on D15 when compared to D0, illustrated by the decreased polysome/ribosome-free ratio (Fig. 7d and e). Among them, there are translation initiation factors (*EIF4A3*, *EIF4E*, *EIF4B*), ribosomal proteins (*RPL6*, *RPL14*) and RNA helicases (*DDX52*), whose ratios are plotted in Fig. 7f (Additional file 6). Some of these observations were confirmed by qPCR (Additional file 2: Figure S9).

To further confirm that the down-regulation of translation-related genes after cardiomyogenic differentiation could affect protein synthesis, we performed a protein synthesis quantification assay. Cells on D0 (hESC) and D15 (cardiomyocytes) were treated with O-propargyl-puromycin (OPP) which is incorporated into newly translated proteins and then fluorescently labeled. Quantification of fluorescence intensity showed a decrease in protein synthesis after cardiac commitment compared to undifferentiated cells (Fig. 7g and h). Taken together, these findings suggest a translation adjustment during hESC-to-cardiomyocyte differentiation.

Interestingly, other crucial cellular processes also showed regulation by post-transcriptional mechanisms.

For instance, on D1 down-buffered and D15 up-buffered GO analysis, the Reactome pathway terms related to cellular metabolism were enriched, suggesting an important post-transcriptional regulation of this process during cardiomyogenic differentiation. Metabolic properties differ between cardiomyocytes and hESCs [35], thus, to understand the regulation of metabolic genes, we grouped the genes annotated as the Reactome pathway terms related to cellular metabolism and called them “Metabolism-related genes” (Additional file 2: Figure S10A and Additional file 7). Comparing the ribosome-free and polysome-bound data, 494 genes were up-regulated on D15, among which 243 (49.2%) were up-loaded, 46 (9.3%) were up-buffered and 205 (41.5%) were up coordinated (Additional file 2: Figure S10B). This likely reflects the post-transcriptional contribution to cardiomyocyte metabolic remodeling demonstrated by metabolic gene recruitment to polysomes.

Discussion

In this report, we differentiated hESC to cardiomyocytes and analyzed distinct time-points during this process to assess the temporal cell fate transition. The polysome profiling approach followed by ribosome-free and polysome-bound RNA-seq allowed us to evaluate gene regulation during the cardiogenic commitment. Polysome profiling analysis has been used as a robust method to assess the association of ribosomes with mRNAs, providing information about their translational status [23–25]. Here, we show that polysome-bound RNAs reflect the cardiac commitment phenotype, illustrated by down-regulation of pluripotency core regulatory circuitry (*OCT4*, *SOX2* and *NANOG*) [36] followed by up-regulation of cardiomyogenesis-related genes. The recapitulation of developmental steps is a powerful strategy to control a specific cell fate [1], where the first step is the transition into one of the three embryonic germ layers. The heart originates from the mesoderm emerging from the primitive streak [30]. Mesodermal development genes were found to be strongly regulated on D4, which represents the cardiac mesoderm commitment time-point. The massive number of DEGs between D4 and D9, added to the up-regulation of developmental pathways and pattern specification genes on D4 followed by their down-regulation at progenitor specification stage (D9) are consistent with the complexity of multiple mesodermal lineage choices, recently mapped by Loh et al. (2016) [37].

Moreover, we showed that 60–80% of DEG along cardiomyogenic differentiation were under some degree of post-transcriptional regulation. Uncoupling between the transcriptome and translome changes, characterized by the independency of the machineries responsible for mRNA availability and engagement in translation, has

been observed in most cell types [38]. Here, we observed that many metabolic and cellular processes are regulated exclusively at the level of polysomal association, probably because of post-transcriptional regulatory mechanisms in cellular development [25, 39, 40]. Surprisingly, genes with coordinated regulation (additive regulation) represent a minority of DEGs. In addition to that, changes in the abundance of mRNAs do not directly influence their polysomal engagement, represented by the buffered expression genes. The mRNA access to the translation machinery is governed by the interaction of translation factors with its sequence and structural motifs and is mediated by a complex network of trans-acting proteins and regulatory RNAs [41, 42]. While mRNA levels in the cytoplasm can randomly fluctuate due to transcriptional bursts, the final gene expression levels are adjusted at the time of protein synthesis [43]. McManus et al. (2013) [44] showed that buffering effects on gene expression (translation efficiency opposite to mRNA abundance) were common and would reduce divergent interspecies expression at the protein level. Translational regulation has already been reported in mouse mesoderm commitment [45] and hESC neuronal differentiation [46]. Here, we show unprecedented data about post-transcriptional regulation during human cardiomyogenic commitment.

We also found that the translational activity is reduced during differentiation; this became more evident when comparing hESC on D0 to cardiomyocytes on D15. Protein synthesis rate in cardiac tissue decreases from fetal to adult development, and in adult heart is lower than in other tissues [47, 48]. Chorghade et al. (2017) showed that the translation initiation RNA-binding protein PABPC1 is post-transcriptionally down-regulated in the adult heart tissue, which possibly decreases the translational capacity of the heart. Our data showed that the D15 down-regulated genes were enriched in RNA processing and translation genes, which were mostly down-regulated only in the polysome-bound fraction. Translation-related genes also showed polysomal dissociation on D15 when compared to D0, illustrated by the decreased polysome/ribosome-free ratio, further suggesting a strong post-transcriptional regulation and corroborating previous observations [38, 49, 50].

Protein synthesis is an indispensable process for maintaining homeostasis in the cell, and aberrations in its regulation may contribute to a wide range of diseases [51]. We have previously investigated the translational regulation during cell commitment to adipogenesis in hASCs. Our results showed the translational control as a key mechanism regulating the early steps of adipogenic differentiation, with a significant reduction in protein synthesis and a lower translational efficiency of ribosomal proteins [52]. In vitro differentiation of other cell types also showed decreasing translation rate, due to,

e.g., eIF2a phosphorylation or rRNA biogenesis [53–57]. Blair et al. (2017) reported the translational down-regulation of translation-related genes after hESC neuronal differentiation (including translation initiation factors and ribosomal proteins) [46], indicating that this mechanism doesn't seem to be cardiac-specific. In contrast, Blanco et al. (2016) showed that adult stem cells have lower protein synthesis rates than committed cells, using skin as a model. In normal skin, the RNA methyltransferase NSUN2 expression is restricted to committed hair follicle populations. They have demonstrated that loss of NSUN2 causes hypomethylation of tRNAs, accumulation of 5' tRNA fragments, which repress cap-dependent protein translation [58–60]. This reduced translation in adult stem cells is probably related to the biology of these specific cell types. Adult stem and progenitor cells are metabolically quiescent and low, if at all, proliferative. After stimulation by injury, they enter an activation state where they proliferate and show increased translational rates [61]. These particular characteristics make the comparison of the regulatory pathways and mechanisms underlying translational regulation in adult and embryonic stem cells not possible as these cells are in different stages of differentiation and in completely different biological environments. Our model recapitulates embryonic differentiation, where highly proliferative stem cells commit and differentiate into cell types with defined gene expression patterns and, in most cases, lower proliferative rates.

Noncontractile stem cells require less energy than beating cardiomyocytes, therefore, a transition in energetic infrastructure is necessary to support the increased energetic needs during cardiac differentiation [62, 63]. This switch in energy metabolism is associated with increased mitochondrial maturation and oxygen consumption and reduced glycolysis, due to a metabolic transcriptome remodeling [35, 62]. We showed that the metabolic reconfiguration is also a consequence of post-transcriptional level gene regulation. Almost 50% of metabolism-related genes regulated on D15 were increased only in polysome-bound fraction, characterizing the higher recruitment of these genes to the translational machinery. Increasing evidence has been showing the role of translational control in regulating metabolic function and that its defect is implicated in the pathogenesis of metabolic disorders [64, 65]. Additionally, cardiomyocyte mRNA interactome revealed that metabolic enzymes can act as RBPs [66], and energy metabolism proteins were found associated with ribosomes in ESCs, for instance, controlling the translation of mRNAs [67]. All these lines of evidence suggest a close relationship between cellular metabolism and translational control which deserves further exploration.

Conclusions

Polysome profiling followed by ribosome-free and polysome-bound RNA-seq of hESC cardiogenic commitment allowed us to evaluate gene expression regulation, revealing a massive control of developmental steps. Post-transcriptional regulation might be acting on 60–80% of DEG along cardiomyogenic differentiation, contributing to the final phenotype. Cardiomyocytes translational activity was reduced when compared to undifferentiated cells showing that translation regulation is also controlled during differentiation. Understanding the biological processes that trigger differentiation is crucial for an improvement of in vitro differentiation strategies and discovery of endogenous regenerative pathways. Our results helped to uncover transcriptional and post-transcriptional regulation involved in cardiomyogenic commitment gene expression, collected unprecedented data regarding polysome-bound RNAs and provided advances in comprehension of differentiation processes. Altogether, this work could prospect fundamental tools to develop new therapy and research approaches.

Methods

Cell culture and cardiomyocyte differentiation

hES-NKX2-5^{eGFP/w} cell line was a gift from Dr. David A. Elliott (Monash Immunology and Stem Cell Laboratories, Monash University, Australia) [27]. Cells were maintained on irradiated MEFs (mouse embryonic fibroblasts) using DMEM-F12 (Gibco™) with 20% KSR (Gibco™), 0.1 mM non-essential amino acids, 55 μM β-mercaptoethanol, 100 μg/mL penicillin-streptomycin, 2 mM L-glutamine and 10 ng/mL bFGF. The cardiomyocyte differentiation protocol was adapted from the previously published one [2]. Seventy-two hr. before differentiation, 7×10^5 cells/well were passaged onto a 6-well plate coated with Growth Factor Reduced BD® Matrigel Matrix for feeder depletion. Then, cells were dissociated with collagenase I (1 mg/mL) for 20 min followed by trypsin-EDTA (0.05%) for approximately 30 s and scraped to form small clusters. After washing, the clusters were cultured in 6-well Ultralow Attachment Corning plates using StemPro-34 (Invitrogen), supplemented with 100 μg/ml penicillin-streptomycin, 2 mM L-glutamine, 150 μg/mL transferrin, 50 μg/mL ascorbic acid and 0.45 mM monothioglycerol (MTG) and kept in humid incubator at 37 °C, 5% CO₂ and 5% O₂. For embryoid bodies (EBs) aggregation (D0-D1), the basal medium was supplemented with 1 ng/mL BMP4. On D1, the medium was replaced with supplementation of 10 ng/mL BMP4, 6 ng/mL Activin A and 5 ng/mL bFGF; on D4 with 10 μg/mL VEGF and 10 μM XAV 939, and starting on D8, the medium was replaced every two or three days supplemented with 10 μg/mL VEGF and 1 ng/mL BMP4. Three

independent differentiation assays were used as experimental replicates.

Flow cytometry

EBs were dissociated on D4 using trypsin-EDTA (0.05%) and incubated with the surface marker PE-conjugated anti-CD56 (1:25 in 0.5% PBS/BSA, BD cat. 347,747) and 1 $\mu\text{g}/\mu\text{L}$ DAPI. On D9, cells were disaggregated with trypsin-EDTA (0.05%) for 5 min and resuspended in PBS to evaluate eGFP expression. On D15, EBs were disaggregated using 1 mg/mL collagenase I for 16 h and trypsin-EDTA (0.05%) for 5 min, fixed with 4% paraformaldehyde, permeabilized with 0.5% Triton X-100 and incubated with anti-troponin T antibody (1:100 in 0.5% PBS/BSA, cardiac isoform Ab-1, Thermo Scientific™, cat. #MS-295-P0) followed by Pacific Blue-conjugated anti-mouse antibody (11000). Analyses were carried out using a FACSCanto II flow cytometer and FlowJo software.

Immunofluorescence and fluorescent microscopy

On D15, EBs were disaggregated using 1 mg/mL collagenase I for 16 h and trypsin-EDTA (0.05%) for 5 min and plated on Matrigel-coated wells. After 2–5 days, cardiomyocytes were fixed with 4% paraformaldehyde, permeabilized and blocked with 0.5% Triton X-100 and 1% PBS-BSA. Overnight incubation with anti-troponin I antibody (1:100 in 0.5% PBS/BSA, Santa Cruz Biotechnology, cat.: sc-15,368) followed by Alexa 546-conjugated anti-rabbit IgG (1:800, Invitrogen) and 1 $\mu\text{g}/\mu\text{L}$ DAPI. EBs or fixed cardiomyocytes were visualized using a Leica DMI6000B optical microscope and images and videos acquired by LAS AF software.

Polysome profile and RNA isolation

At the indicated time points, monolayer hESC on D0 or differentiating EBs were treated with 0.1 mg/mL cycloheximide (Sigma-Aldrich) for 10 min at 37 °C, disaggregated with trypsin-EDTA (0.05%) for 10 min and washed twice with PBS. Cells were resuspended in polysome lysis buffer (15 mM Tris HCl, pH 7.4, 15 mM MgCl₂, 300 mM NaCl, 1% Triton X-100, 40 U/ μL RNAse Out, 24 U/mL DNase and 100 $\mu\text{g}/\text{mL}$ cycloheximide), incubated for 10 min on ice and centrifuged at 12000 \times g for 10 min at 4 °C. For the puromycin control, cells were treated with 2 mg/mL puromycin for 1 h before disaggregated and the buffers were used without cycloheximide. The supernatants were loaded onto 10 to 50% sucrose gradients (prepared with BioComp model 108 Gradient Master) and centrifuged at 150000 \times g (SW40 rotor, HIMAC CP80WX HITACHI) for 160 min at 4 °C. Different sucrose gradient fractions were separated using ISCO gradient fractionation system (ISCO Model 160 Gradient Former Foxy Jr. Fraction Collector), connected to a UV detector, which monitored the absorbance at

275 nm to record the polysome profile. RNA from ribosome-free and pooled polysomal fractions was isolated using the Direct-zol RNA MiniPrep (Zymo Research), following the manufacturer's instructions.

High-throughput sequencing and data analysis

For cDNA library preparation, 200 to 500 ng of ribosome-free or 2 μg of polysome-bound RNA were used to perform three independent sample replicates. The cDNA libraries were prepared using the TruSeq Stranded mRNA Sample Preparation kit (Illumina, Inc.), and RNA-seq was carried out in an Illumina HiSeq platform. Mapping and counting of sequencing data was performed with the Rsubread package [68] against the new version of the human genome GRCh38. Mapping parameters were set for unique mapping of the reads (the rest were default). For some calculations, CPM values (counts per million) were determined, where each sample was normalized to one million reads to account for the library size (Additional file 8).

To assess the quality of the experiment and reproducibility of results, we performed a correspondence analysis (COA), a dimension reduction method of the matrix of counts. In COA, it is possible to simultaneously visualize samples and genes, revealing associations between them.

Differential expression analysis was done using the Bioconductor R package edgeR [69]. Several comparisons were performed for both ribosome-free and polysome-bound RNA fractions - each sample against the preceding time-point: D0 vs D1, D1 vs D4, D4 vs D9 and D9 vs D15; and D0 vs D15. For these analyses, we retained only those genes with at least one count per million in at least three samples. After a normalization procedure using three recommended methods (estimateGLMCommonDisp, estimateGLMTrendedDisp, estimateGLMTagwiseDisp), differential expression analysis for all 10 comparisons was performed using the generalized linear mixed model (glmFit and glmLRT). Correction for multiple testing was performed with FDR and RPKM values for each sample were also determined to compare the expression between the samples (Additional file 3, Additional file 4 and Additional file 7).

Clustering was performed with the k-means algorithm on the table of read log-counts of each gene. Biological replicates were averaged (mean of 3 replicates) and only differentially expressed genes in either condition (D0 vs D1, D1 vs D4, D4 vs D9, D9 vs D15) were considered. Different numbers of k were tested by measuring the within sum of squares. After visual inspection of within sum of squares distribution, we decided on k = 9.

Polysome/ribosome-free ratio was determined by dividing RPKM values derived from the polysomal fraction by the ones derived from the ribosome-free experiment (three biological replicates were averaged).

Gene ontology (GO) analysis was performed using Enrich R (<http://amp.pharm.mssm.edu/Enrichr/>).

Protein synthesis rate analysis

Protein synthesis was measured using Click- iT[®] Plus OPP Protein Synthesis Assay (Molecular Probes, Grand Island, NY). Cells on D0 (hESC) or D15 (cardiomyocytes) were seeded in matrigel-coated 96-well plates, and after 48 h, the staining and detection was performed following the manufacturer's instructions. The quantitative analysis was performed using an Operetta HTS imaging system (PerkinElmer, Waltham MA, USA). Images of 25 fields per well were evaluated with Harmony 3.5.2 software (PerkinElmer). Fluorescence intensities were measured at the cytoplasm regions around the nucleus. For each condition, 1400 cells were randomly chosen for intensity analysis.

Additional files

Additional file 1: Video of beating D15 EBs. (MP4 1974 kb)

Additional file 2: Figure S1. Summary of high-throughput ribosome-free and polysome-bound RNA-seq of distinct cardiomyogenic differentiation time-points ($n = 3$). **Figure S2.** (A) Polysome profiling of D15 cells treated with cycloheximide or puromycin. Ribosome-free (fractions 1–3), monosome (fractions 5–7), light polysomes (fractions 9–15) and heavy polysomes (fractions 16–22) fractions were pooled and isolated. (B) Cardiomyocyte markers evaluated by qPCR on distinct polysome fractions. **Figure S3.** Polysome-bound validation of developmental markers expression using qPCR. **Figure S4.** (A) Number of differentially expressed genes on each differentiation time-point, compared to previous time-point ($FDR < 0.05$, $-2 > \log_2 FC > 2$) on ribosome-free samples. Number of protein-coding and non-coding genes are also shown (bottom panel). (B) Non-coding genes categories of DEGs (all time-points combined, each time-point against its previous for analysis) ($FDR < 0.05$, $-2 > \log_2 FC > 2$) on polysome-bound samples. **Figure S5.** Gene Ontology EnrichR BP enriched terms for up ($FDR < 0.05$, $\log_2 FC > 2$) and down ($FDR < 0.05$, $\log_2 FC < -2$) polysome-bound regulated genes. **Figure S6.** Genes coordinately regulated are under control of transcriptional and post-transcriptional regulation during cardiomyogenic differentiation. **Figure S7.** qPCR validation of DEG during cardiomyogenic differentiation. **Figure S8.** Non-DEGs showed differences on polysome recruitment and dissociation on D0 vs. D1 and D9 vs. D15. **Figure S9.** RNA related-genes validation by qPCR. **Figure S10.** Cardiomyocytes (D15) showed up-regulation of cellular metabolism genes. (DOCX 2575 kb)

Additional file 3: Polysome-bound RNA-seq data of comparisons of each sample against the preceding time-point ($FDR < 0.05$). (XLSX 5476 kb)

Additional file 4: Ribosome-free RNA-seq data of comparisons of each sample against the preceding time-point ($FDR < 0.05$). (XLSX 1508 kb)

Additional file 5: Polysome-bound vs. ribosome-free analysis. (XLSX 41 kb)

Additional file 6: Polysome-bound vs. ribosome-free ratios ($FDR < 0.05$). (XLSX 17874 kb)

Additional file 7: Polysome-bound and ribosome-free RNA-seq data of D0 vs. D15 comparison ($FDR < 0.05$). (XLSX 3849 kb)

Additional file 8: Counts and RPKM values of all mapped genes. (XLSX 30119 kb)

Abbreviations

BP: Biological process; COA: Correspondence analysis; CPM: Counts per million mapped reads; DEG: Differentially expressed genes; EB: Embryoid body; FC: Fold change; FDR: False discovery rate; GO: Gene Ontology;

hASCs: Human adipose stem cells; hESC: Human embryonic stem-cell; OPP: O-propargyl-puromycin; RPKM: Reads per kilobase per million mapped reads

Acknowledgements

The authors would like to thank the Program for Technological Development in Tools for Health-RPT-FIOCRUZ for the use of flow cytometry and microscopy facilities at Carlos Chagas Institute – Fiocruz/PR, the National Cancer Institute (Rio de Janeiro, Brazil) for the use of sequencing facility and Dr. David Elliot from Monash University (Australia) for kindly providing the cells.

Funding

Financial support from CNPq: INCT-REGENERA grant 465656/2014–5. ITP and AWR received fellowships from CAPES, and BD from CNPq.

Availability of data and materials

The datasets generated and/or analysed during the current study are available in the Short Read Archive, NCBI (<https://www.ncbi.nlm.nih.gov/sra/SRP150416>). This site serves as a landing page for the study: description of the project, metadata and raw sequencing files can be found there. Counts data and RPKM can be found in Additional file 8. One tab corresponds to the read counts of each sample, and the other, to RPKM values. Each column of each file is labeled with the sample condition, e.g., 1D4P corresponds to biological replicate 1, at day 4 and polysomal RNA fraction.

Authors' contributions

ITP and AWR performed the experiments. ITP and BD wrote the manuscript. LS performed the bioinformatic analyses and wrote the manuscript. RAH performed the data mapping. MAS coordinated the cardiac differentiation experiments. HN coordinated the bioinformatic analyses. BD conceived and coordinated the work. All authors read and approved the final manuscript.

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Competing interests

The authors declare that they have no competing interests.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Author details

¹Basic Stem-cell Biology Laboratory, Instituto Carlos Chagas - FIOCRUZ-PR, Rua Professor Algacyr Munhoz Mader, 3775, Curitiba, PR 81.350-010, Brazil. ²Bioinformatics Unit, Institut Pasteur de Montevideo, Mataojo 2020, 11400 Montevideo, Uruguay.

Received: 27 July 2018 Accepted: 20 February 2019

Published online: 15 March 2019

References

- Murry CE, Keller G. Differentiation of embryonic stem cells to clinically relevant populations: lessons from embryonic development. *Cell*. 2008; 132:661–80.
- Kattman SJ, Witty AD, Gagliardi M, Dubois NC, Niapour M, Hotta A, et al. Stage-specific optimization of activin/nodal and BMP signaling promotes cardiac differentiation of mouse and human pluripotent stem cell lines. *Cell Stem Cell*. 2011;8:228–40. <https://doi.org/10.1016/j.stem.2010.12.008>.
- Kehat I, Kenyagin-Karsenti D, Snir M, Segev H, Amit M, Gepstein A, et al. Human embryonic stem cells can differentiate into myocytes with structural and functional properties of cardiomyocytes. *J Clin Invest*. 2001;108:407–14. <https://doi.org/10.1172/JCI200112131>.
- Laflamme MA, Chen KY, Naumova AV, Muskheili V, Fugate JA, Dupras SK, et al. Cardiomyocytes derived from human embryonic stem cells in pro-survival factors enhance function of infarcted rat hearts. *Nat Biotechnol*. 2007;25:1015–24.

5. Xu C, Police S, Rao N, Carpenter MK. Characterization and enrichment of cardiomyocytes derived from human embryonic stem cells. *Circ Res*. 2002; 91:501–8.
6. BurrIDGE PW, Keller G, Gold JD, Wu JC. Production of de novo cardiomyocytes: human pluripotent stem cell differentiation and direct reprogramming. *Cell Stem Cell*. 2012;10:16–28. <https://doi.org/10.1016/j.stem.2011.12.013>.
7. Bergmann O, Zdunek S, Felker A, Salehpour M, Alkass K, Bernard S, et al. Dynamics of cell generation and turnover in the human heart. *Cell*. 2015; 161:1566–75.
8. Olson EN. Gene Regulatory Networks in the Evolution and Development of the Heart. *Science*. 2006;313:1922–7. <https://doi.org/10.1126/science.1132292>.
9. Gan L, Schwengberg S, Denecke B. Transcriptome analysis in cardiomyocyte-specific differentiation of murine embryonic stem cells reveals transcriptional regulation network. *Gene Expr Patterns*. 2014;16:8–22.
10. Ben-David U, Kopper O, Benvenisty N. Expanding the boundaries of embryonic stem cells. *Cell Stem Cell*. 2012;10:666–77. <https://doi.org/10.1016/j.stem.2012.05.003>.
11. Young RA. Control of the embryonic stem cell state. *Cell*. 2011;144:940–54. <https://doi.org/10.1016/j.cell.2011.01.032>.
12. Beqqali A, Kloots J, Ward-van Oostwaard D, Mummery C, Passier R. Genome-wide transcriptional profiling of human embryonic stem cells differentiating to cardiomyocytes. *Stem Cells*. 2006;24:1956–67. <https://doi.org/10.1634/stemcells.2006-0054>.
13. den Hartogh SC, Wolstencroft K, Mummery CL, Passier R. A comprehensive gene expression analysis at sequential stages of in vitro cardiac differentiation from isolated MESP1-expressing-mesoderm progenitors. *Sci Rep*. 2016;6. <https://doi.org/10.1038/srep19386>.
14. Li Y, Lin B, Yang L. Comparative transcriptomic analysis of multiple cardiovascular fates from embryonic stem cells predicts novel regulators in human Cardiogenesis. *Sci Rep*. 2015;5. <https://doi.org/10.1038/srep09758>.
15. Liu Q, Jiang C, Xu J, Zhao MT, Van Bortle K, Cheng X, et al. Genome-wide temporal profiling of transcriptome and open chromatin of early cardiomyocyte differentiation derived from hiPSCs and hESCs. *Circ Res*. 2017;121:376–91.
16. Tompkins JD, Jung M, Chen CY, Lin Z, Ye J, Godatha S, et al. Mapping Human Pluripotent-to-Cardiomyocyte Differentiation: Methyloomes, Transcriptomes, and Exon DNA Methylation “Memories”. *EBioMedicine*. 2016; 4:74–85. <https://doi.org/10.1016/j.ebiom.2016.01.021>.
17. Xu XQ, Soo SY, Sun W, Zweigerdt R. Global expression profile of highly enriched cardiomyocytes derived from human embryonic stem cells. *Stem Cells*. 2009;27:2163–74. <https://doi.org/10.1002/stem.166>.
18. Schwanhäusser B, Busse D, Li N, Dittmar G, Schuchhardt J, Wolf J, et al. Global quantification of mammalian gene expression control. *Nature*. 2011; 473:337–42. <https://doi.org/10.1038/nature10098>.
19. Van Der Kelen K, Beyaert R, Inzé D, De Veylder L. Translational control of eukaryotic gene expression. *Crit Rev Biochem Mol Biol*. 2009;44:143–68. <https://doi.org/10.1080/10409230902882090>.
20. King HA, Gerber AP. Translatome profiling: Methods for genome-scale analysis of mRNA translation. *Brief Funct Genomics*. 2014;15:22–31.
21. Kolle G, Shepherd JL, Gardiner B, Kassahn KS, Cloonan N, Wood DLA, et al. Deep-transcriptome and ribonome sequencing redefines the molecular networks of pluripotency and the extracellular space in human embryonic stem cells. *Genome Res*. 2014;21:2014–25.
22. Sampath P, Pritchard DK, Pabon L, Reinecke H, Schwartz SM, Morris DR, et al. A hierarchical network controls protein translation during murine embryonic stem cell self-renewal and differentiation. *Cell Stem Cell*. 2008;2: 448–60.
23. Faye MD, Graber TE, Holcik M. Assessment of Selective mRNA Translation in Mammalian Cells by Polysome Profiling. *J Vis Exp*. 2014:1–8. <https://doi.org/10.3791/52295>.
24. Panda AC, Martindale JL, Gorospe M. Polysome fractionation to analyze mRNA distribution profiles. *Bio Protoc*. 2017;7:e2126.
25. Spangenberg L, Shigunov P, Abud APR, Cofré AR, Stimamiglio MA, Kuligovski C, et al. Polysome profiling shows extensive posttranscriptional regulation during human adipocyte stem cell differentiation into adipocytes. *Stem Cell Res*. 2013;11:902–12.
26. Zych J, Spangenberg L, Stimamiglio MA, Abud APR, Shigunov P, Marchini FK, et al. Polysome profiling shows the identity of human adipose-derived stromal/stem cells in detail and clearly distinguishes them from dermal fibroblasts. *Stem Cells Dev*. 2014;23:2791–802. <https://doi.org/10.1089/scd.2013.0496>.
27. Elliott DA, Braam SR, Koutsis K, Ng ES, Jenny R, Lagerqvist EL, et al. NKK2-5eGFP/w hESCs for isolation of human cardiac progenitors and cardiomyocytes. *Nat Methods*. 2011;8:1037–40. <https://doi.org/10.1038/nmeth.1740>.
28. Yang L, Soonpaa MH, Adler ED, Roepke TK, Kattman SJ, Kennedy M, et al. Human cardiovascular progenitor cells develop from a KDR+ embryonic-stem-cell-derived population. *Nature*. 2008;453:524–8. <https://doi.org/10.1038/nature06894>.
29. Evseenko D, Zhu Y, Schenke-layland K, Kuo J, Latour B, Ge S, et al. Mapping the first stages of mesoderm commitment during differentiation of human embryonic stem cells. *PNAS*. 2010;107:13742–7.
30. Bruneau BG. Signaling and transcriptional networks in heart development and regeneration. *Cold Spring Harb Lab Perspect Biol*. 2013;5:a008292.
31. DeGregori J. The genetics of the E2F family of transcription factors: shared functions and unique roles. *Biochim Biophys Acta - Rev Cancer*. 2002;1602: 131–50.
32. Luo W, Li G, Yi Z, Nie Q, Zhang X. E2F1-miR-20a-5p/20b-5p auto-regulatory feedback loop involved in myoblast proliferation and differentiation. *Sci Rep*. 2016;6:1–13. <https://doi.org/10.1038/srep27904>.
33. Thiele BJ, Doller A, Kähne T, Pregla R, Hetzer R, Regitz-Zagrosek V. RNA-binding proteins heterogeneous nuclear ribonucleoprotein A1, E1, and K are involved in post-transcriptional control of collagen I and III synthesis. *Circ Res*. 2004;95:1058–66.
34. Svobodová K, Horák P, Stratil A, Bartschlagler H, Van Poucke M, Chalupová P, et al. Porcine EEF1A1 and EEF1A2 genes: genomic structure, polymorphism, mapping and expression. *Mol Biol Rep*. 2015;42:1257–64.
35. Tohyama S, Hattori F, Sano M, Hishiki T, Nagahata Y, Matsuura T, et al. Distinct metabolic flow enables large-scale purification of mouse and human pluripotent stem cell-derived cardiomyocytes. *Cell Stem Cell*. 2013; 12:127–37. <https://doi.org/10.1016/j.stem.2012.09.013>.
36. Boyer LA, Tong IL, Cole MF, Johnstone SE, Levine SS, Zucker JP, et al. Core transcriptional regulatory circuitry in human embryonic stem cells. *Cell*. 2005;122:947–56.
37. Loh KMM, Chen A, Koh PWW, Deng TZZ, Sinha R, Tsai JMM, et al. Mapping the pairwise choices leading from pluripotency to human bone, heart, and other mesoderm cell types. *Cell*. 2016;166:451–68. <https://doi.org/10.1016/j.cell.2016.06.011>.
38. Tebaldi T, Re A, Viero G, Pegoretti I, Passerini A, Blanzieri E, et al. Widespread uncoupling between transcriptome and translational variations after a stimulus in mammalian cells. *BMC Genomics*. 2012;13:220. <https://doi.org/10.1186/1471-2164-13-220>.
39. Sonenberg N, Hinnebusch AG. New modes of translational control in development, behavior, and disease. *Mol Cell*. 2007;28:721–9.
40. Tian Q, Stepanians SB, Mao M, Weng L, Feetham MC, Doyle MJ, et al. Integrated Genomic and Proteomic Analyses of Gene Expression in Mammalian Cells. *Mol Cell Proteomics*. 2004;3:960–9.
41. Jackson RJ, Hellen CUT, Pestova TV. The mechanism of eukaryotic translation initiation and principles of its regulation. *Nat Rev Mol Cell Biol*. 2010;11:113–27. <https://doi.org/10.1038/nrm2838>.
42. Stefl R, Skrisovska L, Allain FH-T. RNA sequence- and shape-dependent recognition by proteins in the ribonucleoprotein particle. *EMBO Rep*. 2005;6: 33–8. <https://doi.org/10.1038/sj.embor.7400325>.
43. Raj A, Peskin CS, Tranchina D, Vargas DY, Tyagi S. Stochastic mRNA synthesis in mammalian cells. *PLoS Biol*. 2006;4:1707–19.
44. McManus CJ, May GE, Spealman P, Shteyman A. Ribosome profiling reveals post-transcriptional buffering of divergent gene expression in yeast. *Genome Res*. 2013;24:422–30.
45. Fujii K, Shi Z, Zhulyn O, Denans N, Barna M. Pervasive translational regulation of the cell signalling circuitry underlies mammalian development. *Nat Commun*. 2017;8:1–13. <https://doi.org/10.1038/ncomms14443>.
46. Blair JD, Hockemeyer D, Doudna JA, Bateup HS, Floor SN. Widespread translational remodeling during human neuronal differentiation. *Cell Rep*. 2017;21:2005–16. <https://doi.org/10.1016/j.celrep.2017.10.095>.
47. Chorghade S, Seimetz J, Emmons R, Yang J, Bresson SM, De Lisio M, et al. Poly(a) tail length regulates PABPC1 expression to tune translation in the heart. *Elife*. 2017;6:e24139. <https://doi.org/10.7554/eLife.24139>.
48. Lewis SE, Kelly FJ, Goldspink DF. Pre- and post-natal growth and protein turnover in smooth muscle, heart and slow- and fast-twitch skeletal muscles of the rat. *Biochem J*. 1984;217:517–26.

49. Meyuhas O. Synthesis of the translational apparatus is regulated at the translational level. *Eur J Biochem.* 2000;267:6321–30.
50. Ingolia NT, Lareau LF, Weissman JS. Ribosome profiling of mouse embryonic stem cells reveals the complexity and dynamics of mammalian proteomes. *Cell.* 2011;147:789–802. <https://doi.org/10.1016/j.cell.2011.10.002>.
51. Hershey JWB, Sonenberg N, Mathews MB. Principles of translational control: an overview. *Cold Spring Harb Perspect Biol.* 2012;4:a011528. <https://doi.org/10.1101/cshperspect.a011528>.
52. Marcon BH, Holetz FB, Eastman G, Origa-Alves AC, Amorós MA, de Aguiar AM, et al. Downregulation of the protein synthesis machinery is a major regulatory event during early adipogenic differentiation of human adipose-derived stromal cells. *Stem Cell Res.* 2017;25:191–201. <https://doi.org/10.1016/j.scr.2017.10.027>.
53. Collier AE, Wek RC, Spandau DF. Human keratinocyte differentiation requires translational control by the eIF2 α kinase GCN2. *J Invest Dermatol.* 2017. <https://doi.org/10.1016/j.jid.2017.04.029>.
54. Gerlitz G, Jagus R, Elroy-Stein O. Phosphorylation of initiation factor-2 α is required for activation of internal translation initiation during cell differentiation. *Eur J Biochem.* 2002;269:2810–9.
55. Krichevsky AM, Metzger E, Rosen H. Translational control of specific genes during differentiation of HL-60 cells. *J Biol Chem.* 1999;274:14295–305.
56. Miyata M, Raven JF, Baltzis D, Koromilas AE, Sabe H. IRES-mediated translational control of AMAP1 expression during differentiation of monocyte U937 cells. *Cell Cycle.* 2008;7:3273–81.
57. You KT, Park J, Kim VN. Role of the small subunit processome in the maintenance of pluripotent stem cells. *Genes Dev.* 2012;29:2004–9.
58. Blanco S, Bandiera R, Popis M, Hussain S, Lombard P, Aleksic J, et al. Stem cell function and stress response are controlled by protein synthesis. *Nature.* 2016;534:335–40. <https://doi.org/10.1038/nature18282>.
59. Ivanov P, Emara MM, Villen J, Gygi SP, Anderson P. Angiogenin-induced tRNA fragments inhibit translation initiation. *Mol Cell.* 2011;43:613–23. <https://doi.org/10.1016/j.molcel.2011.06.022>.
60. Tuorto F, Liebers R, Musch T, Schaefer M, Hofmann S, Kellner S, et al. RNA cytosine methylation by Dnmt2 and NSun2 promotes tRNA stability and protein synthesis. *Nat Struct Mol Biol.* 2012;19:900–5. <https://doi.org/10.1038/nsmb.2357>.
61. Rodgers JT, King KY, Brett JO, Cromie MJ, Charville GW, Maguire KK, et al. mTORC1 controls the adaptive transition of quiescent stem cells from G0to G1Alert. *Nature.* 2014;510:393–6. <https://doi.org/10.1038/nature13255>.
62. Chung S, Dzeja PP, Faustino RS, Perez-Terzic C, Behfar A, Terzic A. Mitochondrial oxidative metabolism is required for the cardiac differentiation of stem cells. *Nat Clin Pract Cardiovasc Med.* 2007;4:S60–7. <https://doi.org/10.1038/ncpcardio0766>.
63. Lopaschuk GD, Jaswal JS. Energy metabolic phenotype of the cardiomyocyte during development, differentiation, and postnatal maturation. *J Cardiovasc Pharmacol.* 2010;56:130–40. <https://doi.org/10.1097/FJC.0b013e3181e74a14>.
64. Adeli K. Translational control mechanisms in metabolic regulation: critical role of RNA binding proteins, microRNAs, and cytoplasmic RNA granules. *AJP Endocrinol Metab.* 2011;301:E1051–64. <https://doi.org/10.1152/ajpendo.00399.2011>.
65. Arif W, Datar G, Kalsotra A. Intersections of post-transcriptional gene regulatory mechanisms with intermediary metabolism. *Biochim Biophys Acta - Gene Regul Mech.* 1860;2017:349–62. <https://doi.org/10.1016/j.bbagr.2017.01.004>.
66. Liao Y, Castello A, Fischer B, Leicht S, Föehr S, Frese CK, et al. The cardiomyocyte RNA-binding proteome: links to intermediary metabolism and heart disease. *Cell Rep.* 2016;16:1456–69.
67. Simsek D, Tiu GC, Flynn RA, Byeon GW, Leppke K, Xu AF, et al. The Mammalian Ribo-interactome Reveals Ribosome Functional Diversity and Heterogeneity. *Cell.* 2017;169:1051–65.e18. <https://doi.org/10.1016/j.cell.2017.05.022>.
68. Liao Y, Smyth GK, Shi W. The subread aligner: fast, accurate and scalable read mapping by seed-and-vote. *Nucleic Acids Res.* 2013;41:e108.
69. Robinson MD, McCarthy DJ, Smyth GK. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics.* 2010;26:139–40.

Ready to submit your research? Choose BMC and benefit from:

- fast, convenient online submission
- thorough peer review by experienced researchers in your field
- rapid publication on acceptance
- support for research data, including large and complex data types
- gold Open Access which fosters wider collaboration and increased citations
- maximum visibility for your research: over 100M website views per year

At BMC, research is always in progress.

Learn more biomedcentral.com/submissions



CAPÍTULO III

Prefácio

Neste capítulo, três genes com expressão diferencial na cardiomiogênese de CTE – *RXRG*, *SHISA3* e *CSDC2* – foram escolhidos para a construção de ferramentas que visaram sua avaliação funcional. Um sistema de super expressão induzida, controlada por doxiciclina, foi introduzido em células-tronco pluripotentes, as quais foram submetidas à diferenciação cardíaca. Metodologia e resultados preliminares desta abordagem foram realizados durante Doutorado Sanduíche, na Universidade de Minnesota (EUA), e estão descritos neste capítulo.

INTRODUÇÃO

Tendo em vista a necessidade de desvendar novos genes atuantes no comprometimento à linhagem cardíaca e auxiliar na melhor compreensão desse processo biológico, alguns genes foram escolhidos com base nos dados de RNA-seq (Capítulos I e II) e de acordo com seu padrão de expressão, função biológica e descrição na literatura para serem avaliados quanto à sua participação na diferenciação cardiomiogênica. O estabelecimento de progenitores cardíacos e cardiomiócitos é de grande interesse, tanto para a pesquisa básica quanto para a medicina translacional, e por isso embasou essa escolha.

O gene *CSDC2* codifica para uma proteína de ligação ao RNA e se mostrou expresso em cardiomiócitos (D15) (Capítulo II). Proteínas da família das cold shock estão envolvidas na regulação de RNAs e na regulação pós-transcricional da expressão gênica (MIHAILOVICH *et al.*, 2010). A *CSDC2*, também conhecida como PIPPin, foi descrita apenas em tecido cerebral de murinos (NASTASI *et al.*, 1999).

A sinalização por ácido retinóico é de grande importância durante o desenvolvimento e, inclusive, apareceu como regulada pós-transcricionalmente nos dados deste trabalho (Capítulo II). Entretanto, o papel de um de seus receptores nucleares, *RXRG*, ainda não é bem esclarecido. Os receptores de ácido retinóico (RARs e RXRs) atuam como heterodímeros, ligando-se a elementos responsivos no DNA e desencadeando a transcrição de genes (KREZEL *et al.*, 1996; MANGELSDORF *et al.*, 1992). A expressão do gene de *RXRG* é sutilmente

aumentada em D4 e possui um pico nas células diferenciadas (D15) (Capítulo II). Diante disso, *RXRG* foi um dos genes escolhidos para a avaliação funcional durante a diferenciação cardíaca.

Outro gene selecionado foi o *SHISA3*, o qual parece estar envolvido na regulação negativa da sinalização de Wnt e FGF, importantes vias do desenvolvimento (FURUSHIMA *et al.*, 2007; YAMAMOTO *et al.*, 2005). Há indícios que Shisa3 promove a degradação de β -catenina e, assim, inibe a via canônica de Wnt em células tumorais (CHEN *et al.*, 2014). A função de SHISA3 foi descrita em outros organismos, mas ainda não em humanos, e sua expressão apresenta variação ao longo da diferenciação, com picos nos dias D4 e D15 (Capítulo II).

A estratégia para avaliação funcional dos genes escolhidos foi o estabelecimento de linhagens com super expressão induzida e submissão dessas células à diferenciação cardiomiogênica. O sistema de expressão induzida escolhido foi um sistema integrativo que utiliza o vetor AAVS1-TRE-FLAG-rtTA, e possibilita a expressão controlada através da indução com doxíciclina.

4 METODOLOGIA

4.1 Cultura de células-tronco pluripotentes humanas (CTPs)

A linhagem de células embrionárias humanas H1 (WiCell) e as células induzidas pluripotentes humanas PLZ (IPRN13.13, derivadas de fibroblastos de derme) foram cultivadas em meio mTeSR1 ou TeSR-E8 (Stem Cell Technologies) sobre uma camada de Geltrex (ThermoFischer, cat A1413302) diluído 1:100. As células eram cultivadas até atingirem 80% de confluência, com troca de meio diária, e a passagem feita através da dissociação enzimática utilizando Accumax (Millipore, cat 12932715).

4.2 Protocolo de diferenciação cardiomiogênica em monocamada

As células pluripotentes H1 e PLZ foram submetidas à diferenciação cardíaca de acordo com o protocolo seguinte (Figura 1.4B) (LIAN *et al.*, 2012). Inicialmente (dia -3), 500.000 células foram plaqueadas por poço sobre uma camada de Geltrex em

placa aderente de 12 poços e mantidas com 2 ml de meio mTeSR1. Esse número de células é suficiente para que o poço fique com 90% a 100% de confluência no dia seguinte ao plaqueamento. A confluência é essencial para a eficiência do protocolo. Então, as células foram mantidas por 3 dias, com troca de meio diária. No dia 0, a diferenciação é iniciada através da troca do meio por RPMI suplementado com B-27 sem insulina e 12 μ M de CHIR99021 (Tocris, diluído em DMSO), com incubação por 24h. No dia 1, houve a troca de meio por RPMI/B-27 sem insulina. No dia 3, 1 ml do meio condicionado das células foi misturado com 1 ml de meio fresco RPMI/B-27 sem insulina e suplementado com 10 μ M de XAV939 (Tocris, diluído em DMSO). No dia 5, o meio foi trocado por RPMI/B-27 sem insulina. A partir do dia 7, o meio utilizado foi o RPMI/B-27, com troca a cada 3 dias. Células pulsantes aparecem no cultivo a partir do dia 10 e foram analisadas nos dias conforme indicado.

4.3 Citometria analítica

As células foram dissociadas utilizando tripsina (0,25%) por 20 min, coletadas com DMEM suplementado com 10% de SFB (soro fetal bovino) e centrifugadas a 1200 rpm por 3 min. O pellet de células foi ressuspenso com paraformaldeído (PFA) a 4% para fixação por 15 min e, então, lavadas com PBS 1X. As células foram incubadas com metanol 90% gelado por, pelo menos, 16 h. Para a marcação, as células foram incubadas com PBS/BSA 0,5% + Triton-X 0,1% + anticorpos primários contra cTnT (CT3 – DSHB AB-528495 mouse, diluição 1:50), Nkx2-5 (Santa Cruz, sc-8697 goat, diluição 1:100), Nanog (Sigma, AF1997 goat) ou Brachyury (R&D, AF2085 goat) por 1 h. Depois de lavadas, foram incubadas com PBS/BSA 0,5% + Triton-X 0,1% + anticorpos secundários Alexa-Fluor 555 donkey anti-mouse PE-conjugado (1:1000) e Alexa-Fluor 488 donkey anti-goat FITC-conjugado (1:1000) por 1 h. Amostras controle foram incubadas apenas com os anticorpos secundários. As amostras foram lidas no equipamento FACS Aria (BD Biosciences) e as análises foram feitas utilizando o software FlowJo.

4.4 Imunofluorescência

Para marcação das células por imunofluorescência, o meio foi retirado e as células fixadas com PFA 4% por 20 min. Após lavagem com PBS, foram permeabilizadas com PBS/Triton-X 0.3% por 30 min, e bloqueadas com PBS/BSA 3%

por 1 h. A marcação foi feita com os anticorpos primários: Oct4 (Santa Cruz, sc-5279), Nanog (Sigma, AF1997 goat), Sox2 (Santa cruz, sc-17320), cTnT (CT3 - DSHB AB-528495 mouse), Nkx2-5 (Santa Cruz, sc-8697 goat), α -SMA (Abcam, ab-7817 mouse), RXRG (Santa Cruz, sc-365252 mouse), PIPPIN/CSDC2 (Santa Cruz, sc-376693 mouse), α -sarcomeric actinin (Sigma, A7811 mouse) ou Cx43 (Sigma, C6219 rabbit), conforme indicado; todos diluídos 1:100 em PBS/BSA 3% e incubados por 16 h a 4 °C. Depois de lavadas 2X com PBS 1X, as células foram incubadas com os anticorpos secundários correspondentes Alexa-Fluor 555 anti-mouse, anti-goat ou anti-rabbit (1:1000) por 1h e depois com DAPI (1 μ g/ μ L) por 10 min. As imagens foram adquiridas no centro de microscopia do Leililhei Heart Institute, University of Minnesota (EUA).

4.5 Extração de RNA e síntese de cDNA

O RNA total das células em diferenciação foi extraído com o kit Quick-RNA™ MiniPrep (Zymo Research), conforme especificações do fabricante. Ao final do processo, o RNA obtido foi eluído em água livre de RNase e quantificado por espectrofotometria (NanoDrop ND-100). A síntese de cDNA foi feita utilizando o kit Verso cDNA Synthesis (Thermo Fischer, AB1453A) e 1 μ g de RNA por reação.

4.6 Quantificação da expressão gênica por PCR quantitativo (qPCR)

Reações de qPCR (do inglês, *quantitative polymerase chain reaction*) foram feitas utilizando o kit Premix Ex Taq (Takara, Probe qPCR) e Taqman probes (Thermo Fischer), listados na Tabela 4.1. Para cada reação em triplicata técnica, foram preparadas soluções contendo: 6,6 μ l de Premix Ex Taq 2X, 0,15 μ l de ROX dye II, 0,75 μ l de probe Taqman e 7,5 μ l de cDNA (concentração final aprox. 37,5 ng cDNA). Também foram realizadas reações de qPCR utilizando o kit SYBR Premix Ex Taq (Takara) e oligos listados na Tabela 4.2. Para cada reação em triplicata técnica, foram preparadas soluções contendo: 7,5 μ l de SYBR Premix 2X, 0,15 μ l de ROX dye II, 200 nM de cada oligo *forward* e reverso, e 5 μ l de cDNA (concentração final aprox. 25 ng cDNA). O equipamento QuantStudio 6 Flex Real-Time PCR System foi usado na performance das reações, com placas de 384 poços. O gene endógeno GAPDH foi usado como normalizador em todas as amostras.

Tabela 4.1 Lista de genes e suas respectivas Taqman probes utilizadas.

Gene alvo	Código probe Taqman
CSDC2	Hs00411093_m1
EOMES	Hs00172872_m1
GAPDH	Hs99999905_m1
GATA4	Hs00171403_m1
HCN4	Hs00975492_m1
HEY2	Hs01012057_m1
IRX4	Hs00212560_m1
ISL1	Hs00158126_m1
KCNA5	Hs00969279_s1
LINC00890	Hs04273890_m1
MESP1	Hs00251489_m1
MYH6	Hs01101425_m1
MYH7	Hs01110593_g1
MYL2	Hs00166405_m1
MYL7	Hs01085598_g1
NKX2-5	Hs00231763_m1
RXRG	Hs00199455_m1
SHISA3	Hs01380806_m1
T	Hs00610080_m1
TBX5	Hs01052563_m1
TNNI1	Hs00913333_m1
TNNI3	Hs00165957_m1
TNNT2	Hs00165960_m1

Tabela 4.2 Lista de genes e seus respectivos oligos *forward* (F) e reverso (R) utilizados.

Gene alvo		Oligo 5' - 3'	Tamanho do amplicon	Referência
CSDC2	F	GTTCAAGGGCGTCTGTAAGC	112 pb	Primer Blast
	R	GTACTCCCCCTCGATGTCAG		
GAPDH	F	GGCGATGCTGGCGCTGAGTAC	149 pb	(SPANGENBERG <i>et al.</i> , 2013)
	R	TGGTTCACACCCATGACGA		
KCNJ2	F	GTACTGGTCAAGATGGGTCAAG	120 pb	(ZENG <i>et al.</i> , 2018)
	R	GGTCACATGGAGACATGGTTAG		
LCAD	F	TTCCACAGGAAAAGGCTGTT	144 pb	(GARCÍA-RÚA <i>et al.</i> , 2012)
	R	CTGCTAATTTATGTTGCACTG		
LINC00890	F	GGGCCAGGAGAGGAAAATCAAT	145 pb	Primer Blast
	R	CACTGTAGGCTGGAGGGAAA		
NRF2R2	F	CAACCAGCCGACGAGATT	80 pb	(ZHU <i>et al.</i> , 2016)
	R	ATTGCTCTATGACTGAGGAGGAGAC		
SERCA2	F	TGTATGGCAGGAAAAGAAATG	159 pb	(MAYUZUMI <i>et al.</i> , 2005)
	R	TTGTCACCAACAGCAATTTCT		

4.7 Construção de linhagens celulares com super expressão induzida

4.7.1 Clonagem dos genes candidatos no vetor AAVS1-TRE-FLAG-rtTA

O vetor AAVS1-TRE-FLAG-rtTA é um vetor de integração sítio-específica, que possui braços de homologia com o locus AAVS1 do genoma humano, localizado no cromossomo 19 (LOMBARDO *et al.*, 2011; OCEGUERA-YANEZ *et al.*, 2016). Mediante a clivagem sítio-específica decorrente da ação das nucleases *zinc fingers* recombinantes (ZFN), a integração do transgene é direcionada para o sítio de interesse (HOCKEMEYER *et al.*, 2009), no caso o locus AAVS1. O vetor AAVS1-TRE-FLAG-rtTA foi construído no laboratório do Dr Michael Kyba (Universidade de Minnesota – EUA) e cedido para este trabalho. Compreende um sistema de expressão induzida de transgene, através da ativação por doxiciclina, e por isso possui um sítio de elemento responsivo a tetraciclina (TER, do inglês *tetracyclin responsive element*) localizado no promotor do transgene, e expressão constitutiva da proteína auxiliar rtTA (do inglês, *reverse tetracycline-controlled transactivator*). O vetor também introduz o gene de resistência a puomicina, para facilitar a seleção de células efetivamente transfectadas. Além disso, uma etiqueta FLAG é introduzida na porção 5' do produto do transgene (Figura 4.1).

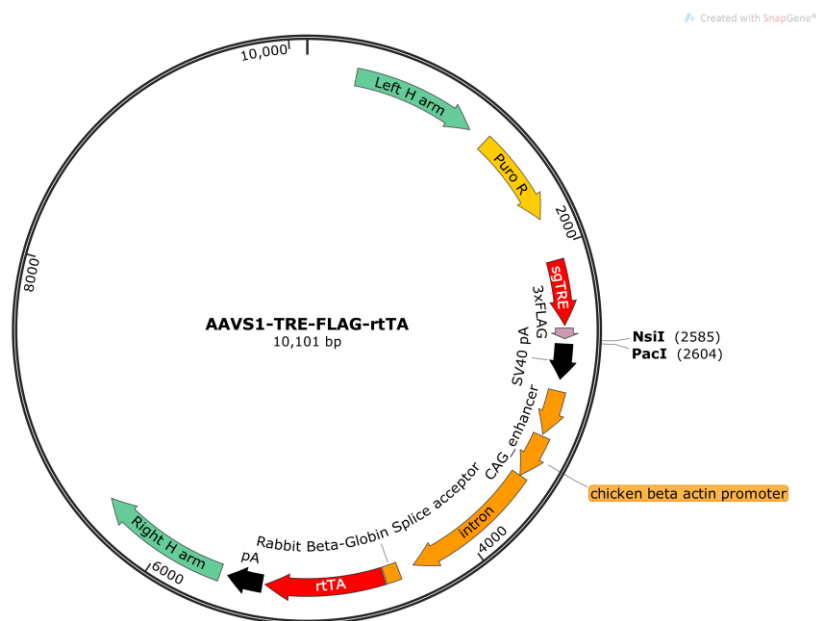


Figura 4.1 Mapa do vetor AAVS1-TRE-FLAG-rtTA. Esse vetor possui braços de homologia com o locus AAVS1 do genoma humano (*Left H arm*, *Right H arm*, em verde), sítio TRE (vermelho) no promotor do transgene que será inserido entre os sítios de restrição de NsiI e PacI, e expressão constitutiva de rtTA (vermelho) e gene de resistência a puomicina (laranja). Mapa criado com Snapgene.

Os sítios de restrição Nsil e PacI foram usados para a clonagem das sequências codificantes (CDS) dos genes *CSDC2*, *RXRG* e *SHISA3*. Oligos foram desenhados de maneira que incluíssem 16-18 nucleotídeos de homologia com o vetor e restaurassem os sítios de restrição, além de excluir o primeiro códon ATG e manter o último códon de parada da CDS (ANEXO I – Tabela 1). Estoques de bactéria com as sequências de cDNA clonadas e verificadas por sequenciamento foram adquiridos da Open Biosystems (Dharmacon), sob os códigos: *CSDC2* cDNA (GE Dharmacon MHS6278-202802532), *RXRG* cDNA (GE Dharmacon MHS6278-202831995) e *SHISA3* cDNA (GE Dharmacon MHS6278-211691041). A amplificação das CDSs foi feita por PCR (do inglês, *polymerase chain reaction*) usando o kit da enzima LA Taq (Takara) com 500 ng de DNA plasmidial e 0.8 μ M de cada oligo (*forward* e reverso). A ciclagem foi feita com 95 °C por 5 min de desnaturação inicial; 95 °C por 30 seg, 50 °C por 30 seg, 72 °C por 2 min (7X) para amplificação; e 72 °C por 10 min para extensão final.

Os produtos das PCRs foram verificados quanto ao tamanho esperado em pares de base (pb) em gel de agarose 1% e purificados do mesmo gel, utilizando o kit Nucleo Spin Gel & PCR Clean-up (Clontech). O vetor AAVS1-TRE-FLAG-rtTA foi linearizado através da reação de restrição utilizando as enzimas Nsil e PacI e também purificado através do gel de agarose. A inserção do fragmento amplificado por PCR contendo a CDS dos genes candidatos, acoplada aos nucleotídeos de homologia, no vetor linear foi feita utilizando o kit In-Fusion HD Cloning (Clontech).

4.7.2 Transformação de bactérias e extração de DNA plasmidial

Bactérias *E. coli* da cepa HST08 termo competentes (Stellar, Takara) foram transformadas utilizando o protocolo seguinte. Uma alíquota de 20 μ l de bactérias competentes foi misturada com 3 μ l da reação de In-Fusion (vetor+fragmento) e incubada no gelo por 30 min. Então, foi feito o choque térmico a 42 °C por 30 seg, seguido de nova incubação no gelo por 2 min. As bactérias foram recuperadas em meio SOC (Takara) a 37 °C por 1 h, sob agitação. Depois foram plaqueadas em placas de LB/ágar com antibiótico específico e incubadas em estufa a 37 °C por 16 h. Colônias isoladas foram crescidas em meio LB líquido, sob agitação, a 37 °C por 16 h, e submetidas a extração de DNA plasmidial utilizando o kit QIAprep Spin Miniprep (Qiagen). O DNA plasmidial foi submetido à confirmação por reação de restrição e sequenciamento de Sanger.

4.7.3 Clonagem do gene *CSDC2* com deleção no vetor AAVS1-TRE-FLAG-rtTA

Com a intenção de estudar domínios específicos da proteína *CSDC2*, planejou-se a construção de uma linhagem celular com super expressão induzida do gene *CSDC2*, com a deleção de dois motivos de ligação ao RNA, no domínio CSD (Figura 4.2) (NCBI NP_055275.1). Foram planejados oligos adjacentes a região a ser deletada, com nucleotídeos adicionais complementares a outra região adjacente (ANEXO I - Tabela 2). Eles foram usados em conjunto com os oligos mencionados anteriormente, para a amplificação da região CDS completa. Dessa maneira, duas PCR foram feitas com o intuito de amplificar as duas porções da CDS adjacentes a região a ser deletada: (1) *forward* AAVS1 + reverso deleção e (2) *forward* deleção + reverso AAVS1 (Figura 4.2). Os produtos das PCR foram purificados em gel de agarose e misturados para formarem duplas fitas híbridas e, então, servirem de molde para a PCR seguinte, utilizando os oligos AAVS1 e amplificar a sequência toda (agora, com a deleção). O produto final foi o fragmento da região CDS de *CSDC2* com a deleção planejada, incluindo as sequências de homologia com o vetor AAVS1-TRE-FLAG-rtTA. A inserção no vetor foi feita utilizando o kit In-Fusion HD Cloning (Clontech), conforme descrito anteriormente. Após a clonagem em bactérias, o DNA plasmidial purificado foi avaliado por sequenciamento de Sanger, confirmando a deleção.

A NCBI Reference Sequence: NP_055275.1
cold shock domain-containing protein C2 [Homo sapiens]

1 mtsestspv vpplhspksp vwptf **pfhre gsrwerggv pprdlpslp tkrtrtysat**
61 **arasagpvfk gvckqfsrsg ghgfitpeng sedifvhsd iegeyvpveg devtykmcpi**
121 **ppknqkfqav evvltqlaph tphetwsgqv vgs**

Legenda:

domínio PIP-1 **domínio CSD** **domínio PIP-2**
Motivos de ligação ao RNA **deleção**

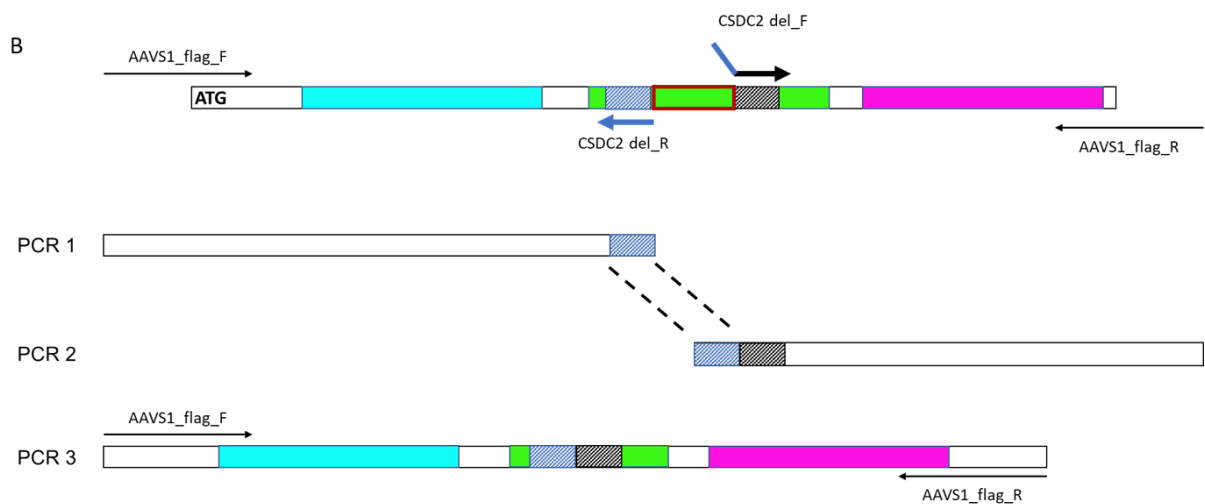


Figura 4.2 Estratégia de deleção dos motivos de ligação ao RNA da proteína codificada pelo gene *CSDC2*. (A) Sequência da proteína *CSDC2* com os domínios PIP-1, CSD e PIP-2, os motivos de ligação ao RNA e a porção a ser deletada estão indicados como na legenda. (B) Estratégia de PCR para amplificação da sequência de *CSDC2* com deleção para posterior clonagem no vetor AAVS1-TRE-FLAG-rtTA. Fonte: o autor.

4.7.4 Transfecção dos vetores AAVS1 em células-tronco pluripotentes humanas

Células da linhagem de CTE humanas H1 foram transfectadas com os vetores AAVS1 construídos para a integração do sistema de expressão induzida. As células foram plaqueadas em baixa confluência (20-30%) em placas de 6 poços, sobre uma camada de Geltrex, com meio de cultivo TEsR E8 (Stem Cell Technologies). No dia seguinte, os reagentes de transfecção foram preparados: 1 ml de meio IMDM (Iscove's Modified Dulbecco's Medium, Gibco), 200 ng plasmídeo ZFN(R) (Addgene, plasmid #60915), 200 ng plasmídeo ZFN(L) (Addgene, plasmid #60916), 600 ng plasmídeo AAVS1, 5 µl reagente vermelho e 5 µl reagente azul (kit de transfecção GenEln,

Ambio). O meio de cultivo das células foi trocado por meio fresco TEsR E8 e a solução preparada foi adicionada aos poços, seguindo-se incubação por 8 h, até uma nova troca por meio fresco. Como controle positivo de transfecção, o vetor AAV-CAGGS-EGFP (Addgene, plasmid #22212) foi usado em uma das condições, conferindo a expressão de GFP (proteína fluorescente verde) e, conseqüentemente, de fluorescência verde nas células transfectadas após 24-48 h.

4.7.5 Seleção de clones com super expressão induzida

Após cerca de 48 h da transfecção e/ou quando as colônias de CTP estivessem estáveis, iniciou-se a seleção por puromicina. Uma concentração final de 0,2 µg/ml foi usada no meio de cultivo até que as células alcançassem confluência (aprox. 7 dias). Então, fez-se a passagem das células para nova placa de 60 mm, também sobre camada de Geltrex, em baixa confluência (~10%), com a intenção de cultivar células isoladas. A seleção com puromicina continuou até que colônias estáveis se formassem (aprox. 7 dias). Sob um microscópio invertido, dentro do fluxo laminar, as colônias isoladas foram pescadas e transferidas para uma nova placa de 24 poços. Seis clones (colônias) foram pescados de cada condição. Os clones foram expandidos até que tivessem células suficientes para armazenar e fazer os testes de expressão.

Os clones foram, então, submetidos (+DOX) ou não (-DOX) à indução com 500 ng/ml de doxiciclina por 24h. O RNA total foi extraído e a expressão dos genes em questão avaliada por qPCR, conforme descrito anteriormente. Um clone de cada condição foi escolhido para dar continuidade ao trabalho, de acordo com o nível de expressão após a indução e a expressão basal (vazamento) do sistema.

5 RESULTADOS

5.1 Diferenciação cardiomiogênica de CTPs

O modelo de diferenciação cardiomiogênica usado nessa etapa do trabalho foi o de diferenciação em monocamada (LIAN *et al.*, 2012). Para que ele pudesse funcionar como uma ferramenta na análise funcional de genes, era preciso que estivesse bem estabelecido nas linhagens celulares a serem utilizadas. Por isso,

inicialmente fez-se a caracterização desse modelo de diferenciação cardíaca nas células embrionárias H1. Alguns marcadores foram seguidos ao longo da diferenciação, utilizando marcação com anticorpos e análise por citometria de fluxo: Nanog para pluripotência, mostrou redução durante os dias iniciais do protocolo; Brachyury/T para comprometimento da mesoderme, o qual apresentou um pico de expressão no dia 1 (D1), conforme esperado (LIAN *et al.*, 2012); e Nkx2-5, que teve sua expressão aumentada conforme o comprometimento cardíaco entre D7 e D15 (Figura 5.1).

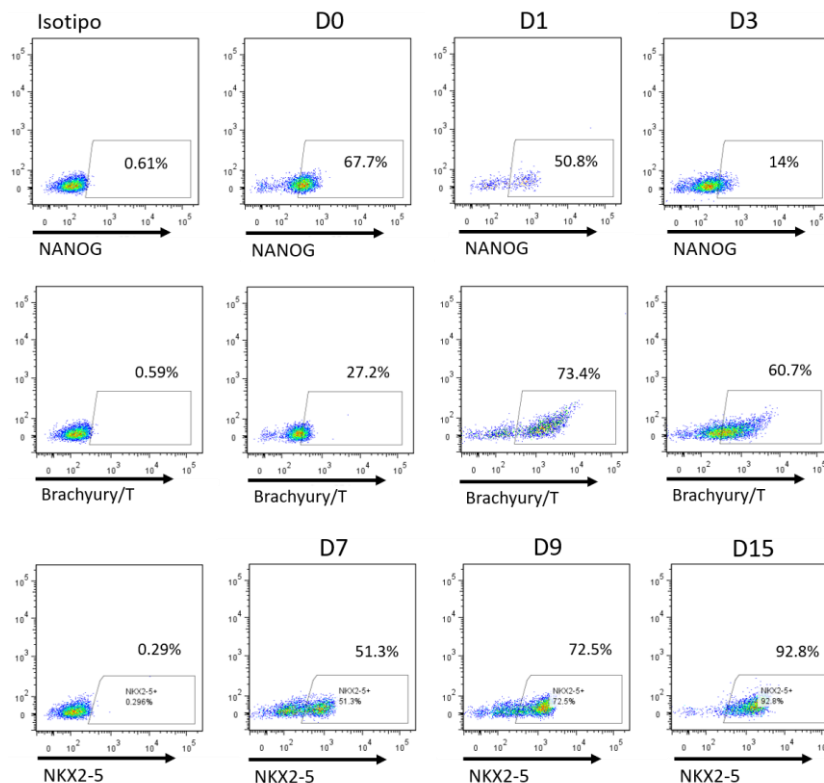


Figura 5.1 Expressão de marcadores ao longo da diferenciação cardiomiogênica em monocamada. Análise por citometria de fluxo usando os marcadores Nanog (pluripotência), Brachyury/T (mesoderme) e Nkx2-5 (progenitores cardíacos). Valores apresentados em porcentagem de células positivas para o respectivo marcador. O controle isotipo foi usado para estabelecer a população de células negativas. Dot plots representativos de um experimento ilustrativo.

A expressão de marcadores cardíacos também foi acompanhada e verificada ao final da diferenciação (Figura 5.2). Citometria de fluxo e imunofluorescência foram usadas para acompanhar a expressão de cTnT durante o protocolo de diferenciação, nos dias indicados na Figura 5.2, até o dia 30 (D30). No D30 também foi verificada a expressão de outros marcadores, como a Cx43, a-actinina sarcomérica e a-SMA. Esses dados ilustram o sucesso na diferenciação cardíaca usando o protocolo de monocamada.

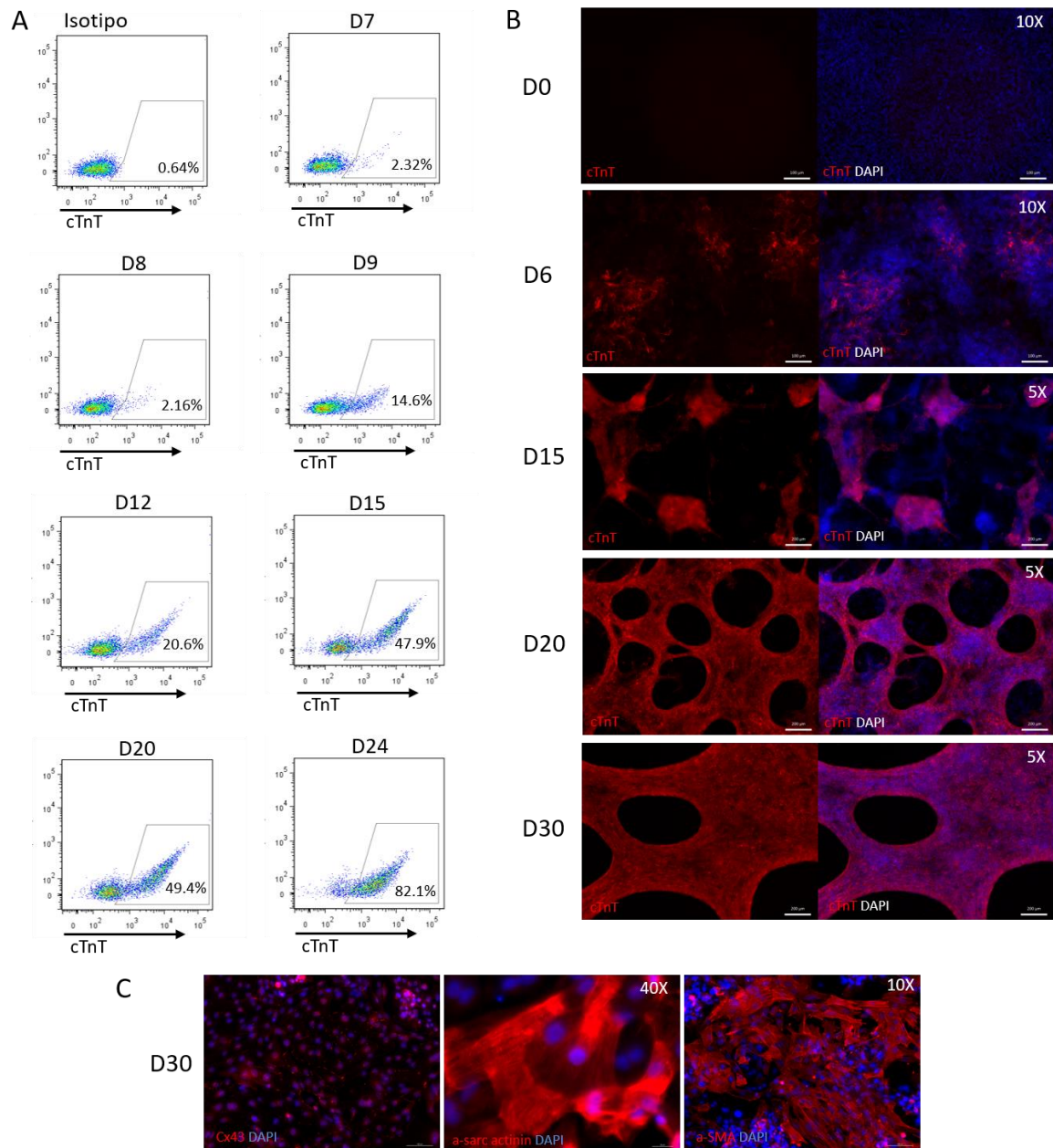


Figura 5.2 Expressão de marcadores cardíacos durante a diferenciação em monocamada. A expressão de cTnT foi acompanhada em diferentes dias do protocolo por (A) citometria de fluxo e (B) imunofluorescência. (C) Cx43, α -actinina sarcomérica e α -SMA também mostraram expressão no D30 da diferenciação.

5.2 Validação dos genes candidatos escolhidos para avaliação funcional

Os genes candidatos a participantes da diferenciação cardíaca foram escolhidos baseado nos dados de RNA-seq (Capítulos I e II). Esses dados foram gerados usando a diferenciação por corpos embrióides (CE) como modelo, e o sequenciamento dos RNAs livres e associados aos ribossomos. Para poder validar a expressão dos genes no modelo de diferenciação em monocamada, qPCRs foram

realizados usando RNA total em diferentes dias da diferenciação (Figura 5.3). Vale lembrar que os dois modelos de diferenciação diferem um pouco com relação ao tempo (dia de diferenciação) e etapa de desenvolvimento, podendo haver pequenas variações na expressão dos genes. Observou-se que, de maneira geral, os genes apresentaram o mesmo padrão de expressão durante a diferenciação cardíaca nos dois protocolos. Com exceção de *SHISA3*, que apresentou maior variação entre as duas.

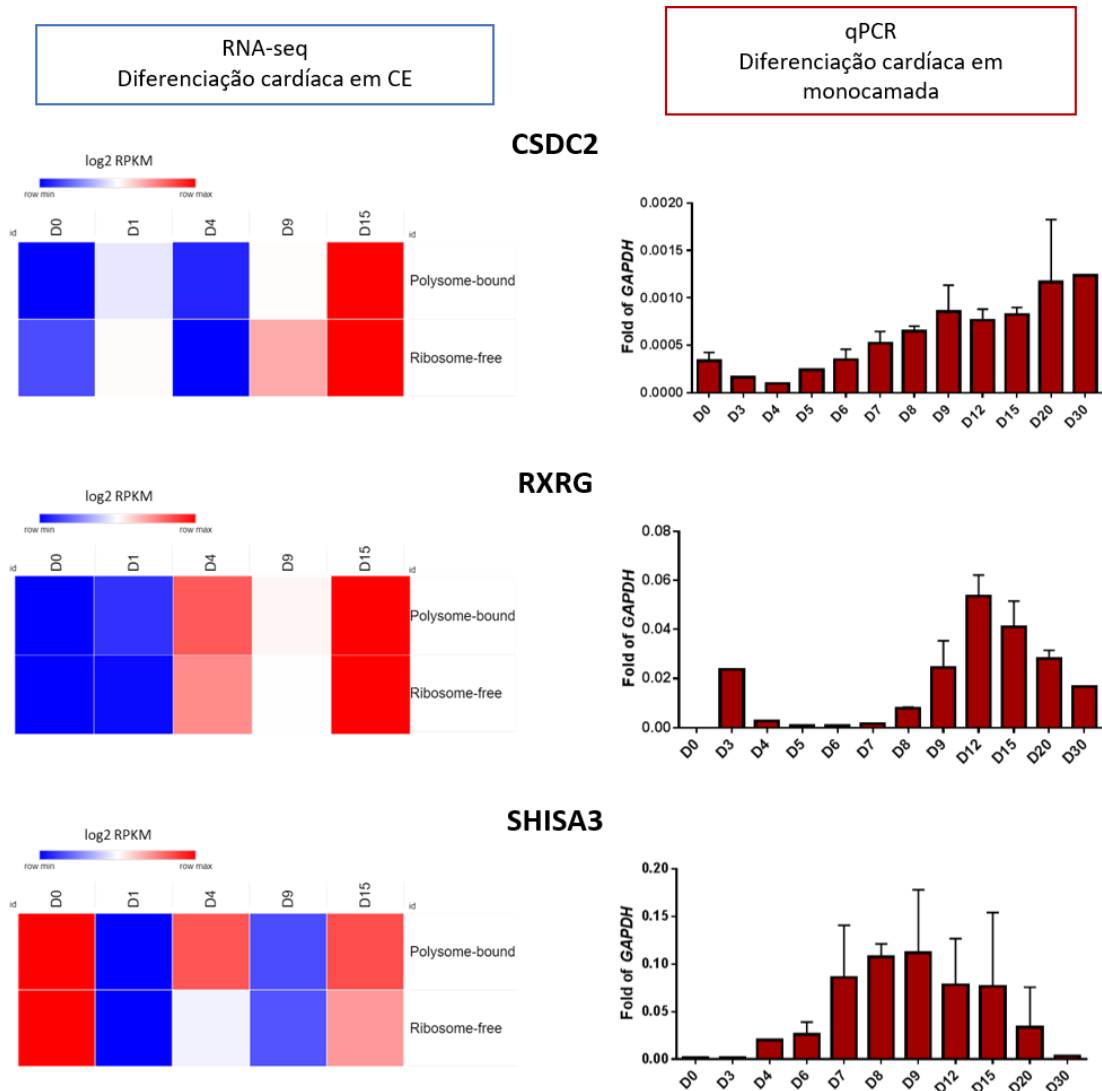


Figura 5.3 Expressão dos genes candidatos ao longo da diferenciação cardíaca em CE e monocamada. Os valores de expressão baseados no RNA-seq estão expressos em log₂ de RPKM, e os de qPCR são valores relativos a expressão do gene constitutivo GAPDH. Os dias de diferenciação estão indicados na figura, assim como as frações de RNA livre (Ribosome-free) ou associado a ribossomos (Polysome-bound).

5.3 Construção de linhagens celulares de super expressão induzida

A fim de avaliar a participação dos genes escolhidos na diferenciação cardiomiogênica, a construção de linhagens de CTP com o sistema de super expressão induzível foi feita para cada um dos genes citados. O vetor AAVS1-TRE-FLAG-rtTA foi clonado com as CDSs de *RXRG*, *SHISA3* e *CSDC2* e confirmados por reação de restrição (ANEXO III – Figura 1) e sequenciamento. Tais vetores foram usados para promover a integração do sistema no locus AAVS1 do genoma das CTP. Após a transfecção, foi feita a seleção com puromicina por, aproximadamente, 15 dias e a seleção clonal. As novas linhagens foram denominadas de acordo com as células parentais, o transgene integrado e com o prefixo “i” indicando induzível: H1 iRXRG, H1 iSHISA3, H1 iCSDC2 e H1 iCSDC2-del. Até seis clones de cada nova linhagem foram selecionados e testados quanto a eficiência de indução da expressão do transgene através da adição de doxiciclina ao meio de cultivo. A quantificação da expressão foi feita através de qPCR, e um clone de cada condição foi escolhido para dar continuidade ao trabalho (Figura 5.4). Além da quantificação da expressão do transgene, a expressão basal de cada clone também teve que ser levada em consideração, uma vez que o sistema, mesmo sem indução, pode ser expresso. Essa expressão basal é chamada de vazamento e pode interferir com a interpretação dos resultados. Os valores de expressão basal foram plotados isoladamente para melhor visualização no ANEXO III – Figura 2.

Os clones selecionados das construções H1 iCSDC2 (clone 4) e H1 iRXRG (clone 4) também tiveram a expressão induzida confirmada através da marcação por imunofluorescência, após 7 dias de indução com 500 ng/ml de doxiciclina (Figura 5.5). Os resultados mostram a expressão citoplasmática de CSDC2 e majoritariamente nuclear de RXRG (Figura 5.5). Além disso, constatou-se a heterogeneidade da população de células, indicando, possivelmente, a presença de células selvagens na cultura.

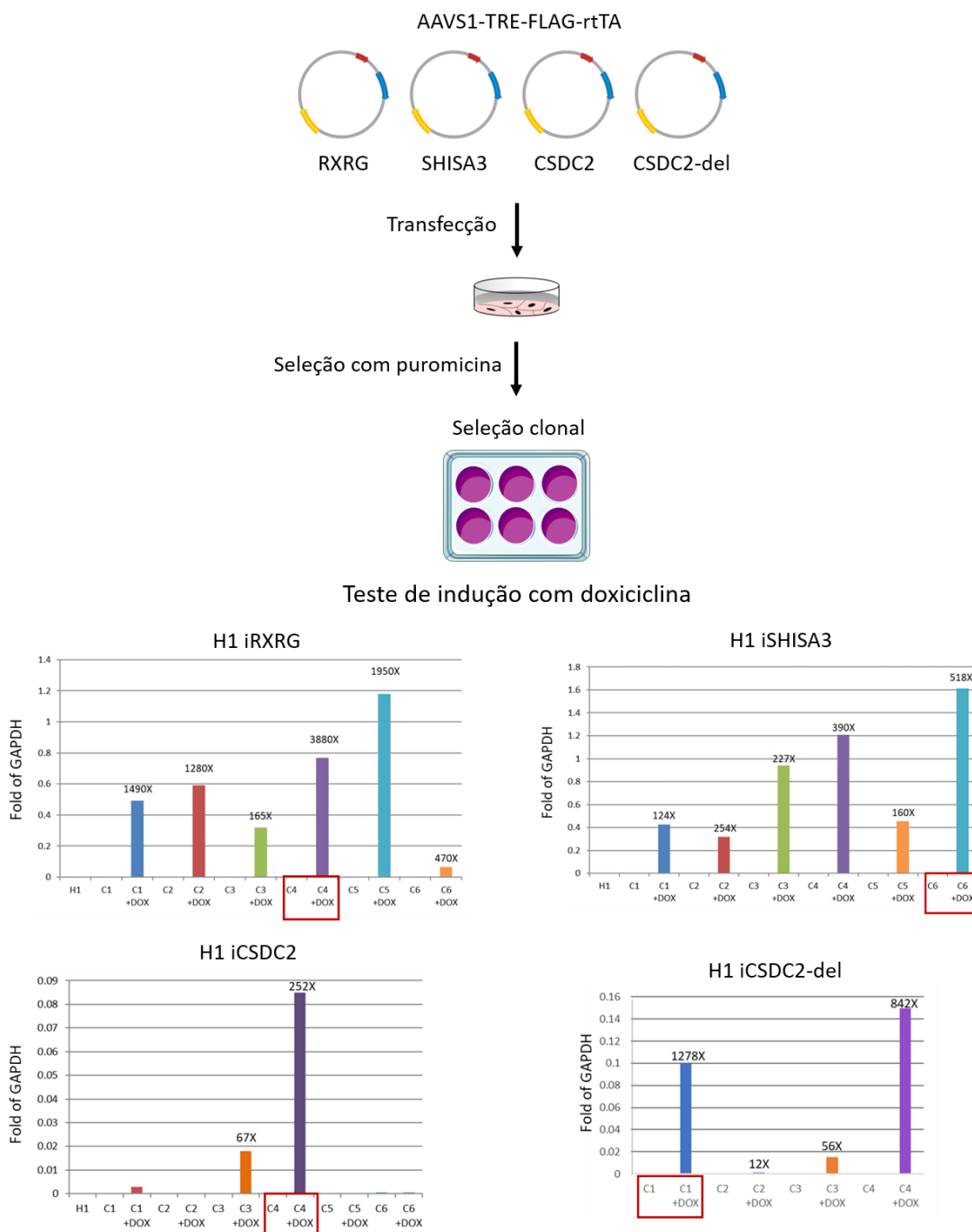


Figura 5.4 Sistema de super expressão induzível com doxiciclina. Após a transfecção com as construções vetoriais, as células foram submetidas à seleção com puromicina e clonal. Os clones selecionados foram testados quanto à expressão dos transgenes através da indução com 500 ng/ml de doxiciclina (+DOX). Valores relativos ao gene constitutivo GAPDH estão mostrados nos gráficos, e foram quantificados por qPCR. Todos os qPCR foram feitos com probes Taqman, exceto iCSDC2-del que foi quantificado usando os oílgos listados na Tabela 4.2. Os clones escolhidos para a continuidade do trabalho estão destacados em vermelho. H1, expressão nas células parentais sem transfecção. Valores indicados nas colunas +DOX representam a expressão quantas vezes maior que o controle sem DOX.

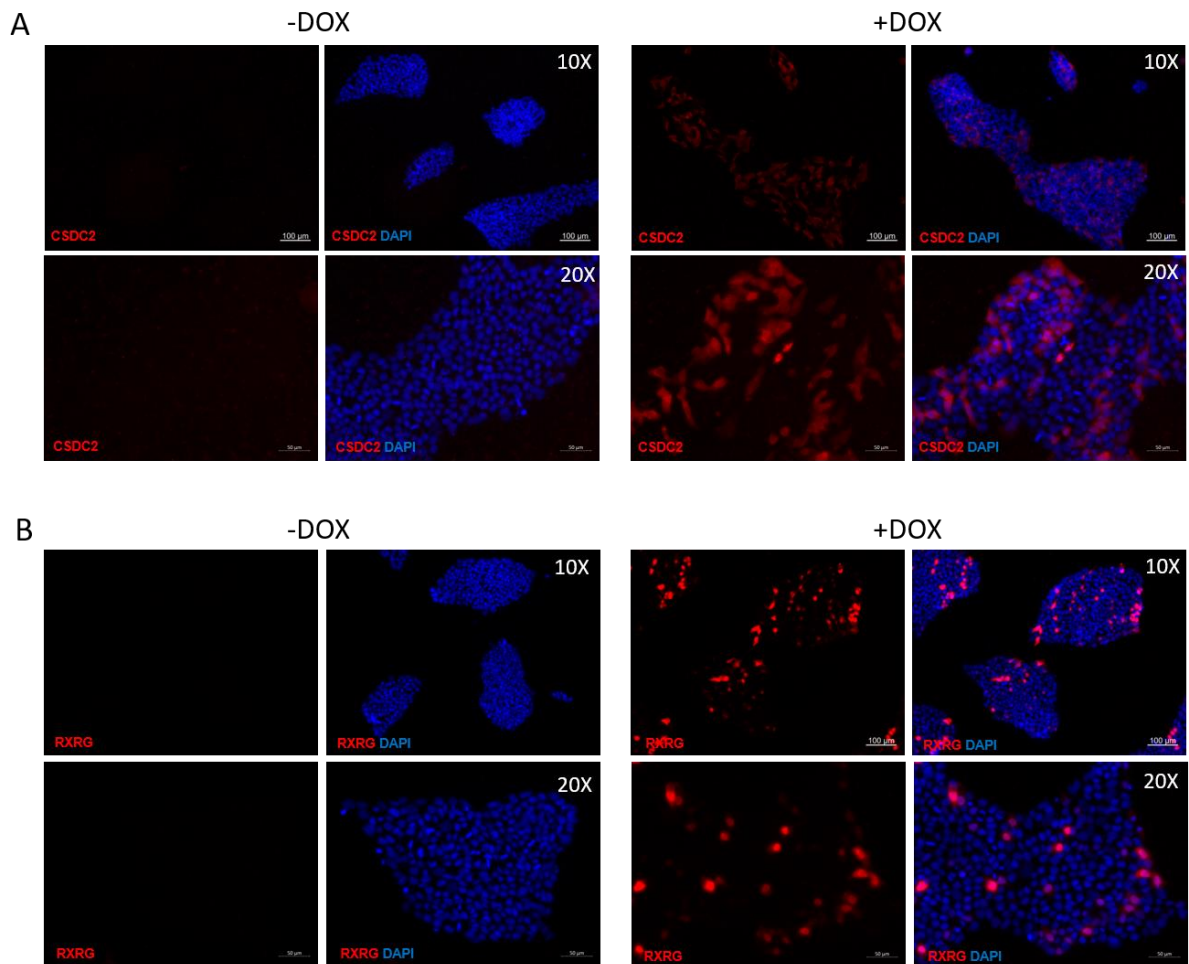
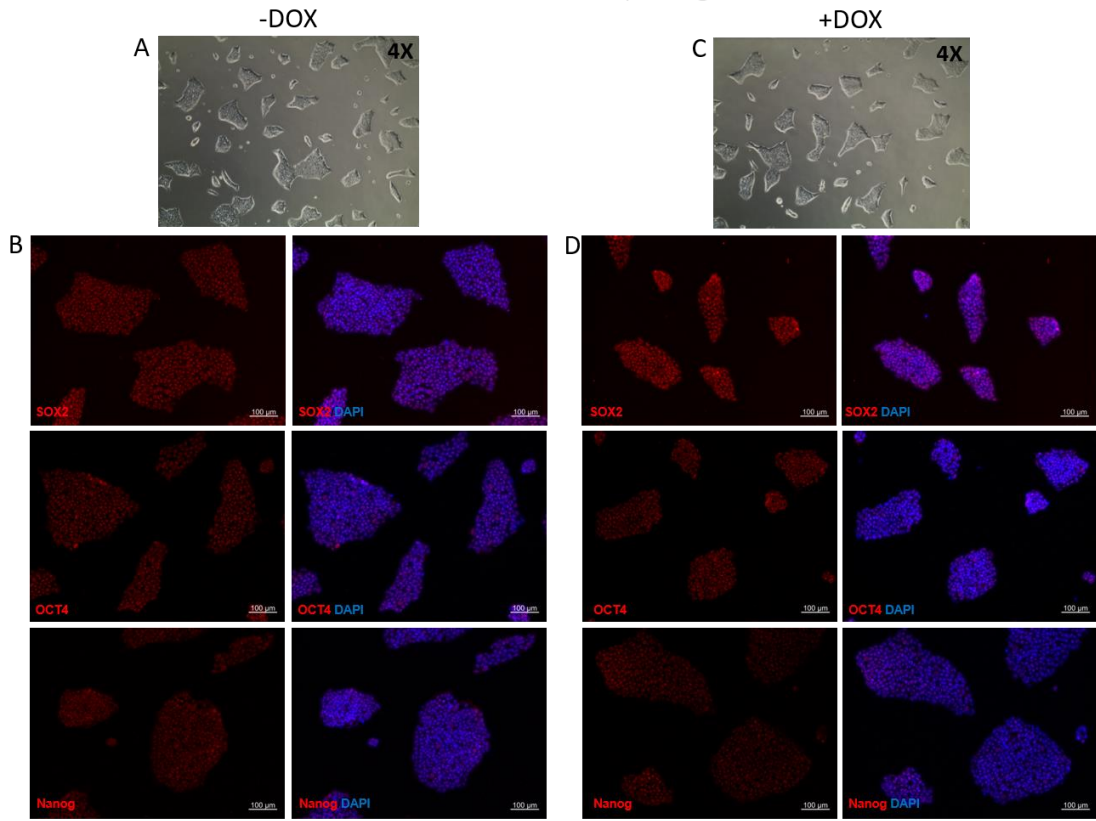


Figura 5.5 Marcação com imunofluorescência mostrando a expressão de CSDC2 e RXRG nas células (A) H1 iCSDC2 (clone 4) e (B) H1 iRXRG (clone 4), respectivamente, após 7 dias de indução com 500 ng/ml de doxiciclina (+DOX). Células não induzidas (-DOX) foram usadas como controle. Coloração nuclear com DAPI.

Após as etapas de seleção com puromicina e clonal, e de algumas passagens em cultura, os clones selecionados mantiveram as características de células pluripotentes (Figura 5.6, painéis da esquerda (-DOX)). Entre essas características estão a morfologia em colônias bem delimitadas (Figura 5.6 A, E e I) e a expressão dos marcadores de pluripotência Oct4, Sox2 e Nanog (Figura 5.6 B, F e J). Além disso, essas características também foram mantidas após a indução com doxiciclina (Figura 5.6, painéis da direita (+DOX)), indicando que os genes *RXRG*, *SHISA3* e *CSDC2* parecem não participar da saída do estado de pluripotência.

H1 iRXRG - clone #4, passagem 8



H1 iSHISA3 - clone #6, passagem 8

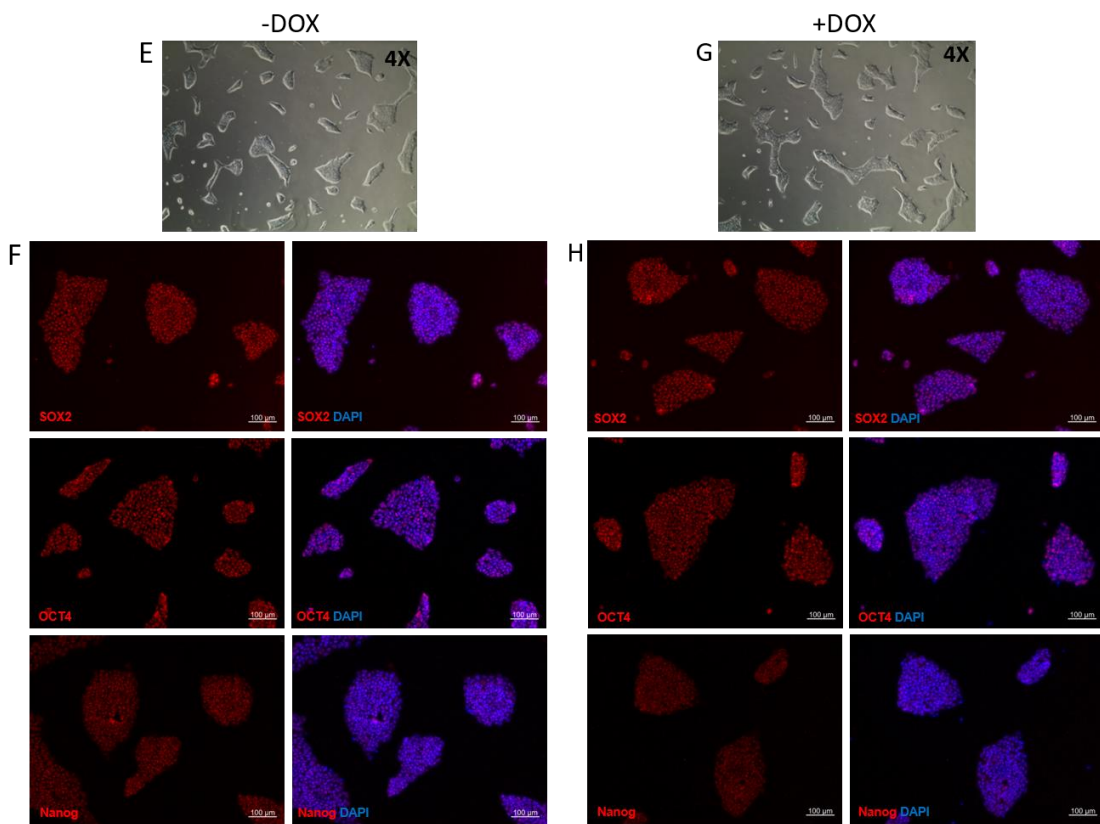


Figura 5.6. Continua na próxima página.

H1 iCSDC2 - clone #4, passagem 9

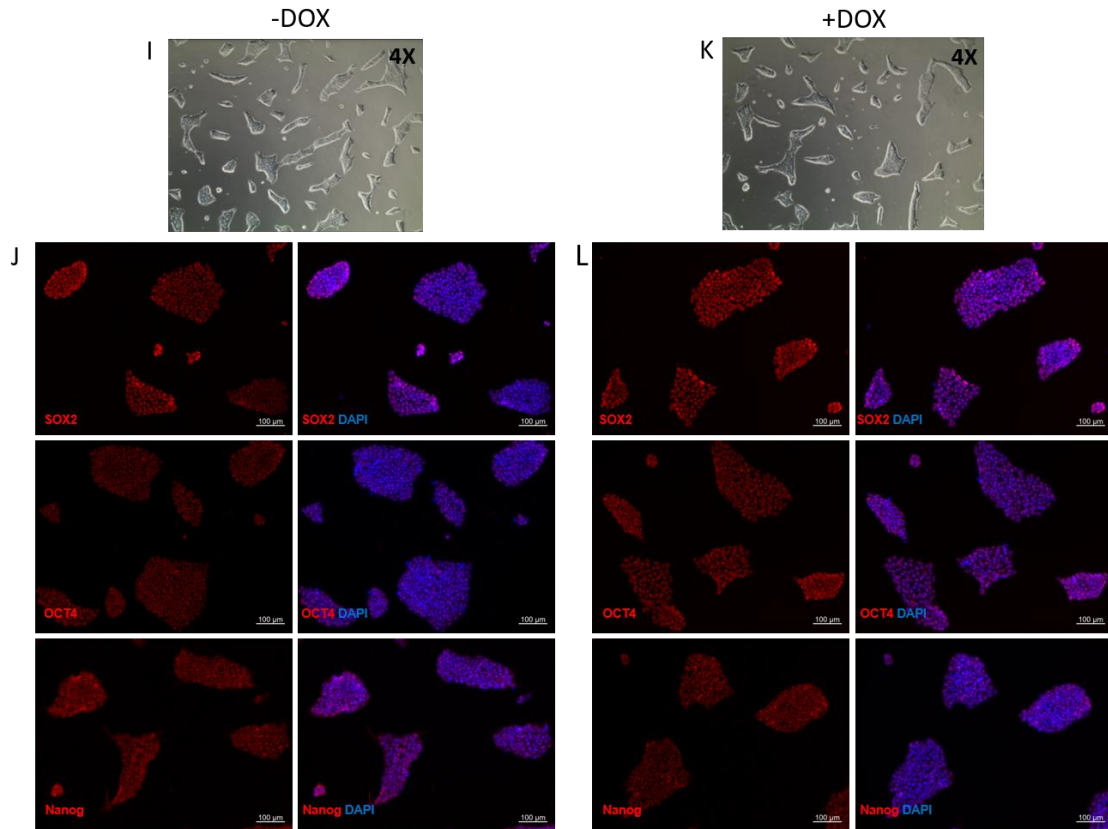


Figura 5.6 Linhagens celulares H1 iRXRG, iSHISA3 e iCSDC2 mantiveram características pluripotentes após algumas passagens em cultivo (-DOX) e após indução com doxiciclina (+DOX). Imagens em campo claro (A, C, E, G, I, K) foram tiradas após 96 h sem (-DOX) ou com indução (+DOX), no aumento de 4X, e mostram a morfologia das colônias de células. Imunofluorescência dos marcadores de pluripotência Oct4, Sox2 e Nanog das células sem indução (-DOX: B, F, J) e induzidas com doxiciclina (+DOX: D, H, L), após 7 dias. Imagens no aumento de 10X. Coloração nuclear com DAPI.

5.4 Ensaio funcionais com a linhagem H1 iRXRG

A linhagem H1 iRXRG (clone #4) foi submetida à diferenciação cardiomiogênica em monocamada a fim de avaliar a influência da super expressão do gene *RXRG* no processo de diferenciação. Diferentes janelas de indução com doxiciclina foram testadas e a dupla expressão dos marcadores NKX2-5 e TNNT2, indicando a presença de cardiomiócitos, foi avaliada no dia 15 da diferenciação através de citometria de fluxo (Figura 5.7). Um primeiro experimento foi realizado (Figura 5.7 A) e pôde se observar um aumento de 10 a 15% no número de células duplo positivas quando a indução foi feita no final da diferenciação, dos dias D7-10 e D10-15. Essas janelas de indução foram, então, escolhidas para os próximos experimentos, para que

pudesse se avaliar a reprodutibilidade do fenótipo observado. Entretanto, a diferenciação cardíaca não funcionou mais adequadamente nessas células, como pode ser observado na Figura 5.7 B, onde o controle (-DOX) gerou apenas em torno de 5% de células duplo positivas (quando o ideal seria em torno de 40%). Mesmo com essa limitação, pode se observar um aumento no número de células duplo positivas quando o gene foi induzido entre os dias D7-10, D7-15 e D10-15, indicando um possível efeito positivo da expressão de *RXRG* na diferenciação cardíaca. O clone #6 também foi usado na tentativa de diferenciar as células a cardiomiócitos, sem sucesso (dados não mostrados). Experimentos adicionais ainda precisam ser realizados para que se possa inferir sobre a funcionalidade de *RXRG* durante a cardiomiogênese.

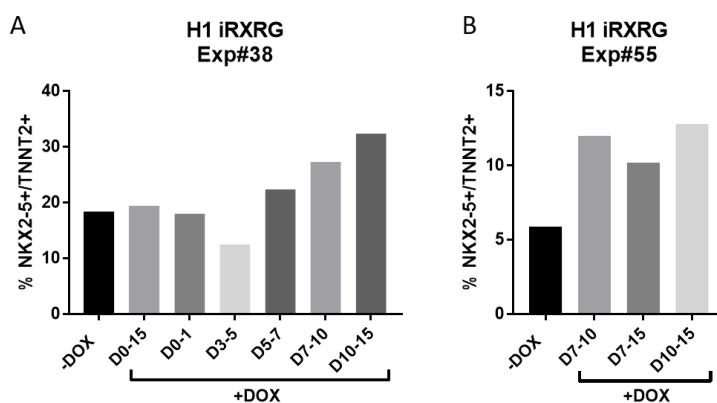


Figura 5.7 Ensaio funcional de diferenciação cardiomiogênica com a linhagem H1 iRXRG. Diferentes janelas de indução com doxiciclina (+DOX) foram testadas (por exemplo, D0-15 a indução foi feita do dia 0 ao dia 15 do protocolo de diferenciação). Células sem indução (-DOX) foram usadas como controle. A porcentagem de células duplo positivas NKX2-5+/TNNT2+ está representada nos gráficos em cada um dos experimentos independentes (A – Experimento #38 e B – Experimento #55) e foi medida através da imuno-marcação e leitura por citometria de fluxo.

5.5 Ensaio funcionais com a linhagem H1 iSHISA3

A linhagem H1 iSHISA3 (clone #6) foi usada para se avaliar a participação do gene *SHISA3* na diferenciação cardiomiogênica. Diferentes janelas de indução também foram testadas e a porcentagem de células NKX2-5+/TNNT2 duplo positivas foi avaliada por citometria de fluxo. Nos dois experimentos iniciais de triagem (Figura 5.8 A e B), observou-se uma pequena variação na porcentagem de cardiomiócitos nas populações que tiveram a expressão de *SHISA3* induzida em diferentes momentos. Nesses dois experimentos, o fenótipo observado quando a indução se deu do dia 0

ao dia 1 (D0-1) se repetiu, indicando uma possível redução na eficiência de diferenciação (n=2) após indução com doxíciclina (Figura 5.8 A, B e C).

Uma vez que o gene *SHISA3* parece estar envolvido na regulação negativa da via de sinalização Wnt, e o protocolo de diferenciação em monocamada se baseia na modulação dessa via através de pequenas moléculas (Figura 1.4B), hipotetizou-se que a indução de *SHISA3* por doxíciclina poderia substituir a molécula XAV939 na inibição de Wnt no dia 3 do protocolo. Para isso, duas condições foram usadas: sem indução por doxíciclina e sem XAV939; e com indução por doxíciclina e sem a utilização de XAV939 (Figura 5.8 D e E). A primeira condição (-DOX/-XAV939) mostrou uma redução na eficiência de diferenciação, ilustrada pela porcentagem de células NKX2-5/TNNT2 duplo positivas, quando comparada com o controle +XAV939. Por sua vez, a condição de indução com doxíciclina (+DOX/-XAV939) não foi capaz de recuperar a eficiência de diferenciação, indicando que a expressão de *SHISA3* nessas condições não substitui o efeito de XAV939.

Experimentos adicionais ainda precisam ser realizados para que se possa inferir sobre a atuação de *SHISA3* na diferenciação cardiomiogênica.

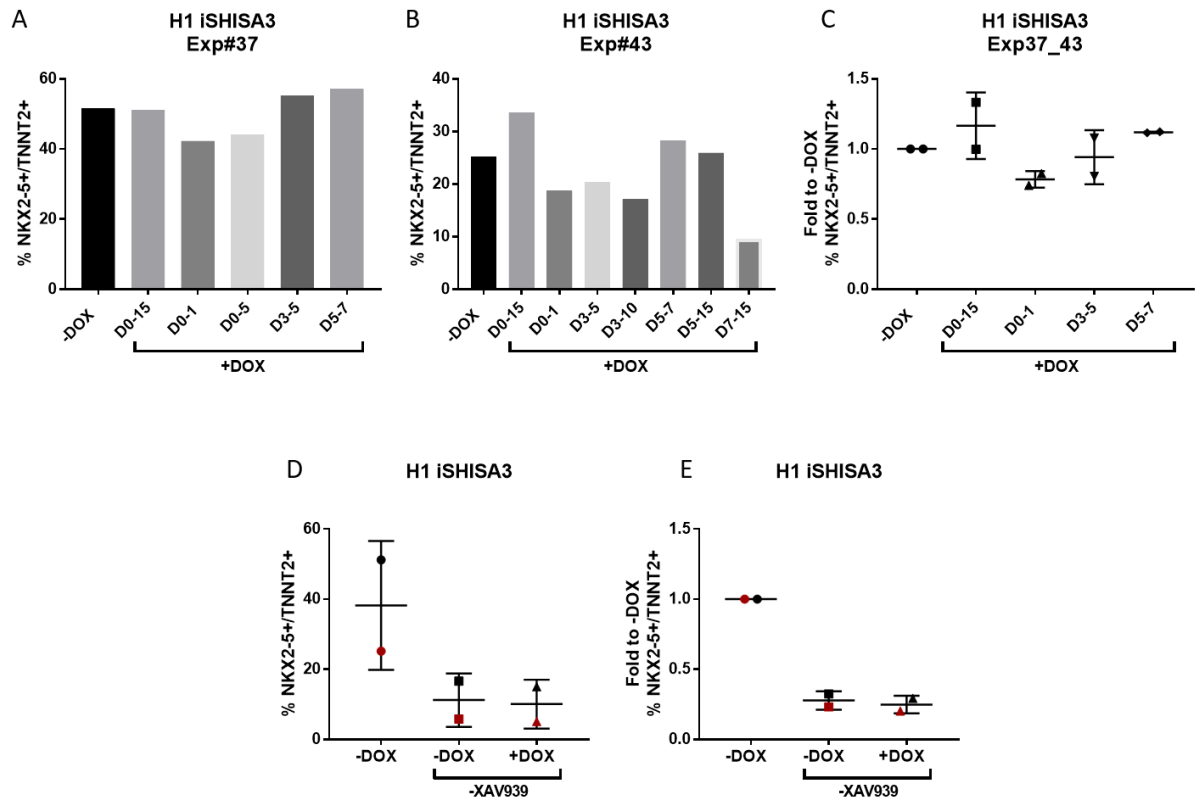


Figura 5.8 Ensaio funcional de diferenciação cardiomiogênica com a linhagem H1 iSHISA3. Diferentes janelas de indução com doxiciclina (+DOX) foram testadas (por exemplo, D0-15 a indução foi feita do dia 0 ao dia 15 do protocolo de diferenciação). Células sem indução (-DOX) foram usadas como controle. A porcentagem de células duplo positivas NKX2-5+/TNNT2+ está representada nos gráficos em cada um dos experimentos independentes (A – Experimento #37 e B – Experimento #43) e foi medida através da imuno-marcação e leitura por citometria de fluxo. (C) Dados mostrados em A e B também foram plotados com relação ao controle -DOX (n=2). Condições sem XAV939 foram testadas e estão mostradas nos gráficos como (D) valores brutos de porcentagem e (E) valores relativos ao controle -DOX (n=2).

5.6 Ensaio funcional com a linhagem H1 iCSDC2

Experimentos de diferenciação cardíaca também foram realizados utilizando a linhagem H1 iCSDC2 (clone #4) para avaliar o possível envolvimento do gene *CSDC2* na diferenciação. Dois experimentos de triagem inicial foram feitos usando diferentes janelas de indução com doxiciclina e a porcentagem de células NKX2-5/TNNT2 duplo positivas foi avaliada por citometria de fluxo (ANEXO IV – Figura 1). Esses resultados mostraram um singular aumento na eficiência de diferenciação cardíaca após a indução da expressão de *CSDC2* em diferentes janelas de tempo. Baseado nesses resultados e na expressão basal de *CSDC2* durante a diferenciação cardíaca (Figura

5.3), a janela de indução entre os dias 7 e 10 (D7-10) do protocolo de diferenciação foi escolhida para a investigação da reprodutibilidade e significância do fenótipo observado.

Experimentos de diferenciação em replicatas independentes foram realizados e avaliados por citometria de fluxo e qPCR no dia 15 do protocolo de diferenciação (Figura 5.9). A indução de *CSDC2* entre os dias 7 e 10 proporcionou um aumento significativo na porcentagem de células NKX2-5/TNNT2 duplo positivas (Figura 5.9A) e na expressão das troponinas cardíacas *TNNT2*, *TNNI1* e *TNNI3* (Figura 5.9B). Além disso, outros marcadores cardíacos, que são utilizados para diferenciar células atriais e ventriculares também foram avaliados. A expressão do gene *IRX4*, típico de células ventriculares, se mostrou significativamente aumentada após indução com doxíciclina (Figura 5.9C). Os genes *ACTA2* e *VWF*, marcadores de células musculares lisas e endoteliais, respectivamente, não tiveram sua expressão alterada, indicando que a super expressão de *CSDC2* contribui apenas para o aumento da diferenciação de cardiomiócitos e não de outras células derivadas dos mesmos progenitores (Figura 5.9D). Também foram avaliados genes envolvidos na função do Ca^{2+} (*SERCA2*), canal de íons (*KCNJ2*) e metabolismo de lipídeos (*LCAD*), indicadores de maturação dos cardiomiócitos. Nenhum desses genes apresentou a expressão alterada após a indução com doxíciclina (Figura 5.9E). Células do clone #3 também foram submetidas à diferenciação e confirmaram os resultados obtidos anteriormente.

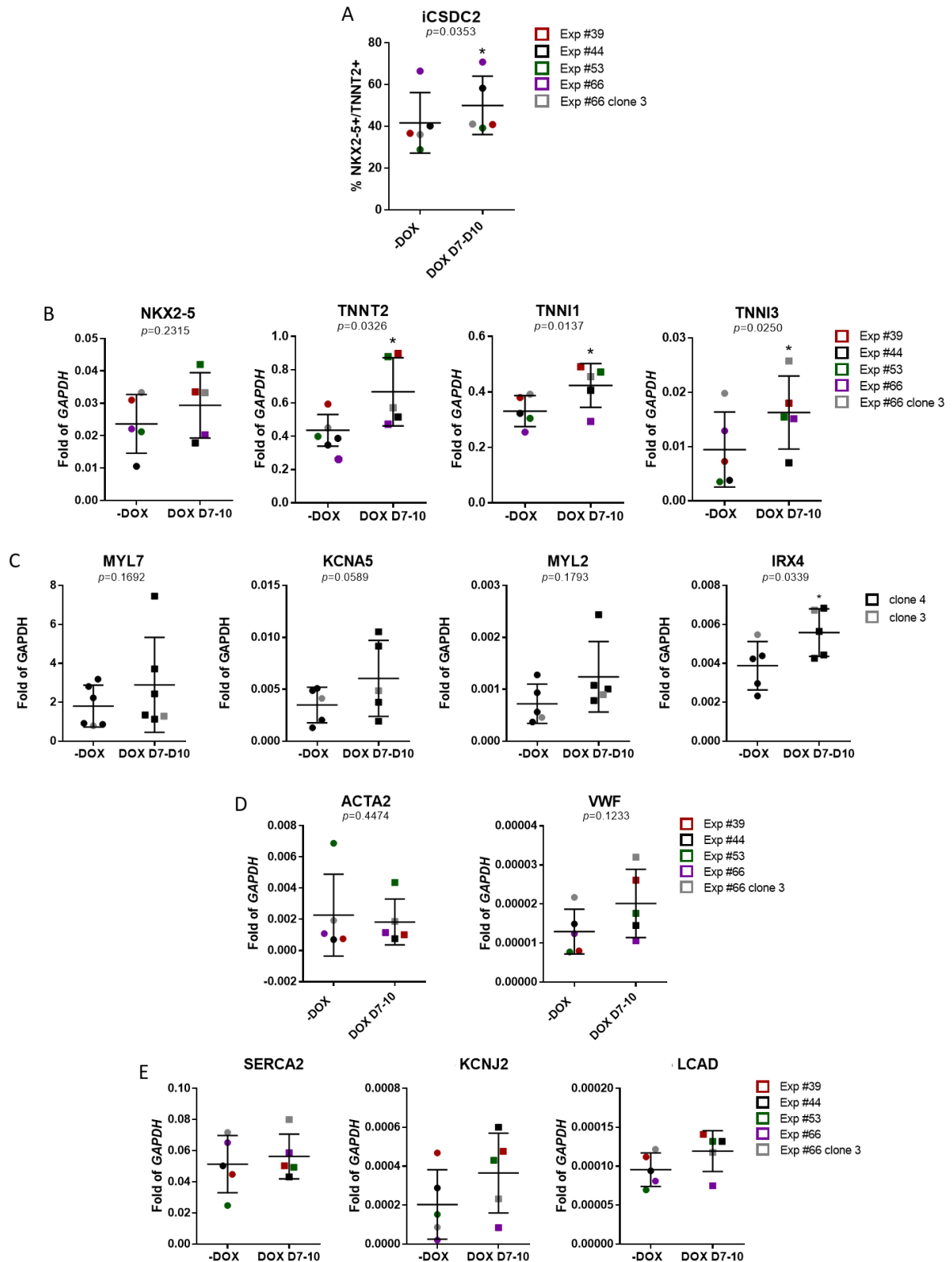


Figura 5.9 Ensaio funcional de diferenciação cardíaca com a linhagem H1 iCSDC2. No dia 15 do protocolo de diferenciação, células não induzidas (-DOX) e induzidas com doxiciclina entre os dias 7-10 (DOX D7-10) foram avaliadas por (A) citometria de fluxo (porcentagem de células NKX2-5⁺/TNNT2⁺) e (B, C, D e E) qPCR (valores de expressão relativa a do gene constitutivo GAPDH). As replicatas independentes estão indicadas por cores, assim como os valores correspondentes ao clone #3. Análise por teste T de Student pareado: * $p < 0,05$.

5.6.1 Perspectivas para a análise do gene *CSDC2*

A análise funcional do gene *CSDC2* será continuada com a avaliação da atividade do domínio CSD e motivos de ligação à RNA através de ensaios com a linhagem H1 *CSDC2*-del em experimentos de diferenciação cardíaca. A expressão da proteína *CSDC2* recombinante também está prevista, através da construção no plasmídeo pET28b, para ser usada em ensaios *in vitro* e para a produção de anticorpos policlonais. Também já foi planejada a estratégia de silenciamento de *CSDC2* utilizando CRISPR/Cas9. Essas abordagens estão previstas para serem desenvolvidas em um novo projeto de Doutorado no Laboratório de Biologia Básica de Células-tronco.

CAPÍTULO IV

Artigo

Polysome-associated lncRNAs along cardiomyogenesis of hESC

Situação – Em preparação para submissão na revista *Non-coding RNA*.

Prefácio

Este capítulo apresenta o artigo escrito para submissão na revista *Non-coding RNA*, que descreve a análise dos dados de RNA-seq ao longo da diferenciação cardíaca de CTE com enfoque nos RNA não-codificantes longos (lncRNAs). Esse manuscrito teve como objetivo a análise da associação diferencial de lncRNAs aos polissomos nas diferentes etapas da diferenciação cardiomiogênica e a regulação pós-transcricional que poderiam estar envolvidos. A expressão de lncRNAs se mostrou interessante nos dados de RNA-seq descritos nesta Tese e por isso essa análise foi incluída.

Polysome-associated lncRNAs along cardiomyogenesis of hESC

Isabela Tiemy Pereira¹, Lucia Spangenberg², Guillermo Cabrera², Bruno Dallagiovanna^{1*}

¹ Basic Stem-cell Biology Laboratory, Instituto Carlos Chagas - FIOCRUZ-PR, Rua Professor Algacyr Munhoz Mader, 3775, Curitiba, PR 81.350-010, Brazil

² Bioinformatics Unit, Institut Pasteur de Montevideo, Mataojo 2020, Montevideo, 11400, Uruguay

*Correspondence: bruno.dallagiovanna@fiocruz.br.

Abstract

Long non-coding RNAs (lncRNAs) have been found to be involved in many biological processes, including the regulation of cell differentiation, but a complete characterization of lncRNA is still lacking. Additionally, there is evidence that lncRNAs interact with ribosomes, raising questions about their functions in cells. Here, we used a developmentally staged protocol to induce cardiogenic commitment of hESCs and then investigated the differential association of lncRNAs with polysomes. Our results identified lncRNAs in both the ribosome-free and polysome-bound fractions during cardiogenesis and showed a very well-defined temporal lncRNA association with polysomes. Clustering of lncRNAs was performed according to the gene expression patterns during the five timepoints analyzed. In addition, differential lncRNA recruitment to polysomes was observed when comparing the differentially expressed lncRNAs in the ribosome-free and polysome-bound fractions or when calculating the polysome-bound vs ribosome-free ratio. The association of lncRNAs with polysomes could represent an additional cytoplasmic role of lncRNAs, e.g., in translational regulation of mRNA expression.

Keywords

long non-coding RNA, hESC, cardiomyocyte, RNA-seq

Introduction

Long non-coding RNAs (lncRNAs) are non-coding RNAs that are arbitrarily characterized as being longer than 200 bp and are structurally indistinguishable from protein-coding and processed messenger RNAs since they are also 5' capped and polyadenylated [1,2]. A plethora of lncRNAs have been identified that are associated with many biological processes, such as embryogenesis, the cell cycle, pluripotency, apoptosis and differentiation [1,3]. lncRNAs are more tissue-specific than protein-coding genes; however, they tend to be expressed at low levels [4–6]. lncRNAs can be classified according to their proximity to protein-coding genes as sense, antisense, divergent, intronic or intergenic [7].

The mechanisms by which lncRNAs act still need to be determined, but there is robust evidence that they play crucial roles in gene expression regulation at different levels, including chromatin organization, transcriptional regulation and post-transcriptional control [7]. Although there is a lack of conserved lncRNA sequences between species [4,8,9], their secondary structure might actively participate in their function through interactions with proteins and other RNAs [10–12].

lncRNAs can also interact with ribosomes and can be found in polysomal complexes [13–16]. The association of lncRNAs with polysomes might indicate the coding potential of these transcripts, representing a source of new peptides [17]. On the other hand, ribosome footprints have been found along lncRNAs as well as in previously known non-coding sequences, suggesting pervasive translation in the cell [18]. In addition to the coding potential of lncRNAs, their association with polysomes could represent a role in translational regulation, for instance, by fine-tuning the speed or specificity of the machinery and titrating out ribosomes [1]. In addition, the base pairing capability of lncRNA indicates that they can interact with and regulate specific mRNAs [7,11].

Cell differentiation requires the activation of specific genetic programs, and lncRNA expression has been found to contribute to cell commitment processes (revised by [3]). Analysis of the murine and human stem cell transcriptomes during cardiomyocyte differentiation revealed thousands of differentially expressed lncRNAs [19–23]. These lncRNAs have temporal-specific expression, suggesting that they play a crucial role in cardiac development regulation [19–24]. Identification of new lncRNAs and their tissue specificity and function could reveal that they are a new class of regulatory molecules in cardiac development [25].

Our group has previously used polysome profiling followed by RNA-seq to analyze the expression and polysome association of lncRNAs in undifferentiated human adipose-derived stem cells and committed to differentiation into adipose cells [14]. Here, we used a hESC cardiomyogenesis model associated with polysome profiling to investigate temporally regulated lncRNAs and their association with polysomes during human cardiac commitment.

Results

1. LncRNAs are expressed during cardiomyogenesis of hESCs

Models of in vitro differentiation of hESCs provide the ability to study developmental biology and address the key events that regulate early lineage commitment [26]. Our group used a developmentally staged protocol to induce cardiogenic commitment of hESCs [27,28] and perform polysome profiling, followed by RNA-seq, to investigate the differential association of RNAs with polysomes, as described in [29]. Here, we focused on analyzing lncRNAs expressed during cardiomyogenesis, mainly those differentially associated with polysomes. Considering an RPKM mean of at least 0.01 between biological triplicates (of each timepoint and distinct RNA fraction), 6693 lncRNAs were identified as being expressed during the five timepoints of cardiomyogenesis analyzed: days D0, D1, D4, D9 and D15 of differentiation, which represent pluripotency, embryoid body (EB) aggregation, cardiac mesoderm, cardiac progenitor and cardiomyocyte stages, respectively (Figure 1A). The lncRNAs identified during cardiomyogenesis represent 17.7% of the total genes expressed (Figure 1B), considering the filter mentioned above. The annotated categories included in this analysis were based on Ensembl gene biotypes: antisense, lincRNA, processed transcript, sense intronic, sense overlapping and TEC (to be experimentally confirmed), and the number of genes in each category is shown in Figure 1C. Principal component analysis (PCA) was performed for the lncRNAs according to the type of RNA fraction (ribosome-free or polysome-bound) (Figure 1D). Ribosome-free samples were dispersed and showed less similarity between the day of differentiation and biological replicates (Figure 1D, left panel). On the other hand, polysome-bound samples showed more distinct groups relative to the day of differentiation, indicating high similarity between polysomal lncRNAs in experimental replicates and more pronounced temporal expression (Figure 1D, right panel).

2. Temporal expression patterns of lncRNAs during cardiomyogenesis

The polysome profiling approach allowed us to isolate and separately sequence polysome-bound and ribosome-free RNAs from cardiomyogenesis during the five timepoints. Comparisons between each differentiation timepoint and the preceding timepoint, considering an overall FDR of ≤ 0.05 and $-2 \geq \log_{2}FC \geq 2$, identified differentially expressed (DE) lncRNAs in each RNA fraction. DE lncRNAs were plotted on a heatmap and showed a clear temporal expression pattern as either polysome-bound or ribosome-free RNAs (Figure 2). Hierarchical clustering grouped more highly expressed lncRNAs into distinct stages, and we classified them into six groups. For instance, some groups that showed high expression in cardiac committed cells (days D9 and D15) - group I; in the mesoderm stage (day D4) - group III; and in the pluripotency state (days D0 and D1) - group V. Not all groups were present in both the ribosome-free and polysome-bound fractions. Mesoderm stage group III only had a few genes in the ribosome-free fraction. Additionally, intermediate expression group IV, which had higher expression on days D1, D4 and D9, was only found in the polysome-bound fraction. Comparing the expression pattern of the ribosome-free and polysome-bound fractions, we observed more stated groups on polysome-bound lncRNAs and more progressive patterns on ribosome-free lncRNAs (Figure 2). Interestingly, the temporal differential expression observed in polysome-bound RNAs could suggest that lncRNAs differentially associate with polysomes during cardiomyogenesis and that they play a role in this process.

In addition, clustering analysis of lncRNAs showed particular expression patterns associated with cardiomyogenic commitment (in the abovementioned time points: D0, D1, D4, D9, D15). Figure 3A shows the results of a PCA analysis of the genes of ribosome-free (right panel) and polysome-bound (left panel) lncRNAs (with at least one CPM in three samples), in which 7 clusters were identified (marked with an ellipse in the panels). The gene expression profiles of the 7 clusters for both RNA fractions were determined (Supplementary Figure 1 and 2, polysomal and free, respectively). Figure 3B shows the expression profiles of two example clusters of polysome-bound lncRNAs. In the left panel, gene expression is high during the initial steps of cardiomyogenic differentiation (D0, D1, D4), with a strong decay afterwards. The second cluster shows genes with low expression during the initial steps of cardiomyogenic differentiation, with a peak at D4 and stabilization during the final

stages (right). Figure 3C shows two example clusters of lncRNA expression in the ribosome-free fraction. These two clusters behave very similarly to those mentioned above. To the left, a constant decrease of gene expression is observed from D4 onwards, and to the right, a steady increase of gene expression is observed, with a peak at D4.

Similarly behaving clusters of lncRNAs in the polysome-bound and ribosome-free fractions have lncRNAs in common, e.g., 38% in the case of the left ones and 14% in the case of the right ones (Figure 3B and 3C, left and right panels, respectively). Interestingly, several lncRNAs are specific to each fraction; the right panel clusters have 107 lncRNAs specific to the ribosome-free RNA fraction and 97 lncRNAs specific to the polysome-bound fraction (the Venn diagrams in Supplementary Figure 3A and 3B). Twenty out of 107 lncRNAs belong to clusters that show no change between time points in the polysomal fraction (Supplementary Figure 1, polysome-bound clusters 1, 2, 6). The rest of the lncRNAs are not even expressed in the polysomal fraction. Forty-five polysomal-specific lncRNAs belong to clusters that show no change in ribosome-free RNA (Supplementary Figure 2, ribosome-free clusters 2, 4 and 7). These results suggest a degree of post-transcriptional regulation of lncRNAs in cardiomyogenic differentiation. In addition, pathway analysis of the lncRNAs in each RNA fraction (polysome-bound and ribosome-free) was performed (Supplementary Table 1). Characteristic significant pathway terms (p -value <0.05) were observed in each fraction.

3. lncRNAs are recruited to polysomes in cardiomyogenic commitment of hESC

Considering the DE lncRNAs, we further investigated whether the lncRNA association with polysomes was only a consequence of transcriptional regulation, e.g., the more transcripts available, the more association with polysomes. We observed that some lncRNAs were only regulated in the polysome-bound fraction, only in the ribosome-free fraction or in both fractions, the latter representing coordinated regulation (Figure 4A). Interestingly, the majority of upregulated genes between days D0 vs D1 were only in the polysome-bound fraction, suggesting the recruitment of those transcripts to the polysome complex. Moreover, between days D9 vs D15, the majority of downregulated lncRNAs were only in the ribosome fraction, suggesting that some transcripts were held in the polysome-bound fraction when their transcription was decreased (Figure 4A). Essentially, all of the timepoints analyzed showed that

more lncRNAs were regulated in only either the polysome-bound or the ribosome-free fraction than in both fractions (2.5 to 40% of lncRNAs). This result could represent a specialized regulation of lncRNA association with polysomes, independent of the transcript abundance.

Once we observed there was differential lncRNA association with polysomes, we asked whether non-DE lncRNAs were also differentially recruited to and dissociated from polysomes during cardiomyogenesis. We calculated the ratio between the polysome-bound and ribosome-free RPKM values to assess the prevalence of lncRNA in each fraction. The results showed that some lncRNAs were affected by polysome recruitment ($FDR \leq 0.05$, $\log FC \geq 2$) or dissociation ($FDR \leq 0.05$, $-2 \geq \log FC$) when comparing each differentiation timepoint and the preceding timepoint during cardiomyocyte differentiation (Figure 4B).

Discussion

Cardiomyogenesis of hESCs provides an exceptional model to investigate the mechanisms of cell fate determination [26]. In this report, we analyzed five distinct timepoints during the differentiation of hESCs into cardiomyocyte to assess the expression of lncRNAs involved in the temporal cell fate transition. Our results identified a large amount of lncRNAs in both ribosome-free and polysome-bound fractions during cardiogenesis (Figure 1). LincRNA and antisense lncRNA were the most representative categories identified in our data, corroborating previous findings that antisense lncRNA may represent one of the largest lncRNA biotypes in mouse and human genomes [30]. The functions of antisense lncRNA are not yet well understood, but because they share sequence similarities with their sense transcripts, it is believed that they could regulate their sense transcripts at transcriptional or post-transcriptional levels [10,30], for instance, processing the sense strand by differential splicing or editing [31].

Transcriptome studies have revealed thousands of differentially expressed lncRNAs during cardiomyocyte differentiation [19–23], and some evidences have demonstrated the crucial roles that lncRNAs play in heart development [32]. The murine lncRNAs Braveheart and Fendrr are classical examples of lncRNAs that function in heart development [33,34]. These lncRNAs interact with the repressor complex Polycomb 2 and regulate gene expression during development. In humans, the antisense transcript of the Braveheart ortholog has been shown to be essential in

cardiac determination, acting by enhancing activity [35]. More recently, murine linc1405 was described in cardiac differentiating cells [36]. Murine linc1405 binds to the transcription factor Eomes and the histone modifiers WDR5 and GCN5 to form a complex that controls *Mesp1* expression. These examples illustrate that lncRNAs can act as regulators of gene transcription in cardiac differentiation. However, evidence showing their involvement in other modes of gene regulation, e.g., controlling mRNA translation, is still limited.

Here, we used polysome profiling and RNA-seq of ribosome-free and polysome-bound RNAs to investigate lncRNAs that were differentially associated with polysomes. Polysome profiling analysis is a robust method that has been used to provide valuable information regarding the association of ribosomes with RNAs [14,37,38]. Principal component analysis showed that polysome-bound lncRNAs were able to better reflect the biology of the samples than ribosome-free lncRNAs regarding the day of differentiation (Figure 1D). Additionally, comparisons between each differentiation timepoint and the preceding timepoint showed dramatic differential expression of lncRNAs during cardiogenesis, illustrating the molecular complexity of cardiac differentiation (Figure 2). Interestingly, we observed a very specific temporal expression of polysome-bound RNAs, suggesting not only that lncRNAs could differentially associate with polysomes during cardiomyogenesis but also that they could play a crucial role in this process. It has already been shown that lncRNAs can be found in polysomal complexes [13–16]; however, their function remains uncharacterized. In addition to their coding potential (e.g., small peptides) [17,39], the association of lncRNAs with polysomes could represent an additional cytoplasmic role in mRNA metabolism, for instance, in translational regulation by acting at the ribosome or mRNA level [1,7,11].

Moreover, we found that similar behaving clusters of lncRNAs in polysome-bound and ribosome-free fractions had few lncRNAs in common (Supplementary Figure 3), and all of the timepoints analyzed showed that more lncRNAs were only regulated in either the polysome-bound or only ribosome-free fraction (Figure 4A). Taken together, these results suggest that the association of lncRNAs with polysomes is independent of the transcript abundance and that specific lncRNAs are able to interact with the polysomal complex. Polysomal lncRNAs could act at the post-transcriptional level of gene expression regulation in different ways. Characterization of the post-transcriptional functions of lncRNAs is still being performed [40]. In human carcinoma HeLa cells, lincRNA-p21 was shown to repress the translation of mRNAs

encoding β -catenin and JunB by partial base-pairing and recruitment of translation repressor proteins [41]. During cell differentiation, the lncRNA TINCR interacts with the STAU-1 protein to form a complex that mediates the stabilization of differentiation mRNAs [42]. Additionally, linc-31 is required to stabilize the translational activator YB-1 and promote ROCK1 protein synthesis, which controls the proliferation to differentiation switch in myoblasts [43]. Recently, the lncRNA Airn, which is known to induce the imprinting of genes on its locus [44], was shown to also act in the cardiomyocyte cytoplasm. A spliced Airn isoform was found bound to the Igf2bp2 transcript and to be involved in the translation control of Igf2bp2 and a number of other genes [45].

Technological advances, including RNA-seq and computational analyses, have allowed improved characterization of non-coding RNAs [32]. These dynamic molecules seem to contribute to the complexity of organisms because non-coding sequences are far more numerous than protein-coding sequences in humans compared to the worm or fruit fly [32,46]. In this report, we characterized the dynamic expression patterns of lncRNAs during cardiomyogenesis and identified a stage-specific pattern expression. LncRNAs expressed in differentiating cells are likely to be involved in cell specialization pathways, while cardiomyocyte lncRNAs might play roles in maturation and cell-fate maintenance. Because RNAs can be rapidly transcribed and degraded, they represent a type of molecule that is perfectly suitable for regulatory mechanisms. In addition, the sequence and structure complexity of lncRNAs allow a plethora of ligand binding that can broaden their functions [31]. Despite some advances, most lncRNAs remain uncharacterized, and it will be important to investigate their mechanisms of action, targets and gene networks in the future.

Taken together, our findings support a role for lncRNAs in the temporal cell fate transition of pluripotent stem cells into cardiac precursor cells and cardiomyocytes.

Material and Methods

Cell culture and cardiomyocyte differentiation

NKX2-5eGFP/w HES3 cell lineage was donated by Monash University (Victoria, Australia) [43] and cultured as previously described [29]. hESC were submitted to a cardiac differentiation protocol adapted from [27,28] and detailed described in [29]. Briefly, cells were mildly dissociated to form small clusters and cultured in 6-well Ultralow Attachment Corning plates using StemPro-34 (Invitrogen), supplemented with

100 µg/ml penicillin-streptomycin, 2 mM L-glutamine, 150 µg/mL transferrin, 50 µg/mL ascorbic acid and 0.45 mM monothioglycerol (MTG) and kept in humid incubator at 37°C, 5% CO₂ and 5% O₂. For embryoid bodies (EBs) aggregation (D0-D1), the basal medium was supplemented with 1 ng/mL BMP4. On D1, the medium was replaced with supplementation of 10 ng/mL BMP4, 6 ng/mL Activin A and 5 ng/mL bFGF; on D4 with 10 µg/mL VEGF and 10 µM XAV 939, and starting on D8, the medium was replaced every two or three days supplemented with 10 µg/mL VEGF and 1 ng/mL BMP4. Three independent differentiation assays were used as experimental replicates.

Samples and sequencing

As previously described [29], monolayer hESC on D0 or differentiating EBs (on D1, D4, D9 and D15) were treated with 0.1 mg/mL cycloheximide and cell lysates were loaded onto 10% to 50% sucrose gradients and centrifuged at 150000 x g for 160 minutes at 4°C. Different sucrose gradient fractions were separated using ISCO gradient fractionation system (ISCO Model 160 Gradient Former Foxy Jr. Fraction Collector). Ribosome-free and polysome-bound fractions were isolated and used to prepare cDNA libraries. RNA-seq was carried out in an Illumina HiSeq platform.

Analysis of sequencing data

Reads were mapped against the reference genome GRCh38 with Rsubreads and features were counted using the function featureCounts as stated [29]. Biotypes were annotated using Ensembl database (Ensembl Genes 78). Only lncRNAs and antisense transcripts with at least one CPM in at least three samples were considered for further analyses.

K-means clustering analysis was made on the matrix of CPMs, with k equals 7. Expression profiles were visualized and genes belonging to the clusters further analyzed separately.

Pathway analysis of the lncRNAs in each RNA fraction (polysomal and free) was performed using the topGene Suite [44].

Polysome/ribosome-free ratio was determined by dividing RPKM values derived from the polysomal fraction by the ones derived from the ribosome-free fraction (three biological replicates were averaged).

Figures

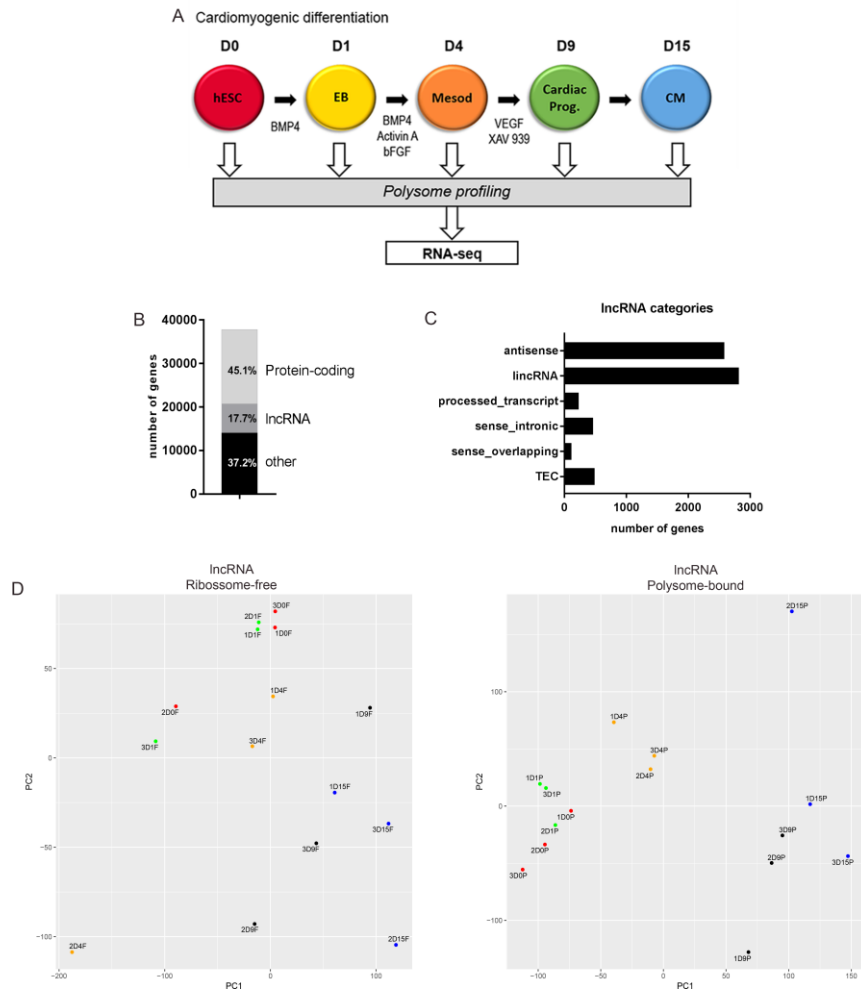


Figure 1. LncRNAs along human cardiomyogenesis. (A) Schematic representation of cardiomyogenic differentiation protocol indicating developmental stages and timing of sample collection. Polysome profiling was performed at each time-point indicated and Ribosome-free and Polysome-bound RNAs were sequenced [29] ($n=3$). (B) Number of genes identified at the five time-points analyzed and at ribosome-free and polysome-bound fractions ($RPKM>0.01$), classified according to protein-coding, long non-coding (lncRNA) and other RNA biotypes. (C) Number of genes identified in each lncRNA category (Ensembl). (D) Principal component analysis (PCA) of ribosome-free and polysome-bound samples at days D0, D1, D4, D9 and D15 ($n=3$).

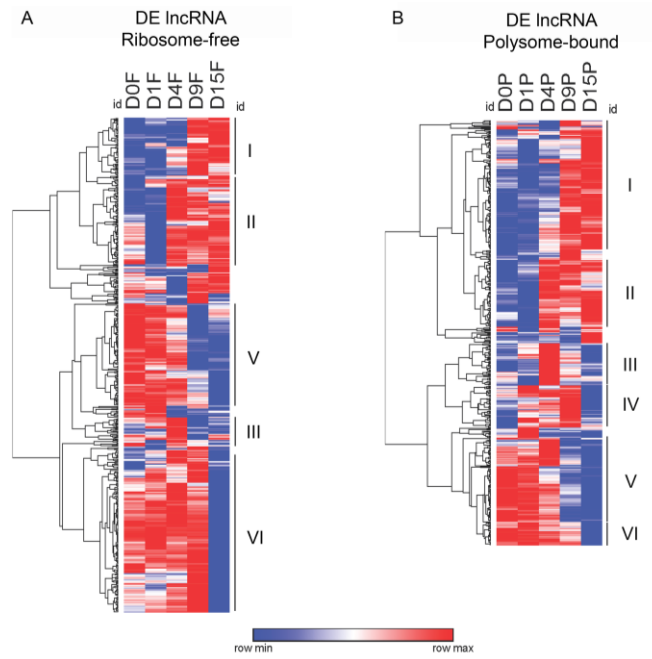


Figure 2. Differentially expressed lncRNAs during cardiomyogenic commitment . Heatmap of DE lncRNAs along hESC cardiomyogenesis in (A) ribosome-free and (B) polysome-bound fractions (each differentiation time-point compared to the preceding time-point, RPKM>0.01, FDR<0.05, $-2 > \log FC > 2$). I-VI groups established according to expression pattern. Heatmap made using <https://software.broadinstitute.org/morpheus/>.

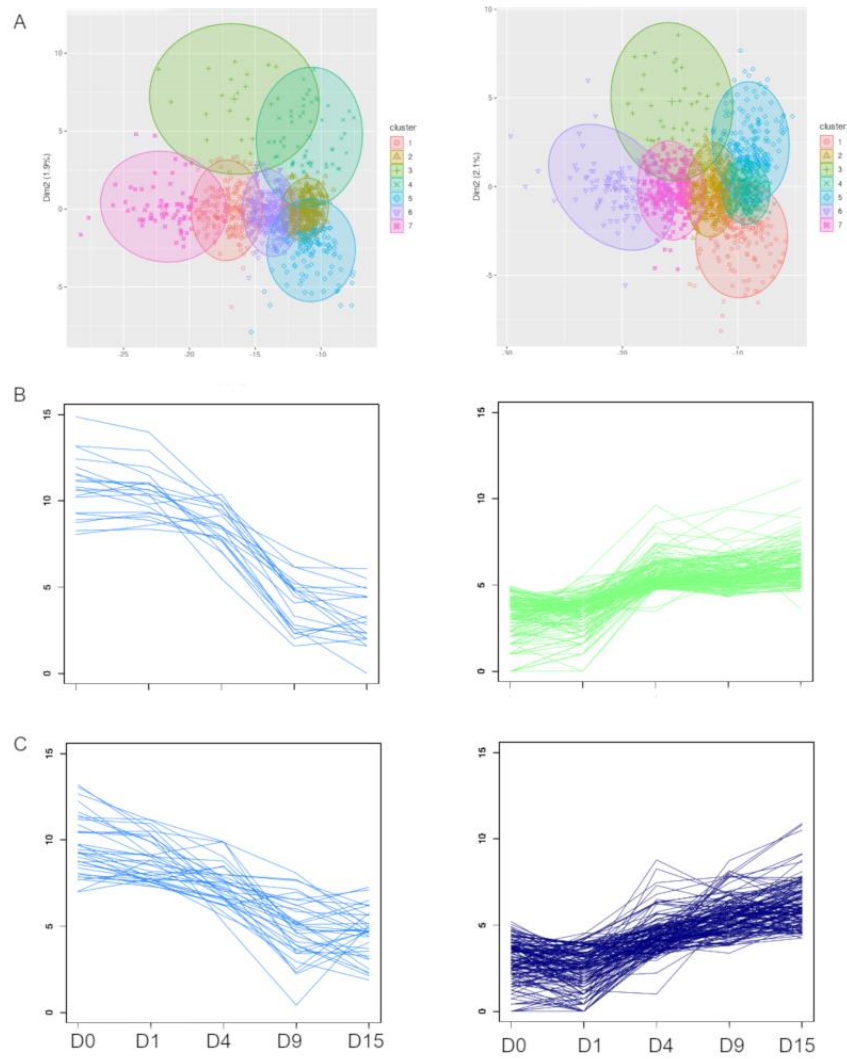


Figure 3. Gene expression pattern clusters of ribosome-free and polysome-bound lncRNAs during cardiomyogenesis. (A) PCA analysis of lncRNAs, identifying 7 distinct clusters (marked with an ellipse). Examples of clusters of polysome-bound (B) and ribosome-free (C) lncRNAs. On the left panel, gene expression is high at initial steps of cardiomyogenic differentiation (D0, D1, D4) with a strong decay afterwards. On the right, genes with low expression at initial steps and stabilization into the final stages.

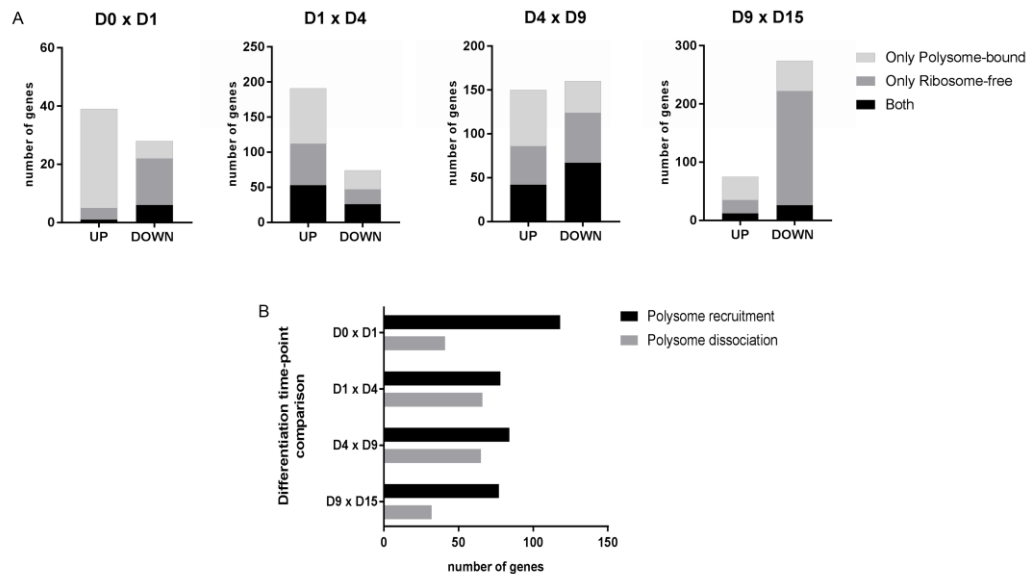
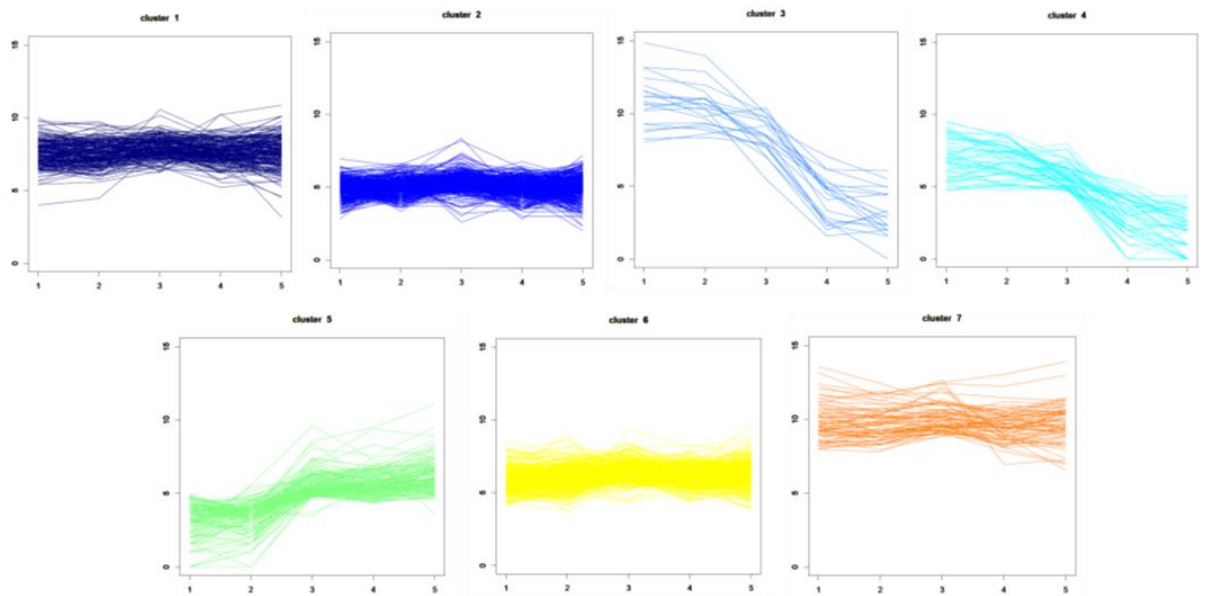


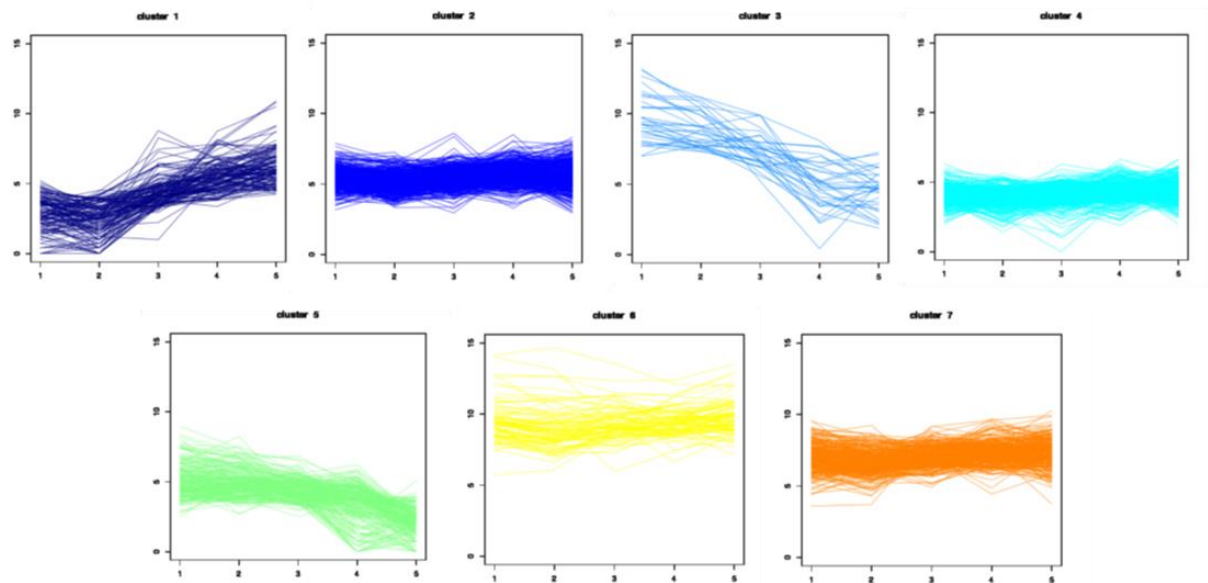
Figure 4. LncRNAs association with polysomes during cardiomyogenesis. (A) LncRNAs differentially expressed (RPKM>0.01, FDR<0.05, $-2 > \log_{2}FC > 2$) only in polysome-bound, in ribosome-free or in both fractions (coordinated). (B) Number of non-DE lncRNA recruited to and dissociated from polysomes (ratio poly/ribo-free, FDR<0.05, $-2 > \log_{2}FC > 2$).

Supplementary Table 1. Pathway analysis of LncRNAs in each RNA fraction. Free is on top with the name of the enriched Pathway term as the first column, second the source from which the information was taken, third FDR, fourth the number of genes from the input that was found in the database, and fifth the genes name. Polysomal fraction is on the bottom.

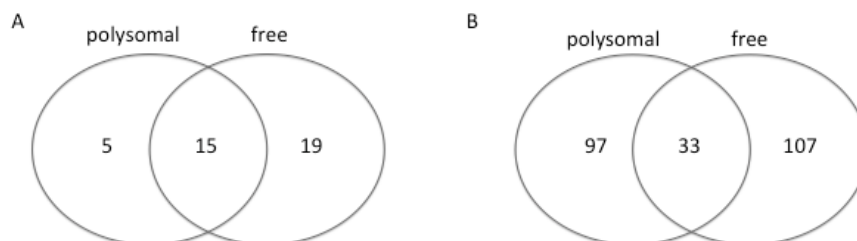
RIBOSOME-FREE RNA				
Name	Source	FDR	#genes	Gene
dermatan sulfate biosynthesis (late stages)	BIOCYC	0,015	<u>1</u>	DSE
Dermatan sulfate biosynthesis	REACTOME	0,015	<u>1</u>	DSE
dermatan sulfate biosynthesis	BIOCYC	0,015	<u>1</u>	DSE
Digestion of dietary lipid	REACTOME	0,015	<u>1</u>	LMF1
Glycosaminoglycan biosynthesis - chondroitin sulfate / dermatan sulfate	KEGG	0,015	<u>1</u>	DSE
Chondroitin sulfate/dermatan sulfate metabolism	REACTOME	0,033	<u>1</u>	DSE
POLYSOME-BOUND RNA				
Digestion of dietary lipid	REACTOME	0,022	<u>1</u>	LMF1
Heterotrimeric G-protein signaling pathway-Gq alpha and Go alpha mediated pathway	PantherDB	0,048	<u>1</u>	RGS5
Lipid digestion, mobilization, and transport	REACTOME	0,048	<u>1</u>	LMF1
Heterotrimeric G-protein signaling pathway-Gi alpha and Gs alpha mediated pathway	PantherDB	0,048	<u>1</u>	RGS5



Supplementary Figure 1. Gene expression profiles of the 7 clusters for polysome-bound lncRNAs.



Supplementary Figure 2. Gene expression profiles of the 7 clusters for ribosome-free lncRNAs.



Supplementary Figure 3. Venn diagrams showing specific and common lncRNAs from (A) Figure 3 left panels and (B) Figure 3 right panels.

References

1. Pircher, A.; Gebetsberger, J.; Polacek, N. Ribosome-associated ncRNAs: An emerging class of translation regulators. *RNA Biol.* 2014, 11, 1335–1339, doi:10.1080/15476286.2014.996459.
2. Ulitsky, I.; Bartel, D.P. LincRNAs: Genomics, evolution, and mechanisms. *Cell* 2013, 154, 26–46, doi:10.1016/j.cell.2013.06.020.
3. Lopez-Pajares, V. Long non-coding RNA regulation of gene expression during differentiation. *Pflugers Arch. Eur. J. Physiol.* 2016, 468, 971–981, doi:10.1007/s00424-016-1809-6.
4. Cabili, M.N.; Trapnell, C.; Goff, L.; Koziol, M.; Tazon-Vega, B.; Regev, A.; Rinn, J.L. Integrative annotation of human large intergenic noncoding RNAs reveals global properties and specific subclasses. *Genes Dev.* 2011, 25, 1915–1927, doi:10.1101/gad.17446611.
5. Schmitz, S.U.; Grote, P.; Herrmann, B.G. Mechanisms of long noncoding RNA function in development and disease. *Cell. Mol. Life Sci.* 2016, 73, 2491–2509, doi:10.1007/s00018-016-2174-5.
6. Bunch, H. Gene regulation of mammalian long non-coding RNA. *Mol. Genet. Genomics* 2018, 293, 1–15, doi:10.1007/s00438-017-1370-9.
7. Angrand, P.O.; Vennin, C.; Le Bourhis, X.; Adriaenssens, E. The role of long non-coding RNAs in genome formatting and expression. *Front. Genet.* 2015, 6, 1–12, doi:10.3389/fgene.2015.00165.
8. Hezroni, H.; Koppstein, D.; Schwartz, M.G.; Avrutin, A.; Bartel, D.P.; Ulitsky, I. Principles of Long Noncoding RNA Evolution Derived from Direct Comparison of Transcriptomes in 17 Species. *Cell Rep.* 2015, 11, 1110–1122, doi:10.1016/j.celrep.2015.04.023.
9. Pang, K.C.; Frith, M.C.; Mattick, J.S. Rapid evolution of noncoding RNAs: Lack of conservation does not mean lack of function. *Trends Genet.* 2006, 22, 1–5, doi:10.1016/j.tig.2005.10.003.
10. Johnsson, P.; Lipovich, L.; Grandér, D.; Morris, K. V. Evolutionary conservation of long non-coding RNAs; Sequence, structure, function. *Biochim. Biophys. Acta - Gen. Subj.* 2014, 1840, 1063–1071, doi:10.1016/j.bbagen.2013.10.035.
11. Karapetyan, A.R.; Buiting, C.; Kuiper, R.A.; Coolen, M.W. Regulatory roles for long ncRNA and mRNA. *Cancers (Basel)*. 2013, 5, 462–490, doi:10.3390/cancers5020462.
12. Li, R.; Zhu, H.; Luo, Y. Understanding the functions of long non-coding RNAs through their higher-order structures. *Int. J. Mol. Sci.* 2016, 17, doi:10.3390/ijms17050702.
13. Carlevaro-Fita, J.; Rahim, A.; Guigó, R.; Vardy, L.A.; Johnson, R. Cytoplasmic long noncoding RNAs are frequently bound to and degraded at ribosomes in human cells. *Rna* 2016, 22, 867–882, doi:10.1261/rna.053561.115.
14. Dallagiovanna, B.; Pereira, I.T.; Origa-Alves, A.C.; Shigunov, P.; Naya, H.; Spangenberg, L. lncRNAs are associated with polysomes during adipose-derived stem cell differentiation. *Gene* 2017, 610, 103–111, doi:10.1016/j.gene.2017.02.004.
15. Ingolia, N.T.; Lareau, L.F.; Weissman, J.S. Ribosome profiling of mouse embryonic stem cells reveals the complexity and dynamics of mammalian proteomes. *Cell* 2011, 147, 789–802, doi:10.1016/j.cell.2011.10.002.
16. van Heesch, S.; van Iterson, M.; Jacobi, J.; Boymans, S.; Essers, P.B.; de Bruijn, E.; Hao, W.; MacInnes, A.W.; Cuppen, E.; Simonis, M. Extensive localization of long noncoding RNAs to the cytosol and mono- and polyribosomal complexes. *Genome Biol.* 2014, 15, R6, doi:10.1186/gb-2014-15-1-r6.
17. Ruiz-Orera, J.; Messeguer, X.; Subirana, J.A.; Alba, M.M. Long non-coding RNAs as a source of new peptides. *Elife* 2014, 3, 1–24, doi:10.7554/eLife.03523.
18. Guttman, M.; Russell, P.; Ingolia, N.T.; Weissman, J.S.; Lander, E.S. Ribosome profiling provides evidence that large noncoding RNAs do not encode proteins. *Cell* 2013, 154, 240–251, doi:10.1016/j.cell.2013.06.009.
19. Wamstad, J.A.; Alexander, J.M.; Truty, R.M.; Shrikumar, A.; Li, F.; Eilertson, K.E.; Ding, H.; Wylie, J.N.; Pico, A.R.; Capra, J.A.; et al. Dynamic and coordinated epigenetic regulation of developmental transitions in the cardiac lineage. *Cell* 2012, 151, 206–220, doi:10.1016/j.cell.2012.07.035.
20. Kurian, L.; Aguirre, A.; Sancho-Martinez, I.; Benner, C.; Hishida, T.; Nguyen, T.B.; Reddy, P.; Nivet, E.; Krause, M.N.; Nelles, D.A.; et al. Identification of novel long noncoding RNAs underlying

- vertebrate cardiovascular development. *Circulation* 2015, 131, 1278–1290, doi:10.1161/CIRCULATIONAHA.114.013303.
21. Li, Y.; Lin, B.; Yang, L. Comparative Transcriptomic Analysis of Multiple Cardiovascular Fates from Embryonic Stem Cells Predicts Novel Regulators in Human Cardiogenesis. *Sci. Rep.* 2015, 5, doi:10.1038/srep09758.
 22. Li, Y.; Zhang, J.; Huo, C.; Ding, N.; Li, J.; Xiao, J.; Lin, X.; Cai, B.; Zhang, Y.; Xu, J. Dynamic Organization of lncRNA and Circular RNA Regulators Collectively Controlled Cardiac Differentiation in Humans. *EBioMedicine* 2017, 24, 137–146, doi:10.1016/j.ebiom.2017.09.015.
 23. Tompkins, J.D.; Jung, M.; Chen, C. yi; Lin, Z.; Ye, J.; Godatha, S.; Lizhar, E.; Wu, X.; Hsu, D.; Couture, L.A.; et al. Mapping Human Pluripotent-to-Cardiomyocyte Differentiation: Methyloomes, Transcriptomes, and Exon DNA Methylation “Memories.” *EBioMedicine* 2016, 4, 74–85, doi:10.1016/j.ebiom.2016.01.021.
 24. Rizki, G.; Boyer, L.A. Lnc ing epigenetic control of transcription to cardiovascular development and disease. *Circ. Res.* 2015, 117, 192–206, doi:10.1161/CIRCRESAHA.117.304156.
 25. Scheuermann, J.C.; Boyer, L.A. Getting to the heart of the matter: long non-coding RNAs in cardiac development and disease. *EMBO J.* 2013, 32, 1805–1816, doi:10.1038/emboj.2013.134.
 26. Murry, C.E.; Keller, G. Differentiation of Embryonic Stem Cells to Clinically Relevant Populations: Lessons from Embryonic Development. *Cell* 2008, 132, 661–680, doi:10.1016/j.cell.2008.02.008.
 27. Kattman, S.J.; Witty, A.D.; Gagliardi, M.; Dubois, N.C.; Niapour, M.; Hotta, A.; Ellis, J.; Keller, G. Stage-specific optimization of activin/nodal and BMP signaling promotes cardiac differentiation of mouse and human pluripotent stem cell lines. *Cell Stem Cell* 2011, 8, 228–240, doi:10.1016/j.stem.2010.12.008.
 28. Yang, L.; Soonpaa, M.H.; Adler, E.D.; Roepke, T.K.; Kattman, S.J.; Kennedy, M.; Henckaerts, E.; Bonham, K.; Abbott, G.W.; Linden, R.M.; et al. Human cardiovascular progenitor cells develop from a KDR+ embryonic-stem-cell-derived population. *Nature* 2008, 453, 524–528, doi:10.1038/nature06894.
 29. Pereira, I.T.; Spangenberg, L.; Robert, A.W.; Amorín, R.; Stimamiglio, M.A.; Naya, H.; Dallagiovanna, B. Polysome profiling followed by RNA-seq of cardiac differentiation stages in hESCs. *Sci. data* 2018, 5, 1–11, doi:10.1038/sdata.2018.287.
 30. Luo, S.; Lu, J.Y.; Liu, L.; Yin, Y.; Chen, C.; Han, X.; Wu, B.; Xu, R.; Liu, W.; Yan, P.; et al. Divergent lncRNAs regulate gene expression and lineage differentiation in pluripotent cells. *Cell Stem Cell* 2016, 18, 637–652, doi:10.1016/j.stem.2016.01.024.
 31. Geisler, S.; Collier, J. RNA in unexpected places: long non-coding RNA functions in diverse cellular contexts. *Mol. Cell Biol. Rev.* 2013, 14, 699–712, doi:10.1038/nrm3679.
 32. Devaux, Y.; Zangrando, J.; Schroen, B.; Creemers, E.E.; Pedrazzini, T.; Chang, C.P.; Dorn, G.W.; Thum, T.; Heymans, S. Long noncoding RNAs in cardiac development and ageing. *Nat. Rev. Cardiol.* 2015, 12, 415–425, doi:10.1038/nrcardio.2015.55.
 33. Klattenhoff, C.A.; Scheuermann, J.C.; Surface, L.E.; Bradley, R.K.; Fields, P.A.; Steinhauser, M.L.; Ding, H.; Butty, V.L.; Torrey, L.; Haas, S.; et al. Braveheart, a long noncoding RNA required for cardiovascular lineage commitment. *Cell* 2013, 152, 570–583, doi:10.1016/j.cell.2013.01.003.
 34. Grote, P.; Wittler, L.; Hendrix, D.; Koch, F.; Währisch, S.; Beisaw, A.; Macura, K.; Bläss, G.; Kellis, M.; Werber, M.; et al. The Tissue-Specific lncRNA Fendrr Is an Essential Regulator of Heart and Body Wall Development in the Mouse. *Dev. Cell* 2013, 24, 206–214, doi:10.1016/j.devcel.2012.12.012.
 35. Ounzain, S.; Micheletti, R.; Arnan, C.; Plaisance, I.; Cecchi, D.; Schroen, B.; Reverter, F.; Alexanian, M.; Gonzales, C.; Ng, S.Y.; et al. CARMEN, a human super enhancer-associated long noncoding RNA controlling cardiac specification, differentiation and homeostasis. *J. Mol. Cell. Cardiol.* 2015, 89, 98–112, doi:10.1016/j.yjmcc.2015.09.016.
 36. Guo, X.; Xu, Y.; Wang, Z.; Wu, Y.; Chen, J.; Wang, G.; Lu, C.; Jia, W.; Xi, J.; Zhu, S.; et al. A Linc1405/Eomes Complex Promotes Cardiac Mesoderm Specification and Cardiogenesis. *Cell Stem Cell* 2018, 22, 893–908.e6, doi:10.1016/j.stem.2018.04.013.
 37. Faye, M.D.; Graber, T.E.; Holcik, M. Assessment of Selective mRNA Translation in Mammalian Cells by Polysome Profiling. *J. Vis. Exp.* 2014, 1–8, doi:10.3791/52295.

38. Panda, A.C.; Martindale, J.L.; Gorospe, M. Polysome Fractionation to Analyze mRNA Distribution Profiles. *Bio Protoc.* 2017, 5, doi:10.1038/nrm3228.All.
39. Bazzini, A.A.; Johnstone, T.G.; Christiano, R.; MacKowiak, S.D.; Obermayer, B.; Fleming, E.S.; Vejnar, C.E.; Lee, M.T.; Rajewsky, N.; Walther, T.C.; et al. Identification of small ORFs in vertebrates using ribosome footprinting and evolutionary conservation. *EMBO J.* 2014, 33, 981–993, doi:10.1002/embj.201488411.
40. Yoon, J.-H.; Abdelmohsen, K.; Gorospe, M. Posttranscriptional Gene Regulation by Long Noncoding RNA. *J. Mol. Biol.* 2013, 425, 3723–3730, doi:10.1016/j.jmb.2012.11.024.
41. Yoon, J.H.; Abdelmohsen, K.; Srikantan, S.; Yang, X.; Martindale, J.L.; De, S.; Huarte, M.; Zhan, M.; Becker, K.G.; Gorospe, M. LincRNA-p21 Suppresses Target mRNA Translation. *Mol. Cell* 2012, 47, 648–655, doi:10.1016/j.molcel.2012.06.027.
42. Kretz, M.; Siprashvili, Z.; Chu, C.; Webster, D.E.; Zehnder, A.; Qu, K.; Lee, C.S.; Flockhart, R.J.; Groff, A.F.; Chow, J.; et al. Control of somatic tissue differentiation by the long non-coding RNA TINCR. *Nature* 2013, 493, 231–235, doi:10.1038/nature11661.
43. Dimartino, D.; Colantoni, A.; Ballarino, M.; Martone, J.; Mariani, D.; Danner, J.; Bruckmann, A.; Meister, G.; Morlando, M.; Bozzoni, I. The Long Non-coding RNA Inc-31 Interacts with Rock1 mRNA and Mediates Its YB-1-Dependent Translation. *Cell Rep.* 2018, 23, 733–740, doi:10.1016/j.celrep.2018.03.101.
44. Latos, P.A.; Pauler, F.M.; Koerner, M. V; Ba, H.; Hudson, Q.J.; Stocsits, R.R.; Allhoff, W.; Stricker, S.H.; Klement, R.M.; Warczok, K.E.; et al. Airn Transcriptional Overlap, But Not Its lncRNA Products, Induces Imprinted Igf2r Silencing. *Science* (80-). 2012, 1469–1473.
45. Hosen, M.R.; Militello, G.; Weirick, T.M.; Ponomareva, Y.; Dassanayaka, S.; Moore, J.B.; Doering, C.; Wysoczynski, M.; Jones, S.P.; Dimmeler, S.; et al. Airn Regulates Igf2bp2 Translation in Cardiomyocytes. *Circ. Res.* 2018, 122, 1347–1353, doi:10.1161/CIRCRESAHA.117.312215.
46. Frith, M.C.; Pheasant, M.; Mattick, J.S. The amazing complexity of the human transcriptome. *Eur. J. Hum. Genet.* 2005, 13, 894–897, doi:10.1038/sj.ejhg.5201459.
47. Elliott, D.A.; Braam, S.R.; Koutsis, K.; Ng, E.S.; Jenny, R.; Lagerqvist, E.L.; Biben, C.; Hatzistavrou, T.; Hirst, C.E.; Yu, Q.C.; et al. NKX2-5eGFP/w hESCs for isolation of human cardiac progenitors and cardiomyocytes. *Nat. Methods* 2011, 8, 1037–1040, doi:10.1038/nmeth.1740.
48. Chen, J.; Bardes, E.E.; Aronow, B.J.; Jegga, A.G. ToppGene Suite for gene list enrichment analysis and candidate gene prioritization. *Nucleic Acids Res.* 2009, 37, 305–311, doi:10.1093/nar/gkp427.

CAPÍTULO V

Prefácio

Neste capítulo, estão descritos a metodologia e os resultados referentes a análise funcional do lncRNA *LINC00890*. Células-tronco pluripotentes foram modificadas a fim de expressar o *LINC00890* sob o controle de doxíciclina ou ter o *LINC00890* silenciado, para avaliação do seu papel durante a diferenciação cardiomiogênica. A expressão de *LINC00890* durante a diferenciação cardíaca se mostrou interessante, colocando-o como candidato a participar deste processo. Por isso, nos propomos a investigar a hipótese da funcionalidade de *LINC00890* na diferenciação cardíaca de células-tronco humanas e este capítulo foi incluído.

INTRODUÇÃO

Na última década, trabalhos vêm demonstrando o potencial dos lncRNAs em participar da regulação da expressão gênica em diferentes níveis, incluindo a organização da cromatina, a regulação transcricional e também em processos pós-transcricionais (ANGRAND *et al.*, 2015). Além disso, os lncRNAs podem contribuir para a especificação da identidade celular (revisado por LOPEZ-PAJARES, 2016), incluindo a diferenciação cardiomiogênica.

Os dados de RNA-seq descritos no Capítulo I também possibilitaram a análise de lncRNAs expressos nos diferentes estágios da cardiomiogênese de CTP (Capítulo IV). A fim de melhor investigar a influência dos lncRNAs na diferenciação celular e buscar desvendar funções específicas de lncRNAs individuais, o gene *LINC00890* foi escolhido para avaliação funcional na cardiomiogênese.

O lncRNA *LINC00890* se apresentou expresso em progenitores cardíacos (D9) e teve a expressão reduzida em cardiomiócitos (D15). Não há nenhuma informação na literatura com relação a sua função, apenas dados de expressão em diferentes tecidos adultos humanos, os quais mostram a expressão de *LINC00890* em endométrio e próstata (FAGERBERG *et al.*, 2014). O *LINC00890* apareceu associado aos polissomos nos nossos dados de RNA-seq, indicando um possível papel na regulação da tradução de mRNAs ou até mesmo como codificante de peptídeos.

A estratégia para avaliação funcional do *LINC00890* foi o estabelecimento de linhagens com super expressão induzida ou silenciamento e submissão dessas células à diferenciação cardiomiogênica. O sistema de expressão induzida escolhido foi um sistema integrativo que utiliza o vetor AAVS1-TRE-GW-rtTA, e possibilita a expressão controlada através da indução com doxiciclina. Para o silenciamento, a técnica de CRISPR/Cas9 foi utilizada para a deleção completa do gene.

6 METODOLOGIA

6.1 Construção de linhagens celulares com super expressão induzida de *LINC00890*

6.1.1 Clonagem do gene *LINC00890* no vetor AAVS1-GW-TRE-rtTA

O vetor AAVS1-GW-TRE-rtTA também é um vetor de integração sítio-específica no locus AAVS1 e possui o sistema de expressão induzida por doxiciclina, conforme descrito anteriormente (item 4.7.1, Capítulo III). As diferenças para o AAVS1-TRE-FLAG-rtTA são a ausência da etiqueta FLAG e o sistema de entrada por Gateway (GW, Thermo Fischer) (Figura 6.1).

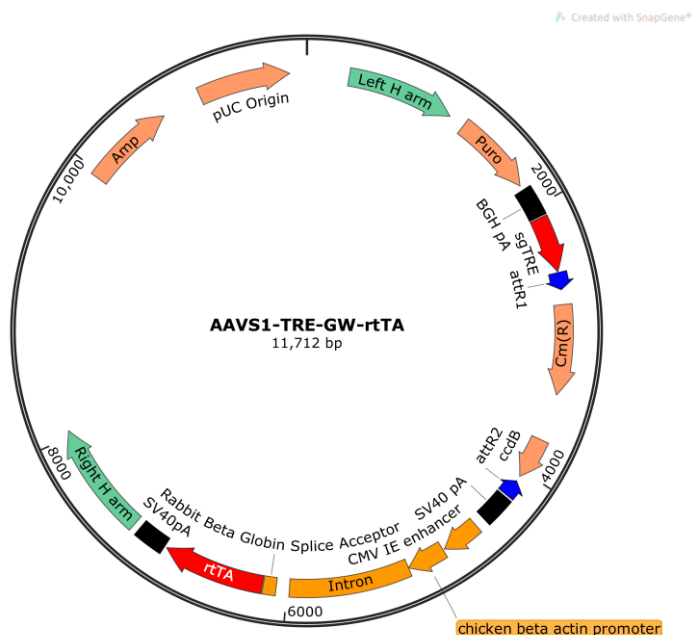


Figura 6.1 Mapa do vetor AAVS1-TRE-GW-rtTA. Esse vetor possui braços de homologia com o locus AAVS1 do genoma humano (*Left H arm*, *Right H arm*, em verde), sítio TRE (vermelho) no promotor do transgene que será inserido entre os sítios attR1 e attR2 (azul), e expressão constitutiva de rtTA (vermelho) e gene de resistência a puromicina (laranja). Mapa criado com SnapGene.

A sequência do gene *LINC00890* foi obtida e anotada pelo grupo liderado pelo Dr. S. Weimann do German Cancer Research Center (Alemanha), que gentilmente nos cedeu o clone de cDNA DKFZp686D0853. O DNA plasmidial do clone citado foi utilizado para amplificação da sequência do *LINC00890* utilizando oligos desenhados de maneira que incluíssem os sítios attB (GW) no fragmento amplificado (ANEXO I - Tabela 1). Uma vez que o *LINC00890* é um lncRNA, a estratégia de clonagem abrangeu o RNA transcrito todo. O fragmento amplificado foi de 4.5 kb, que compreende os 3 éxons do gene. A amplificação foi feita por PCR usando o kit da enzima LA Taq (Takara) com 500 ng de DNA plasmidial e 0.8 μ M de cada oligo (*forward* e *reverso*). A ciclagem foi feita com 95 °C por 5 min de desnaturação inicial; 95 °C por 30 seg, 50 °C por 30 seg, 72 °C por 5 min (7X) para amplificação; e 72 °C por 10 min para extensão final. Depois da PCR e da confirmação do tamanho do amplicon em gel de agarose, a amostra foi incubada com a enzima de restrição Dnpi por 15 min a 37 °C e, então, purificada com o kit Wizard Gel & PCR Clean-up (Promega).

Seguiu-se com a inserção do fragmento amplificado no vetor pDONOR 221 (ANEXO I – Figura 1), através da reação: 150 ng do produto de PCR, 150 ng do vetor pDONOR 221, 2 μ l da enzima BP clonase II (Invitrogen), com incubação por 16 h a

temperatura ambiente (TA). A reação foi parada com a adição de 1 µl de proteinase K e incubação por 15 min a 37 °C. O produto da reação foi usado para transformar bactérias, conforme descrito anteriormente (item 4.7.2, Capítulo III). Colônias isoladas tiveram o DNA plasmidial isolado, o qual foi submetido à confirmação por reação de restrição e sequenciamento de Sanger. O próximo passo foi a transferência do transgene para o vetor final AAVS1-TRE-GW-rtTA, através da reação: 150 ng de pDONOR+*LINC00890*, 150 ng de AAVS1-TRE-GW-rtTA, 2 µl da enzima LR (Invitrogen), com incubação por 16 h a TA. A reação foi parada com a adição de 1 µl de proteinase K e incubação por 15 min a 37 °C. O produto da reação foi usado para transformar bactérias, conforme descrito anteriormente. Colônias isoladas tiveram o DNA plasmidial isolado, o qual foi submetido à confirmação por reação de restrição.

6.1.2 Clonagem do gene *LINC00890* com mutação no vetor AAVS1-GW-TRE-rtTA

Com o intuito de avaliar a função da CDS predita na sequência do lncRNA *LINC00890*, foi planejado a construção de uma linhagem de células com expressão induzida desse gene mutado. A mutação planejada se caracterizou pela inserção de dois nucleotídeos na região próxima ao ATG, a fim de mudar a janela de leitura (Figura 6.2). Foram desenhados oligos *forward* e reverso para a mesma porção da sequência, ambos incluindo os nucleotídeos adicionais (ANEXO I - Tabela 2). Dessa maneira, durante a PCR, a mutação seria inserida na sequência amplificada final. Em virtude da sequência do *LINC00890* ser grande (4,5 kb), foram planejadas diferentes PCRs para que se conseguisse amplificar a sequência por completo, incluindo a mutação (Figura 6.2). Todos os primers estão disponíveis no ANEXO I - Tabelas 1 e 2. O produto final foi a sequência do *LINC00890* com a mutação inserida, e os sítios attB incluídos para seguir com a clonagem com Gateway, conforme descrito anteriormente.

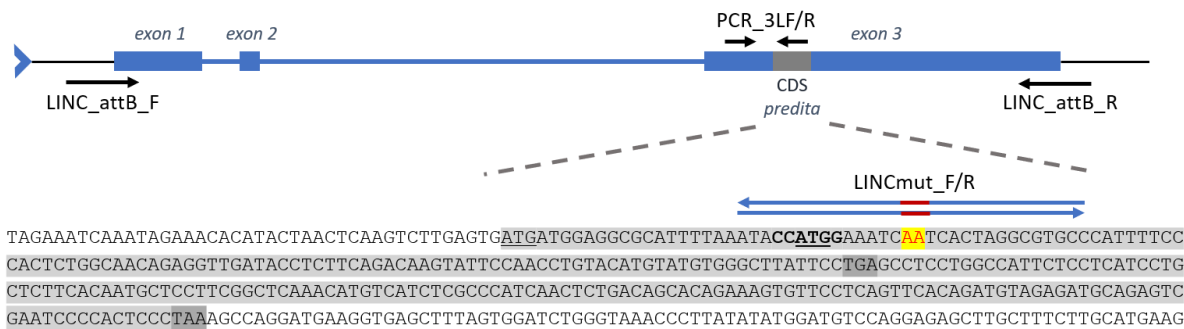


Figura 6.2 Estratégia de mutação na CDS predita para o gene *LINC00890*. Existem duas possíveis CDS (grifadas juntas em cinza claro): os códons de início (ATG) e de parada estão destacados em sublinhado e cinza escuro, respectivamente. A sequência Kozak está destacada em negrito. Através de PCR usando oligos mutados, dois nucleotídeos AA foram inseridos na sequência, conforme destacado em vermelho e amarelo. Os oligos usados nas PCR estão ilustrados na figura. Fonte: o autor.

6.1.3 Transfecção dos vetores AAVS1 em células-tronco pluripotentes humanas e seleção de clones

A linhagem de CTE humanas H1 e as iPSCs humanas PLZ foram transfectadas com os vetores AAVS1 construídos para a integração do sistema de expressão induzida. A transfecção foi feita conforme descrito anteriormente no item 4.7.4, Capítulo III. Após transfecção, as células foram submetidas à seleção pela resistência a puromicina e seleção clonal, conforme descrito no item 4.7.5, Capítulo III. Os clones foram, então, submetidos (+DOX) ou não (-DOX) à indução com 500 ng/ml de doxiciclina por 24h para avaliar a expressão de *LINC00890* por qPCR.

6.2 Silenciamento do gene *LINC00890* usando CRISPR/Cas9

6.2.1 Montagem do sistema Cas9/gRNA no vetor PX458

A estratégia de silenciamento escolhida para o gene *LINC00890* foi a de deleção completa do gene, utilizando um par de gRNAs (RNAs guias) (Figura 6.3). Para isso, quatro gRNAs foram desenhados para que reconhecessem as regiões adjacentes ao gene utilizando a plataforma online do Zhang lab (<http://crispr.mit.edu/>, acesso em julho de 2017) (chamados de A, B, C e D), e testados quanto sua eficiência. Os oligos para construção do sistema de expressão de Cas9/gRNAs foram desenhados de acordo com as instruções de clonagem no vetor pSpCas9(BB)-2A-RFP (PX458 – Addgene, plasmid #48138) (ANEXO II – Figura 1 e Tabela 1).

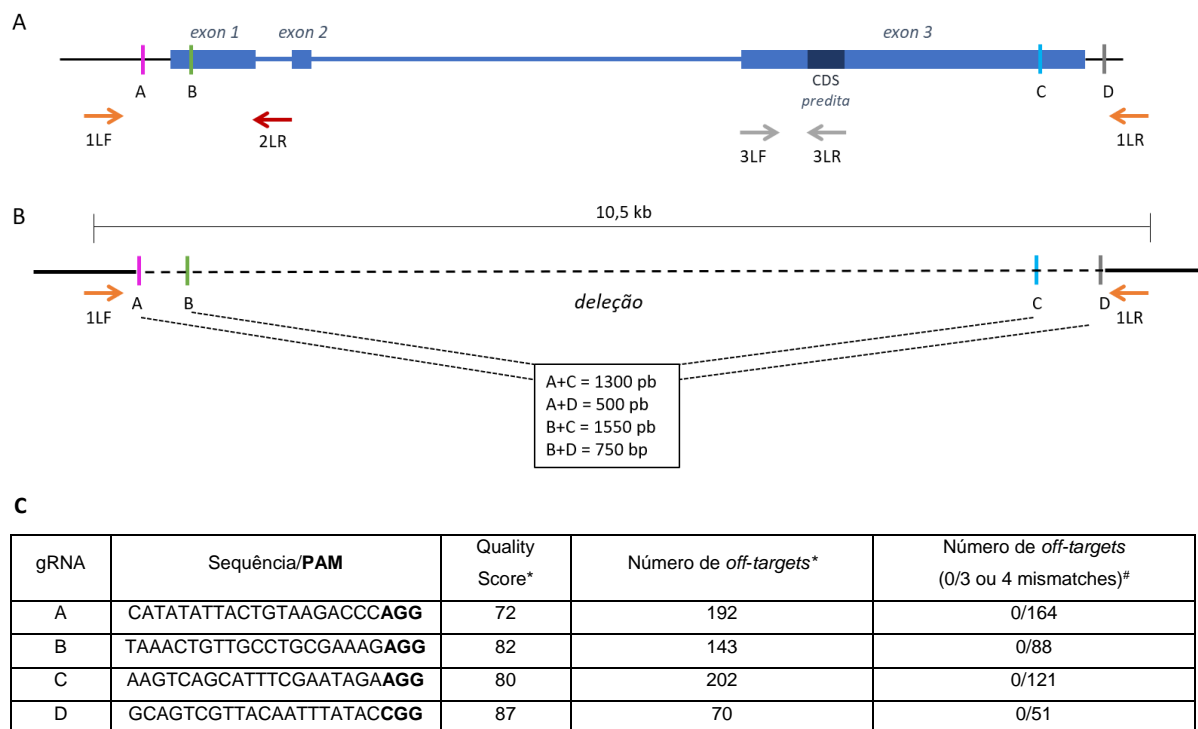


Figura 6.3 Estratégia de silenciamento do gene *LINC00890* através de deleção utilizando um par de gRNAs (dentre os gRNAs A, B, C e D). A deleção foi verificada através de PCR usando os oligos ilustrados na figura. (A) Posicionamento dos gRNAs e oligos no gene *LINC00890*. (B) Esquema da PCR de verificação da deleção, ilustrando os tamanhos esperados dos amplicons quando da deleção usando os diferentes pares de gRNAs. (C) Resumo dos parâmetros dos gRNAs testados, *segundo a plataforma online Zhang Lab, e *off-targets* (alvos inespecíficos) #segundo plataforma Cas-OFFinder (<http://www.rgenome.net/cas-offinder/>). Fonte: o autor.

Os oligos listados na Tabela 1 – ANEXO II foram anelados e fosforilados na reação: 100 μ M oligo F, 100 μ M oligo R, 1 μ l tampão 10X T4, 0,5 μ l de enzima T4 PNK (NEB); termociclada a 37 $^{\circ}$ C por 30 min, 95 $^{\circ}$ C por 5 min e temperatura reduzida a 25 $^{\circ}$ C (5 $^{\circ}$ C/min). Após a digestão do vetor PX458 com a enzima BbsI e purificação em gel de agarose, foi feita a reação de ligação com: 50 ng PX458 digerido, 1 μ l dos oligos anelados (diluídos a 1:250), 5 μ l do tampão Quick ligation e 1 μ l da enzima Quick ligase (NEB); incubada por 10 min a TA. A transformação de bactérias e extração de DNA plasmidial foram feitas conforme descrito anteriormente. A confirmação da inserção dos oligos/gRNAs foi feita por reação de restrição.

6.2.2 Transfecção dos vetores PX458 em células-tronco pluripotentes humanas

Células da linhagem de CTE humanas H1 e as iPSCs humanas PLZ foram transfectadas com os vetores PX458 construídos para expressão de Cas9 e gRNAs (RAN *et al.*, 2013). As células foram plaqueadas em baixa confluência (20-30%) em placas de 24 poços, sobre uma camada de Geltrex, com meio de cultivo TEsR E8 (Stem Cell Technologies). No dia seguinte, os reagentes de transfecção foram preparados: 100 µl de meio IMDM, 500 ng PX458 (1), 500 ng PX458 (2), 2 µl reagente vermelho e 2 µl reagente azul (kit de transfecção GenEln, Amsbio). O meio de cultivo das células foi trocado por meio fresco TEsR E8 e a solução preparada foi adicionada aos poços, seguindo-se incubação por 16 h. Após 24-48h, foi possível observar a expressão de RFP (proteína fluorescente vermelha) e, conseqüentemente, de fluorescência vermelha nas células transfectadas. As células foram mantidas em cultivo e expandidas até que se obtivesse um número suficiente de células para análise e seleção.

6.2.3 Avaliação e validação da deleção por PCR

Para testar se as células foram mutadas e, portanto, tiveram o gene *LINC00890* deletado (KO, do inglês *knockout*), PCRs foram feitas de duas maneiras. (1) Durante a passagem das células, cerca de 5 µl de suspensão de células foram incubados com 100 µl de água deionizada a 98 °C por 10 min. A reação de PCR foi preparada com: 6 µl de 5PRIME HotMasterMix (QuantaBio), 0,1 µM de cada oligo F+R e 1 µl da suspensão de células fervida. (2) O DNA genômico das células foi extraído usando o kit GeneJET Genomic DNA Purification (Thermo Fischer) e a reação de PCR preparada com: 50 ou 100 ng de DNA genômico, 0,8 µM de cada primer F+R e o kit da enzima LA Taq (Takara). Em ambos os casos, a ciclagem foi feita com 95 °C por 2 min de desnaturação inicial; 95 °C por 30 seg, 60 °C por 30 seg, 70 °C por 1 min e 30 seg (40X) para amplificação; e 70 °C por 10 min para extensão final. Os oligos usados estão listados na Tabela 6.1. Os produtos de PCR foram avaliados em gel de agarose 1% corado com brometo de etídeo.

Tabela 6.1 Oligos usados na PCR de avaliação de deleção do gene *LINC00890*. WT, *wild type* (do inglês, selvagem). Tamanho do amplicon considerando a deleção com gRNA A+D.

Nome do oligo	Sequência 5'-3'	Deteção	Tamanho do amplicon
PCR_1LF	5'GGGCCTTTGACTACAAATGGAT	WT	10,5 kb
PCR_1LR	5'GCAGAGGAATGTGGAAGGCT	Mutado (KO)	514 pb
PCR_2LR	5'GCTCGCTCAAGTAGGAGTTG	WT (íntron)	1050 pb (+1LF)
PCR_3LF	5'GCTGGGCAGGAACCTTACAA	WT (éxon 3)	780 pb
PCR_3LR	5'TGTTTGAGCCGAAGGAGCAT		

6.2.4 Seleção de clones com silenciamento gênico

Após a transfecção, as células foram passadas para uma nova placa quando atingiram 80% de confluência. Nessa passagem, uma primeira PCR de avaliação foi feita com suspensão de células. Se essa primeira população inicial total (bulk) fosse positiva, deu-se seguimento ao cultivo e seleção das células. Uma passagem para placa de 60 mm, sobre camada de Geltrex, em baixa confluência (~10%), foi feita com a intenção de cultivar células isoladas. Quando colônias estáveis se formaram (aprox. 10 dias), colônias isoladas foram pescadas e transferidas para uma nova placa de 24 poços, sob um microscópio invertido, dentro do fluxo laminar. Vinte e quatro clones (colônias) foram pescados e expandidos até que tivessem células suficientes para uma nova passagem e PCR de verificação. Clones negativos, ou seja, que não amplificaram o fragmento correspondente a deleção, foram descartados. Clones positivos, que amplificaram o fragmento correspondente a deleção (KO), mas que também amplificaram a sequência selvagem (WT), foram guardados e/ou submetidos a uma nova rodada de seleção clonal. Novas transfecções e seleção clonal foram feitas até se encontrar um clone positivo para a deleção (KO) e negativo para o selvagem (WT).

6.3 Ensaios de fisiologia de cardiomiócitos derivados de CTP

6.3.1 Quantificação da frequência de batimento de cardiomiócitos

Vídeos de cardiomiócitos diferenciados em monocamada foram gravados utilizando uma câmera digital acoplada a um microscópio óptico invertido, em sala

com temperatura controlada. A quantificação da frequência de batimento foi feita através da contagem de batimentos por minuto, nas diferentes condições analisadas.

6.3.2 Eletrofisiologia de cardiomiócitos usando Patch clamp

A técnica de Patch clamp *whole cell* (do inglês, célula completa) foi usada para medir a voltagem dos canais de íons dos cardiomiócitos diferenciados *in vitro*. Micropipetas de vidro (2-3 M Ω) foram preenchidas com uma solução contendo: 5 mM NaCl, 150 mM KCl, 2 mM CaCl₂, 10 mM HEPES, 5 mM EGTA e 5 mM MgATP (com pH ajustado para 7,2 usando KOH). A solução de banho consistiu de: 150 mM NaCl, 5,4 mM KCl, 1,8 mM CaCl₂, 1 mM MgCl₂, 15 mM HEPES, 15 mM glicose (com pH ajustado para 7,4 usando NaOH). Todas as medidas eletrofisiológicas foram realizadas no equipamento amplificador Axopatch 200B e conversor Axon Digidata 1320A A/D, através do sistema pCLAMP (quadros Digidata A/D e D/A e pCLAMP 9.2, Molecular Devices). Os dados foram analisados com Clampfit. As gravações foram filtradas com frequências de corte para evitar o *aliasing* e foram digitalizadas em velocidades pelo menos 5 vezes da frequência de corte, geralmente entre 2 e 10 kHz, respectivamente. Os potenciais de ação (APs) foram adquiridos sob pulsos de 3 ms de duração, 1,2 vezes da intensidade basal sob uma taxa de estimulação de 0.5 Hz. Os métodos análogo e P/4 foram usados para cancelamento de vazamento e capacidade transiente. A resistência em série foi parcialmente compensada pelo circuito de retorno. O ensaio de Patch clamp foi realizado em colaboração com o laboratório do Dr Samuel Dudley (Universidade de Minnesota, EUA).

7 RESULTADOS

7.1 Escolha de *LINC00890* para avaliação funcional

O RNA não-codificante longo transcrito pelo gene *LINC00890* foi escolhido para análise funcional durante a diferenciação cardiomiogênica. Ele se apresentou expresso em progenitores cardíacos (D9) (Figura 7.1A), com expressão similar ou maior que alguns genes sabidamente expressos em progenitores cardíacos, como GATA4, TBX5, MEF2C e ISL1, e que outros lncRNAs (Figura 7.1B). Não há nenhuma

informação na literatura com relação a sua função, apenas dados de expressão em diferentes tecidos adultos humanos, os quais mostram a expressão de *LINC00890* em endométrio e próstata (FAGERBERG *et al.*, 2014). Além disso, aparece associado aos polissomos, indicando um possível papel na regulação da tradução ou até mesmo como codificante de peptídeos. A possibilidade do *LINC00890* estar exercendo algum papel durante a diferenciação cardiomiogênica motivou seu estudo. Sua escolha foi baseada nos dados de RNA-seq (Capítulos I e II), gerados usando a diferenciação por corpos embriões (CE). A expressão de *LINC00890* no modelo de diferenciação em monocamada foi confirmada por qPCR (usando RNA total) (Figura 7.1A).

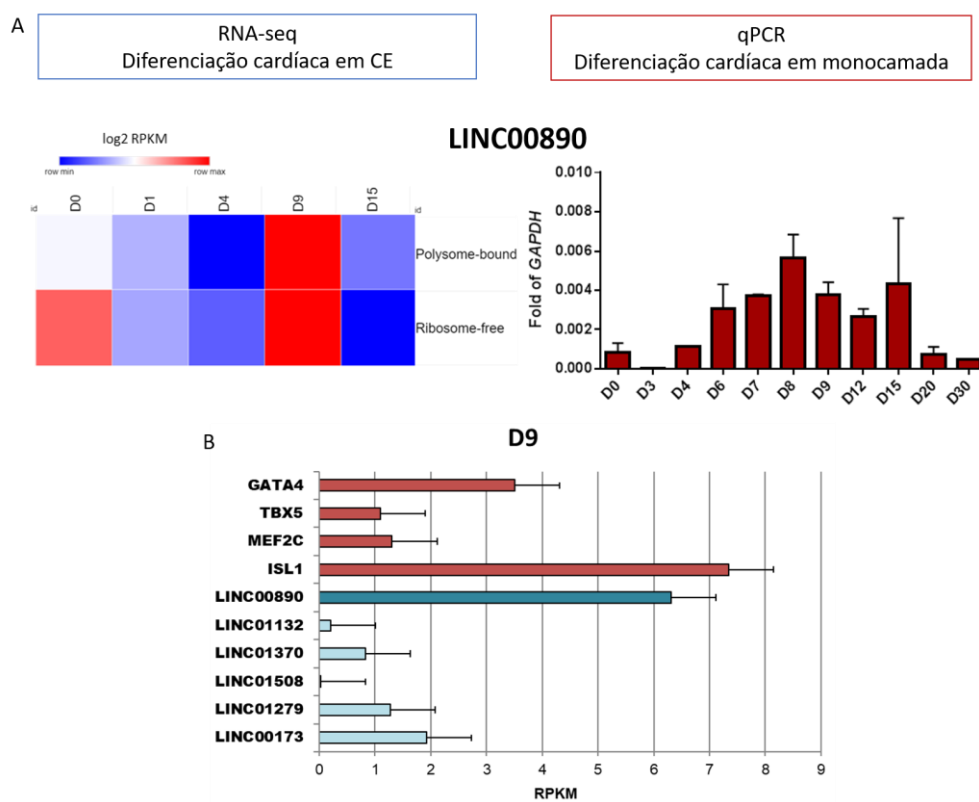


Figura 7.1 Expressão do lncRNA *LINC00890* durante a diferenciação cardiomiogênica em CE e monocamada. Células-tronco embrionárias foram submetidas a diferenciação cardiomiogênica e avaliadas em diferentes dias. (A) Os valores de expressão baseados no RNA-seq estão expressos em log₂ de RPKM e representados no mapa de calor; os de qPCR são valores relativos a expressão do gene constitutivo GAPDH. Os dias de diferenciação estão indicados na figura, assim como as frações de RNA livre (Ribosome-free) ou associado a ribossomos (Polysome-bound). (B) Níveis de expressão de genes marcadores de progenitores cardíacos (vermelho), de *LINC00890* e de outros lncRNAs identificados em D9 (dados de RNA-seq, Polysome-bound).

7.2 Construção de linhagens celulares de super expressão induzida de *LINC00890*

O vetor AAVS1-TRE-GW-rtTA foi clonado com o gene *LINC00890* e confirmado por reação de restrição (ANEXO III – Figura 1) e sequenciamento. Tal vetor foi usado para promover a integração do sistema de expressão induzida no locus AAVS1 do genoma das CTP. Após a transfecção, foram feitas a seleção com puromicina e a seleção clonal. As novas linhagens foram denominadas H1 iLINC e PLZ iLINC, de acordo com a célula parental. Até seis clones de cada nova linhagem foram selecionados e testados quanto a eficiência de indução da expressão do transgene através da adição de doxiciclina ao meio de cultivo. A quantificação da expressão foi feita através de qPCR, e um clone de cada condição foi escolhido para dar continuidade ao trabalho (Figura 7.2).

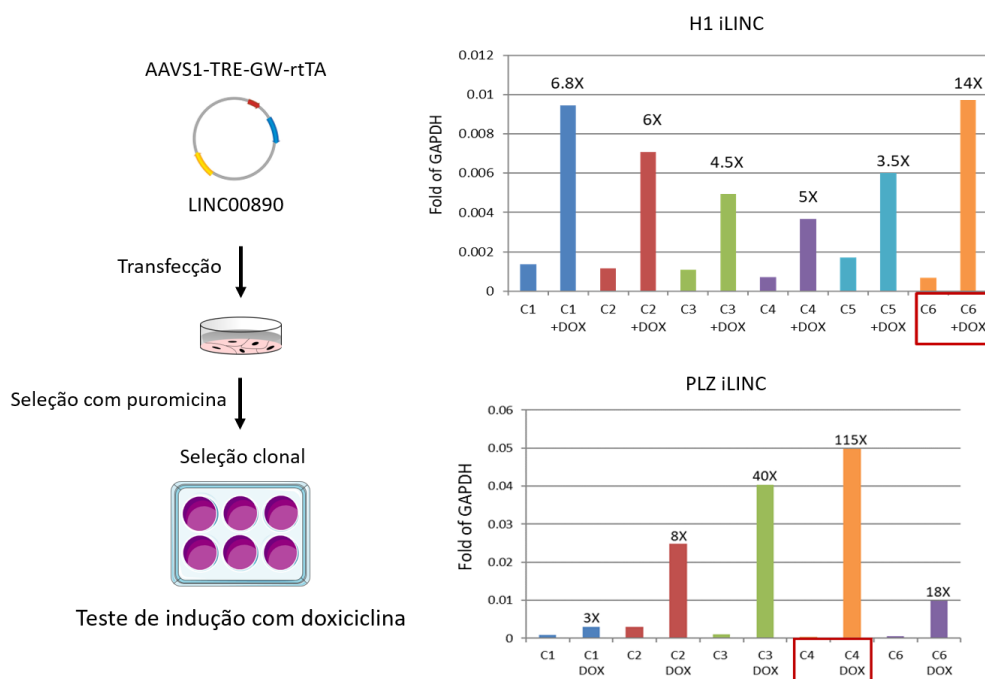


Figura 7.2 Sistema de super expressão induzível com doxiciclina. Após a transfecção com as construções vetoriais, as células foram submetidas à seleção com puromicina e clonal. Os clones selecionados foram testados quanto à expressão de *LINC00890* através da indução com 500 ng/ml ou 1 ug/ml (PLZ iLINC) de doxiciclina (+DOX). Valores relativos ao gene constitutivo GAPDH estão mostrados nos gráficos, e foram quantificados por qPCR. Todos os qPCR foram feitos com probes Taqman. Os clones escolhidos para a continuidade do trabalho estão destacados em vermelho. Valores indicados nas colunas +DOX representam a expressão quantas vezes maior que o controle sem DOX.

Após as etapas de seleção com puromicina e clonal, e de algumas passagens em cultura, a nova linhagem H1 iLINC (clone #6) manteve as características de células pluripotentes (Figura 7.3, painel da esquerda (-DOX)). Entre essas características

estão a morfologia em colônias bem delimitadas e a expressão dos marcadores de pluripotência Oct4, Sox2 e Nanog. Além disso, essas características também foram mantidas após a indução com doxiciclina (Figura 7.3, painel da direita (+DOX)), indicando que o gene *LINC00890* parece não participar da saída do estado de pluripotência.

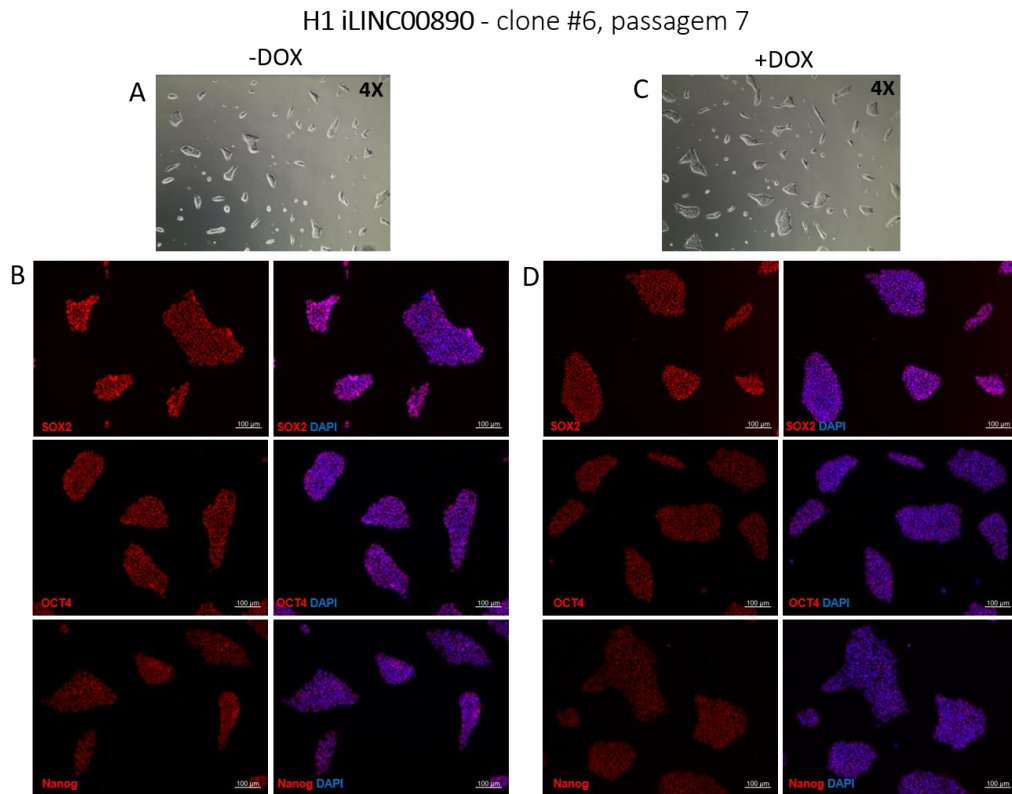


Figura 7.3 Linhagem H1 iLINC manteve características pluripotentes após algumas passagens em cultivo (-DOX) e após indução com doxiciclina (+DOX). Imagens em campo claro (A e C) foram tiradas após 96 h sem (-DOX) ou com indução (+DOX), no aumento de 4X, e mostram a morfologia das colônias de células. Imunofluorescência dos marcadores de pluripotência Oct4, Sox2 e Nanog das (B) células sem indução (-DOX) e (D) induzidas com doxiciclina (+DOX), após 7 dias. Imagens no aumento de 10X. Coloração nuclear com DAPI.

7.3 Ensaios funcionais com as linhagens H1 iLINC e PLZ iLINC

A linhagem H1 iLINC (clone #6) foi usada para avaliar a participação do gene *LINC00890* na diferenciação cardiomiogênica. Experimentos iniciais de triagem avaliaram a eficiência da diferenciação cardíaca e a expressão de marcadores após a indução da expressão de *LINC00890* ao longo de toda a diferenciação (indução contínua: DOX D0-15) (Figura 7.4). Os resultados mostraram uma redução na

porcentagem de células NKX2-5/TNNT2 duplo positivas na condição induzida DOX D0-15, quando comparado com o controle sem indução (Figura 7.4A), indicando, portanto, uma redução na eficiência de diferenciação. Além disso, a monocamada uniforme de células pulsantes observada no controle sem indução (-DOX) não era mais observada nas células induzidas de D0-15, sendo substituída por pequenos aglomerados de células pulsantes (Figura 7.4B). Esses resultados indicam que o *LINC00890* interfere de alguma maneira na diferenciação cardíaca e, por isso, demos continuidade no estudo desse gene.

Ao avaliar a expressão dos marcadores de subtipos de cardiomiócitos atriais e ventriculares, esperava-se que estivessem reduzidos após a indução contínua com doxiciclina (DOX D0-15), uma vez que a eficiência de diferenciação foi menor nessa condição. Entretanto, curiosamente, foi possível observar que os marcadores de células atriais (Figura 7.4D) tendiam a aumentar, enquanto os marcadores de células ventriculares reduziam (Figura 7.4E). Isso nos levou a criar a hipótese de que o gene *LINC00890* atua de diferentes maneiras em momentos distintos da diferenciação: (1) seria um inibidor da formação de mesoderme cardíaca no início da diferenciação e (2) promoveria uma melhora na diferenciação com favorecimento da diferenciação de células atriais a partir de progenitores cardíacos.

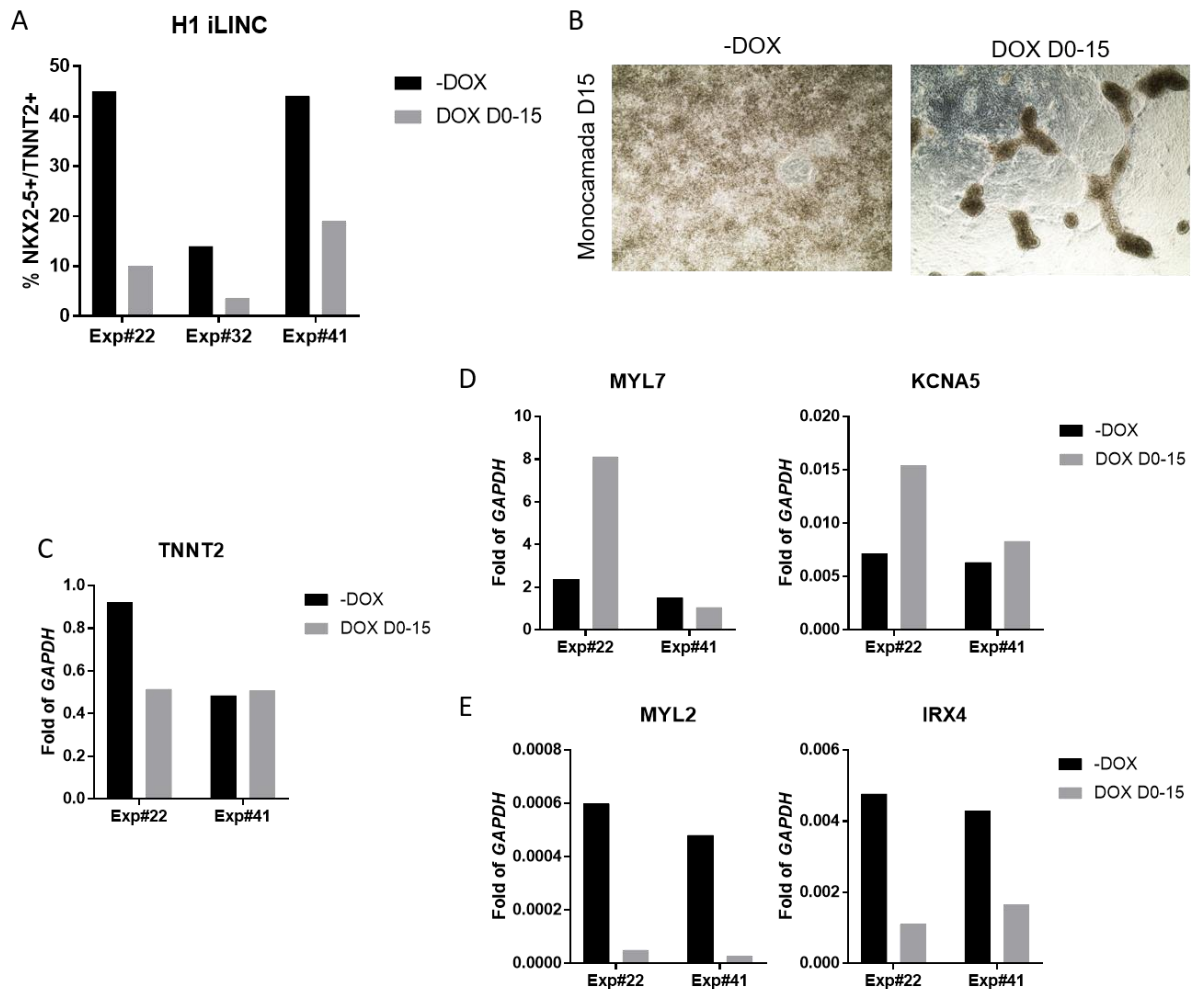


Figura 7.4 Ensaios funcionais de triagem com a linhagem H1 iLINC. A expressão contínua do *LINC00890* (DOX D0-15) ao longo da diferenciação cardiomiogênica reduziu a eficiência da diferenciação. (A) Quantificação por citometria de fluxo da porcentagem de células NKX2-5⁺/TNNT2⁺ duplo positivas no dia 15 da diferenciação nas condições sem indução (-DOX) e indução contínua (DOX D0-15). (B) Imagens representativas da diferenciação em monocamada de cardiomiócitos nas condições sem indução (-DOX) e indução contínua (DOX D0-15). Aumento de 4x. Quantificação da expressão por qPCR de (C) TNNT2 e dos marcadores (D) atriais e (E) ventriculares no dia 15 (valores relativos ao gene constitutivo GAPDH).

Considerando algumas limitações e a escolha, desde o princípio deste trabalho, de investigar o comprometimento final de progenitores cardíacos a cardiomiócitos, optamos por direcionar a investigação sobre o papel do *LINC00890* na segunda parte da hipótese proposta. Para isso, outros experimentos de triagem foram realizados com o intuito de definir uma janela de indução de expressão a ser estudada. Além disso, definiu-se que a avaliação dos parâmetros funcionais se daria no dia 30 da diferenciação, considerando que os cardiomiócitos estariam mais maduros e suas características de subtipos (atrial e ventricular) mais definidas. Dois experimentos independentes testaram diferentes janelas de indução com doxiciclina, focando nos

estágios de progenitores cardíacos (ANEXO V – Figura 1). A avaliação da expressão gênica de marcadores pan-cardíacos, atriais e ventriculares por qPCR mostrou que, de maneira geral, todas as janelas testadas levaram a um aumento dos marcadores após a indução. Entretanto, a indução entre os dias 5 e 10 da diferenciação (D5-10) se destacou pela homogeneidade e/ou maior expressão entre as replicatas. Portanto, essa janela de indução foi escolhida para os experimentos que seguiram.

Ao todo cinco replicatas foram realizadas com a linhagem H1 iLINC e uma com a linhagem PLZ iLINC para avaliar a expressão de marcadores no dia 30, após a indução com doxiciclina do gene *LINC00890* entre os dias 5 e 10 (Figuras 7.5 e 7.6). A expressão do gene *NKX2-5*, expresso a partir de progenitores cardíacos, não apresentou significância estatística quando comparado com as células não induzidas, apesar da média de expressão entre as replicatas ser maior do que a da condição controle. Entretanto, outros marcadores pan-cardíacos, como as troponinas *TNNT2* e *TNNI1* se mostraram estatisticamente aumentadas nas células induzidas, com exceção da troponina *TNNI3* (Figura 7.5A). O gene *ACTA2* é marcador de células musculares lisas, outras células derivadas dos mesmos progenitores cardíacos, e sua expressão se mostrou estatisticamente reduzida nas células induzidas, indicando um possível favorecimento de cardiomiócitos na presença do *LINC00890* (Figura 7.5B). Por outro lado, o gene *VWF*, marcador de células endoteliais também derivadas dos progenitores, não teve sua expressão alterada. A expressão do *LINC00890* também foi avaliada no dia 30, e se mostrou sem diferença estatística quando comparada com a das células não induzidas, indicando que nessa etapa da diferenciação (D30) o sistema de indução não está mais ativo, como esperado (Figura 7.5C).

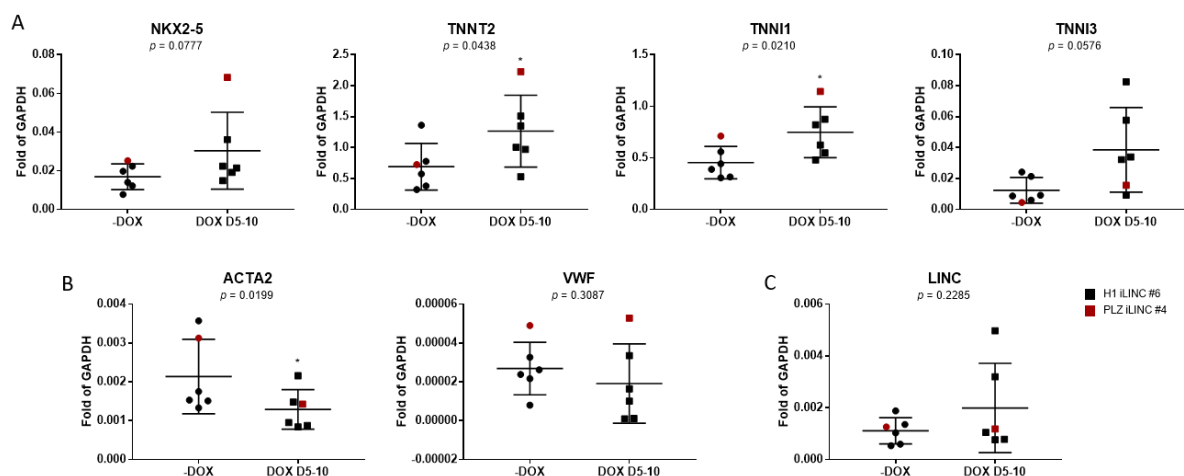


Figura 7.5 Expressão dos marcadores cardíacos nas linhagens H1 iLINC e PLZ iLINC após indução com doxiciclina nos dias 5-10 (DOX 5-10). Células não induzidas (-DOX) foram usadas como controle. (A) Marcadores pan-cardíacos, (B) marcadores de outras linhagens derivadas de progenitores cardíacos e a expressão do *LINC00890* foram avaliados por qPCR no dia 30 da diferenciação. Os valores estão expressos como relativos ao gene constitutivo GAPDH. Os resultados provenientes da linhagem PLZ iLINC estão identificados em vermelho, conforme a legenda. Análise por teste T de Student pareado: * $p < 0,05$.

Os marcadores de células atriais e ventriculares também foram avaliados nas células induzidas (DOX D5-10) e não induzidas (-DOX) (Figura 7.6). Dentre os marcadores atriais, a expressão do gene *MYL7* se mostrou significativamente aumentada nas células induzidas (Figura 7.6A). E entre os marcadores ventriculares, a expressão de *HEY2* também se mostrou aumentada nas células induzidas (Figura 7.6B). Os demais marcadores avaliados não apresentaram diferença estatística quando comparados com as células controle não induzidas, apesar da média de expressão de algum deles ter sido maior. Por outro lado, o marcador de células nodais *HCN4* não apresentou diferença entre as duas condições (Figura 7.6C), assim como os marcadores de maturação *SERCA2*, *KNCJ2* e *LCAD* (Figura 7.6D).

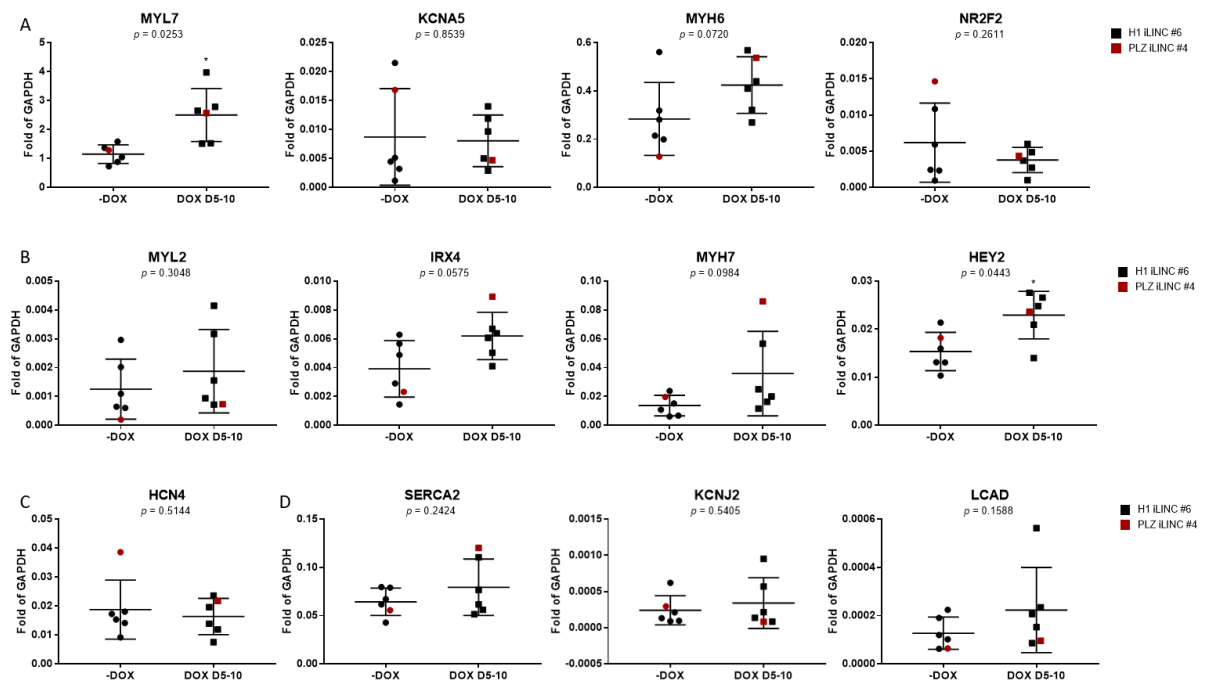


Figura 7.6 Expressão de marcadores de subtipos e de maturação de cardiomiócitos avaliada por qPCR no dia 30 da diferenciação. (A) Genes marcadores de células atriais. (B) Genes marcadores de células ventriculares. (C) Gene marcador de células nodais. (D) Genes marcadores de maturação, representando características como função do Ca^{2+} (SERCA2), canal de íons (KCNJ2) e metabolismo de lipídeos (LCAD). Os valores estão expressos como relativos ao gene constitutivo GAPDH. Os resultados provenientes da linhagem PLZ iLINC estão identificados em vermelho, conforme a legenda. Análise por teste T de Student pareado: * $p < 0,05$.

Um experimento representativo com cada linhagem iLINC foi analisado tanto no dia 10 como no dia 30 da diferenciação (Figura 7.7). Os resultados de expressão gênica mostraram que o gene *LINC00890* estava aumentado no dia 10 nas células induzidas (indução DOX D5-10), quando comparado com as células não induzidas, como esperado. Por outro lado, essa diferença não foi mais observada no dia 30, 20 dias depois da retirada da doxíciclina do meio de cultura e, portanto, interrupção da indução do sistema de super expressão. Por outro lado, os marcadores cardíacos NKX2-5 e TNNT2 não apresentaram diferenças entre as células induzidas e controle no dia 10. Diferenças na expressão desses genes apareceram mais tardiamente na diferenciação das células, após o período de super expressão de *LINC00890* (Figura 7.7).

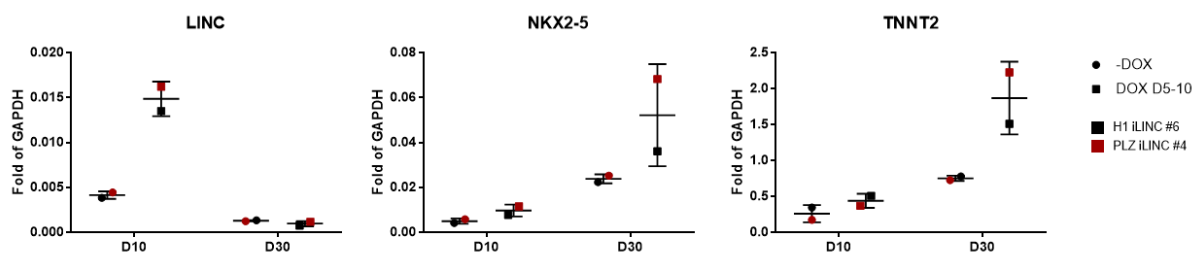


Figura 7.7 Expressão gênica nas linhagens H1 iLINC e PLZ iLINC nos dias 10 e 30 da diferenciação. Células não induzidas (-DOX) e induzidas (DOX D5-10) foram avaliadas por qPCR (n=2). Os valores estão expressos como relativos ao gene constitutivo GAPDH. Os resultados provenientes da linhagem PLZ iLINC estão identificados em vermelho, conforme a legenda.

Em conjunto, esses dados mostram que, de maneira geral, a expressão induzida de *LINC00890* entre os dias 5 e 10 da diferenciação levou ao aumento de alguns marcadores de cardiomiócitos, indicando um possível papel de favorecimento da diferenciação cardíaca do *LINC00890*.

Para continuar investigando o fenótipo dos cardiomiócitos após a indução da expressão do *LINC00890*, alguns ensaios fisiológicos foram realizados. Observou-se que a monocamada de cardiomiócitos diferenciados tinha diferenças na frequência das contrações espontâneas entre as células induzidas e não induzidas. Então, a quantificação da frequência de contrações foi feita usando filmagens das células em cultura. Células que tiveram a expressão de *LINC00890* entre os dias 5-10 da diferenciação apresentaram frequência de batimentos maior que as células controle sem indução (Figura 7.8).

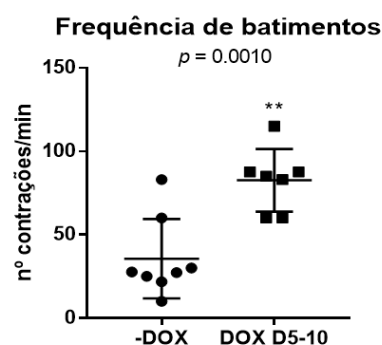


Figura 7.8 Quantificação da frequência de batimentos nos cardiomiócitos derivados da linhagem H1 iLINC. Células não-induzidas (-DOX) e induzidas entre os dias 5-10 da diferenciação cardíaca com doxiciclina (DOX D5-10) foram avaliadas pelo número de contrações por minuto. Análise por teste T de Student não-pareado: ** $p = 0,001$.

Ainda, células induzidas (DOX D5-10) e não induzidas (-DOX) foram submetidas a análise pela técnica de Patch clamp. As informações geradas por essa técnica permitem inferir sobre os tipos de cardiomiócitos que estão presentes numa população de células através dos parâmetros eletrofisiológicos, como o potencial de ação. Os resultados mostraram que na população de células que tiveram a expressão de *LINC00890* induzida haviam mais células com características atriais (41,7%) do que na população não induzida (22,9%) (Figura 7.9A). Interessantemente, o número de células ventriculares foi similar entre as duas condições, enquanto o número de células nodais foi menor nas células induzidas (de 22,9% na condição -DOX e 2,8% na DOX D5-10). Outros parâmetros também apresentaram diferenças entre as células induzidas e não induzidas, como a duração do potencial de ação em 90% de repolarização (APD90), a inclinação máxima do traço ascendente do potencial de ação (dV/dtmax) e a amplitude do potencial de ação (Amp), os quais se mostraram aumentados nos cardiomiócitos ventriculares DOX D5-10. Enquanto o potencial de membrana de repouso mostrou-se reduzido nessas células (Figura 7.9B e C). Na Figura 7.9D está ilustrado a diferença na amplitude dos cardiomiócitos -DOX e DOX D5-10.

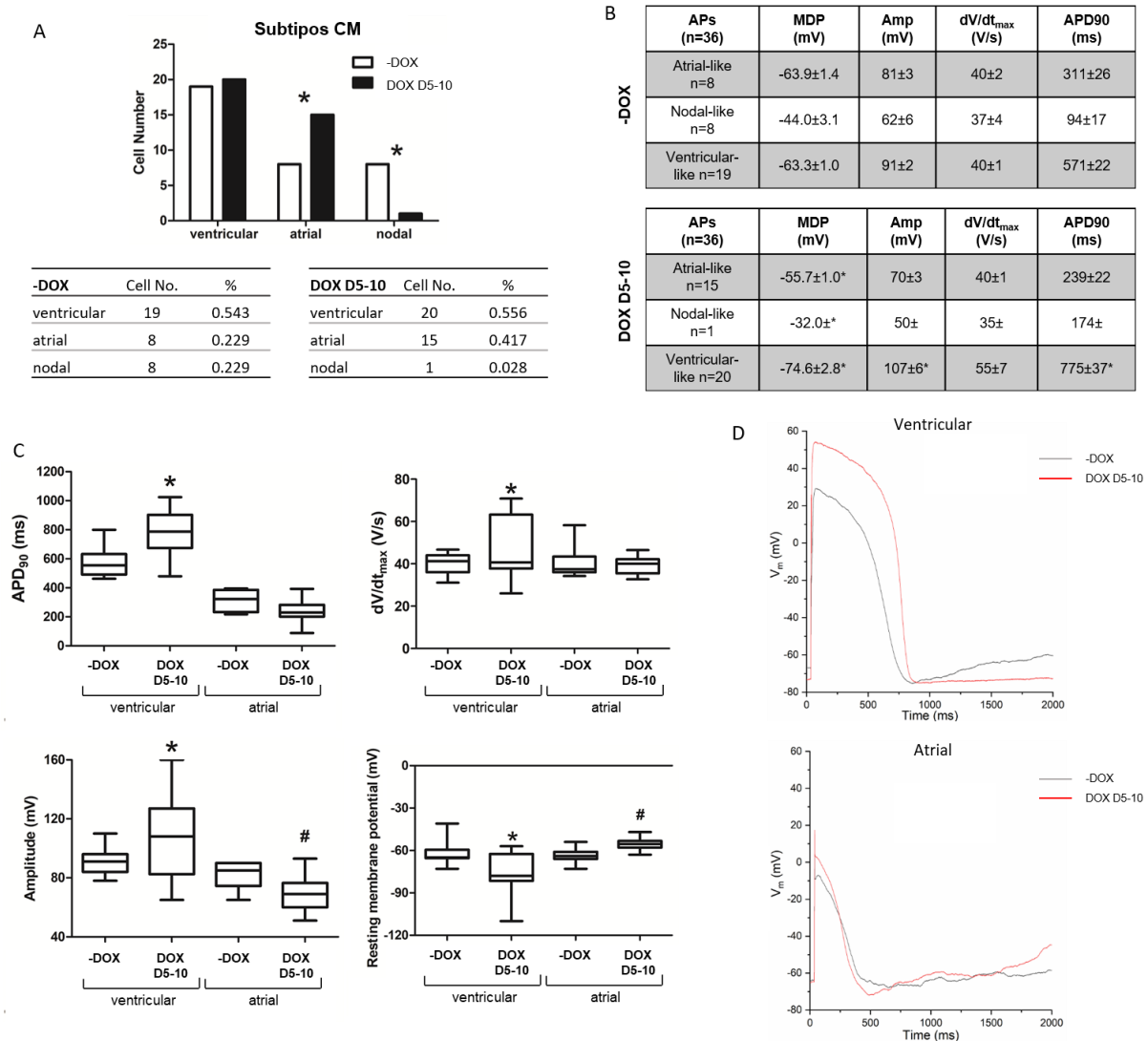


Figura 7.9 Análise eletrofisiológica de cardiomiócitos por Patch clamp. (A) Quantificação dos subtipos ventricular, atrial e nodal de cardiomiócitos nas populações de células não induzidas (-DOX) e induzidas (DOX D5-10) no dia 30 da diferenciação. Análise por teste Chi-quadrado, $*p=0.0225$. O número de células e o percentual que representam estão indicados. (B) Valores dos parâmetros medidos por Patch clamp nos cardiomiócitos das duas condições (n=36, cada condição). APs, potenciais de ação; MDP, potencial diastólico máximo; Amp, amplitude; dV/dt_{max}, inclinação máxima do traço ascendente do potencial de ação; APD90, duração do potencial de ação em 90% de repolarização. (C) Parâmetros descritos em B plotados graficamente. $*p<0,05$ quando comparado com -DOX ventricular. $\#p<0,05$ quando comparado com -DOX atrial. (D) Imagens representativas dos potenciais de ação de cardiomiócitos ventriculares e atriais das condições -DOX e DOX D5-10.

7.4 Construção de linhagens com silenciamento do gene *LINC00890*

Com o intuito de avaliar mais a fundo a participação do *LINC00890* na diferenciação cardiomiogênica, o silenciamento do gene foi feito nas linhagens H1 e PLZ usando a tecnologia de CRISPR/Cas9. Quatro diferentes gRNAs foram

desenhados de maneira que um par pudesse ser usado para deletar o gene do *LINC00890* por completo (Figura 7.10A), e suas sequências foram clonadas no vetor px458 que permite a expressão simultânea da proteína Cas9. A eficiência dos gRNAs em seus pares pode variar, por isso, inicialmente as quatro possibilidades de combinação entre os gRNAs (A, B C e D) foram testadas (Figura 7.10B). O par de gRNAs A+D foi capaz de gerar a deleção esperada, evidenciada pela amplificação do fragmento de 514 pb na PCR usando os oligos 1LF e 1LR. Esse par de gRNAs foi, então, escolhido para dar seguimento ao silenciamento gênico.

A transfecção desses vetores tem muito baixa eficiência, em torno de 5-10% (dados não mostrados). Além disso, o sistema não possui seleção por antibiótico, fazendo com que a geração de uma linhagem com silenciamento dependa exclusivamente da seleção clonal. Por isso, várias tentativas tiveram que ser feitas e inúmeros clones analisados até que se conseguisse isolar células devidamente silenciadas. Esse processo consistiu de algumas etapas: (1) transfecção, (2) teste da população *bulk*, (3) clonagem manual, (4) teste dos clones, (5) subclonagem, (6) teste dos subclones, (7, se aplicável) sub-subclonagem e teste dos sub-subclones. Se a população *bulk* inicial não fosse positiva para a deleção (amplicon de 514 pb - KO), uma nova transfecção deveria ser feita. Ou se nenhum dos 24 clones testados a partir de uma população *bulk* positiva não fossem positivos, uma nova transfecção também deveria ser feita. A amplificação por PCR de um fragmento de 780 pb usando os oligos 3LF e 3LR foi usada como controle e indicativo da presença de células selvagens (WT) na população testada. Essas etapas foram seguidas até que se encontrasse um clone/subclone positivo para a deleção e negativo para o fragmento WT (Figura 7.10).

Desta maneira, a partir de populações H1 e PLZ *bulk* positivas para KO, pelo menos 24 clones foram selecionados e testados. Encontrou-se um clone derivado da linhagem H1 (Figura 7.10C) e três da linhagem PLZ (Figura 7.10D), os quais foram subclonados. Pelo menos um subclone de cada linhagem apresentou a amplificação correspondente a KO e foi negativo para a amplificação WT. Os clones escolhidos foram denominados H1 8.16 LINC KO e PLZ 4.7 LINC KO.

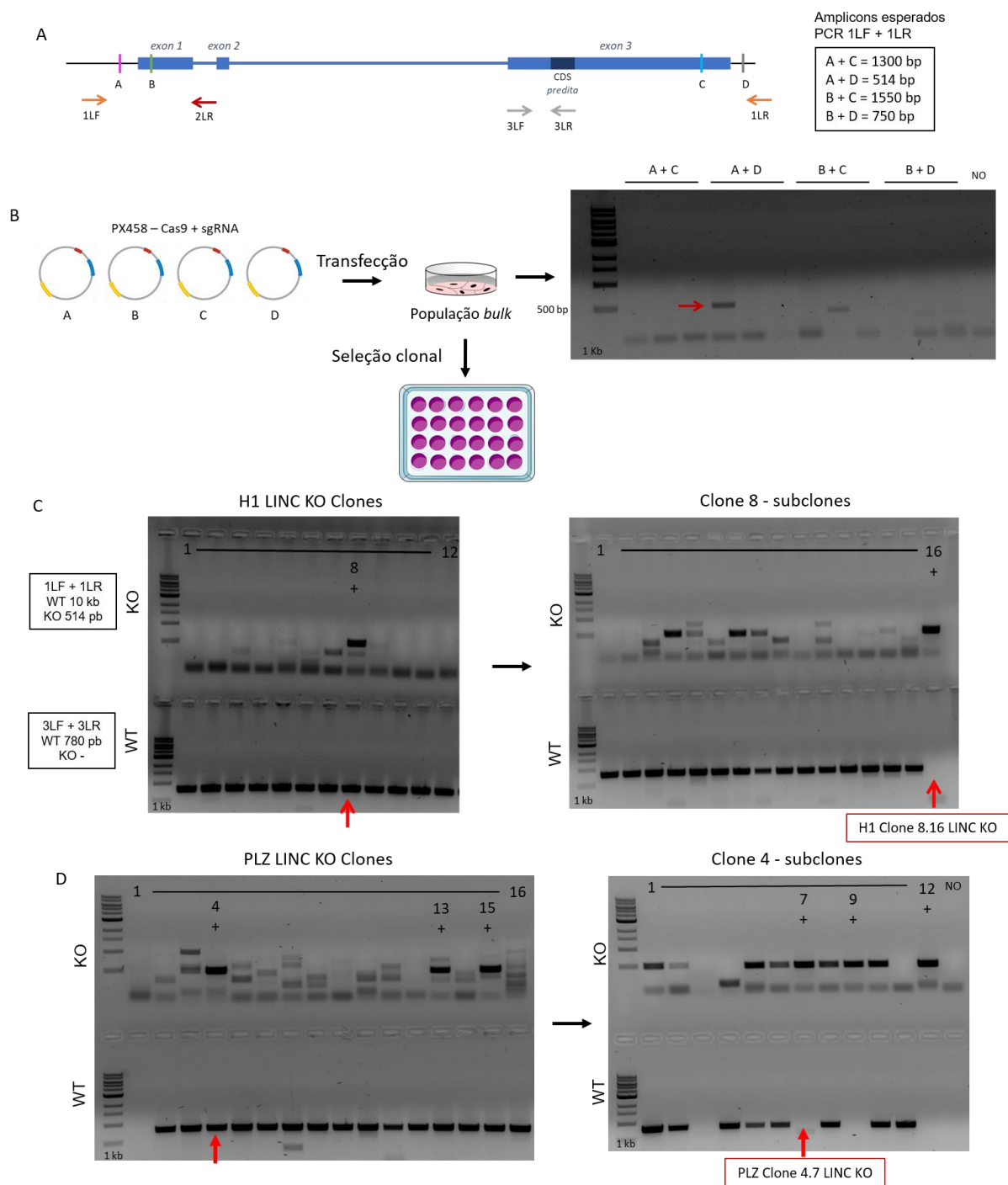


Figura 7.10 Construção de linhagens com silenciamento de *LINC00890*. (A) Esquema representativo do gene *LINC00890*, a posição relativa dos gRNAs e oligos para PCR teste e os tamanhos esperados dos amplicons. (B) Esquema representativo dos procedimentos de transfecção e seleção de clones, e gel de agarose ilustrando a PCR (1LF+1LR) da triagem de eficiência dos pares de gRNAs. A flecha indica a amplificação do fragmento esperado de 514 pb. Imagens dos géis de agarose ilustrando a amplificação dos fragmentos de KO (514 pb) e WT (780 pb) nos clones e subclones derivados da linhagem (C) H1 e (D) PLZ. 1kb, marcador de tamanho em pb. NO, controle negativo. Cada coluna do gel representa a PCR de um clone, numerados na parte superior. KO, amplificação do fragmento após a deleção usando os oligos 1LF e 1LR. WT, amplificação do fragmento interno do gene usando os oligos 3LF e 3LR. As flechas vermelhas indicam os clones que foram subclonados e os subclones escolhidos.

Para facilitar a triagem de inúmeros clones, as PCRs foram feitas usando DNA extraído através da fervura das células em água (vide item 6.2.3). Uma vez que os clones H1 8.16 LINC KO e PLZ 4.7 LINC KO foram escolhidos, e para garantir uma PCR mais sensível, o DNA genômico dos clones foi extraído usando um kit e as PCRs teste foram repetidas. Desta vez, o clone H1 8.16 apresentou amplificação correspondente a células WT (Figura 7.11A), indicando ainda uma possível contaminação com essas células. Por isso, decidiu-se seguir com mais uma etapa de clonagem (sub-subclonagem). Os sub-subclones foram novamente testados por PCR, assim como os subclones positivos de PLZ clone 4 (Figura 7.10D e dados não mostrados), usando DNA genômico extraído com kit (Figura 7.11B). Os resultados mostraram que todos os clones testados apresentaram a amplificação KO, enquanto as células parentais H1 e PLZ foram negativas, como esperado. Com relação as duas reações WT, alguns clones ainda apresentaram amplificação, entretanto foi possível selecionar um clone de cada linhagem que aparentemente parece limpo, indicando que, na sua grande maioria, as células da sua população correspondem a células mutadas KO. Os clones escolhidos foram denominados H1 8.16.10 LINC KO e PLZ 4.7 LINC KO.

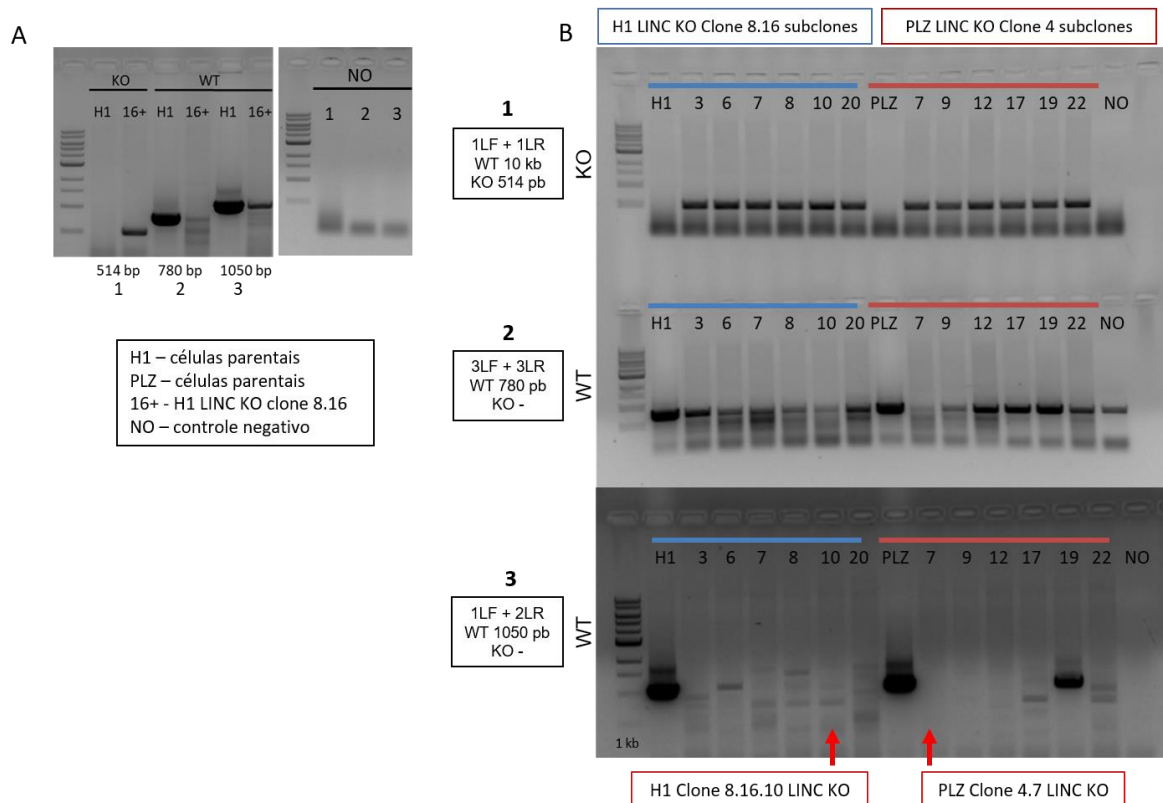


Figura 7.11 Legenda na próxima página.

Figura 7.11 Seleção de clones com silenciamento de *LINC00890* por deleção. (A) O clone H1 8.16 foi testado por PCR usando DNA genômico extraído com kit. As reações KO (1) e WT (2 e 3) estão explicadas na figura. (B) Uma nova clonagem foi feita e os sub-subclones de H1 8.16 foram testados por PCR com DNA extraído com kit, assim como os subclones positivos de PLZ clone 4. Imagens dos géis de agarose ilustrando a amplificação dos fragmentos de KO (514 pb) e WT (780 pb e 1050 bp). 1kb, marcador de tamanho em pb. NO, controle negativo. Cada coluna do gel representa a PCR de um clone, numerados na parte superior. As flechas vermelhas indicam os clones que foram escolhidos.

Mesmo após diversas etapas de clonagem, os clones selecionados H1 8.16.10 LINC KO e PLZ 4.7 LINC KO mantiveram suas características de pluripotência, evidenciado pela morfologia das colônias de células em cultura (Figura 7.12). Ainda, a deleção de *LINC00890* não parece afetar tais características, indicando que esse gene não parece ser fundamental na manutenção do estado pluripotente.

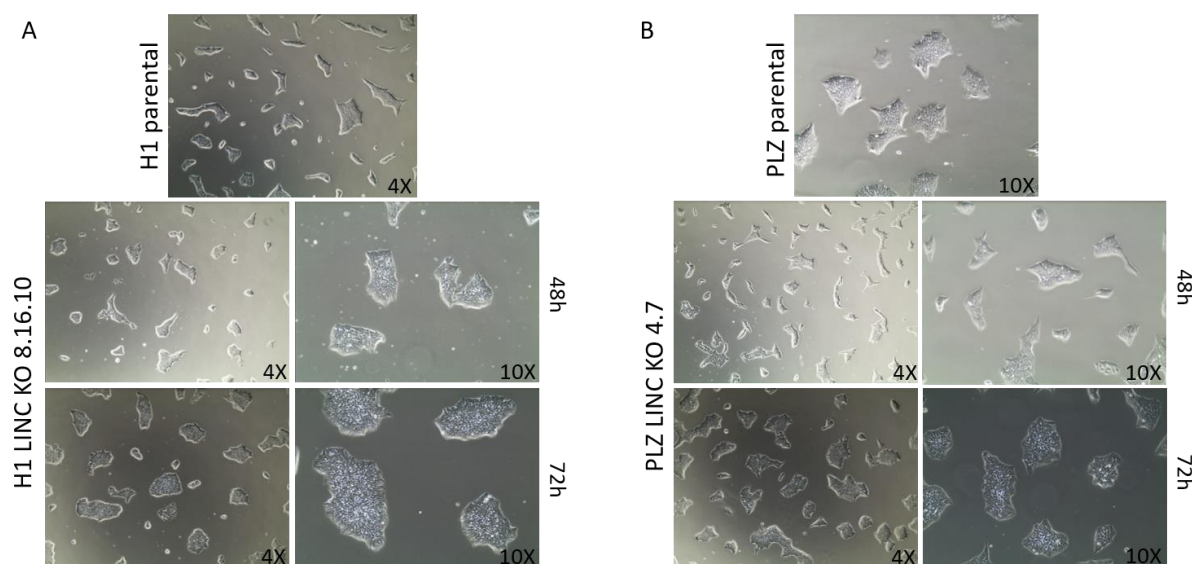


Figura 7.12 Clones silenciados para *LINC00890* mantêm o estado pluripotente. Células (A) H1 parental e clone LINC KO 8.16.10, e (B) PLZ parental e clone LINC KO 4.7 foram fotografadas 48h e 72h após plaqueamento. A morfologia das colônias com bordas bem delimitadas é uma das características principais de células pluripotentes. Imagens nos aumentos de 4x ou 10x.

7.5 Ensaio funcional com as linhagens H1 LINC KO e PLZ LINC KO

As linhagens que tiveram o gene *LINC00890* silenciado foram usadas para avaliar a participação deste na diferenciação cardiomiogênica. Um experimento representativo com cada linhagem foi feito a fim de investigar a expressão de *LINC00890* ao longo da diferenciação e confirmar o silenciamento (Figura 7.13). Na

linhagem H1 KO, os níveis de expressão de *LINC00890* foram muito menores que os das células WT ao longo de toda a diferenciação. Por sua vez, a linhagem PLZ KO apresentou uma pequena expressão de *LINC00890* nos dias 5, 10 e 15 da diferenciação, possivelmente derivada de células WT contaminantes na cultura, mas que ainda assim representam valores muito mais baixos dos expressos pelas células WT.

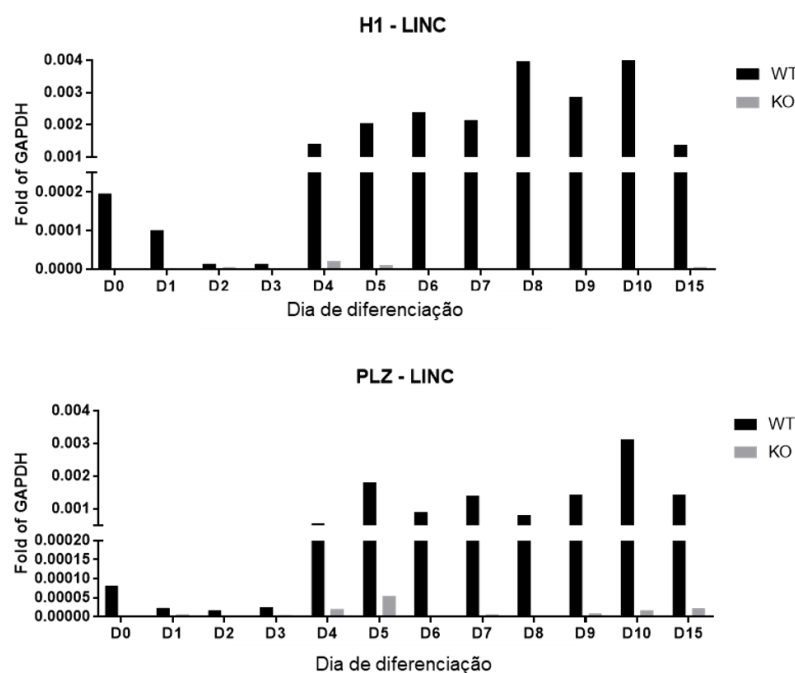


Figura 7.13 Expressão de *LINC00890* nas linhagens H1 LINC KO e PLZ LINC KO ao longo da diferenciação cardiomiogênica. Células parentais H1 ou PLZ (WT) foram usadas como controle. A expressão foi quantificada por qPCR e os valores estão expressos relativos ao gene constitutivo GAPDH. WT, células selvagens parentais. KO, células silenciadas.

Experimentos de diferenciação em monocamada independentes foram realizados com a linhagem H1 LINC KO 8.16.10 e os resultados mostraram que as células não foram capazes de se diferenciar em cardiomiócitos (Figura 7.14A). Células KO no dia 15 da diferenciação não apresentaram expressão de *LINC00890*, enquanto as células H1 parentais (WT) apresentaram, confirmando mais uma vez o silenciamento (Figura 7.14B). Ainda, as células KO apresentaram níveis de expressão significativamente mais baixos do que células WT para os genes cardíacos *NKX2-5* e *TNNT2*, indicando a incapacidade de diferenciação daquelas em cardiomiócitos (Figura 7.14C). Por outro lado, o marcador de células musculares lisas, provenientes dos mesmos progenitores, se mostrou superior nas células KO quando comparadas com as WT, indicando que o silenciamento de *LINC00890* parece prejudicar

especificamente a diferenciação de cardiomiócitos. O marcador de células endoteliais VWF não apresentou diferenças entre as comparações.

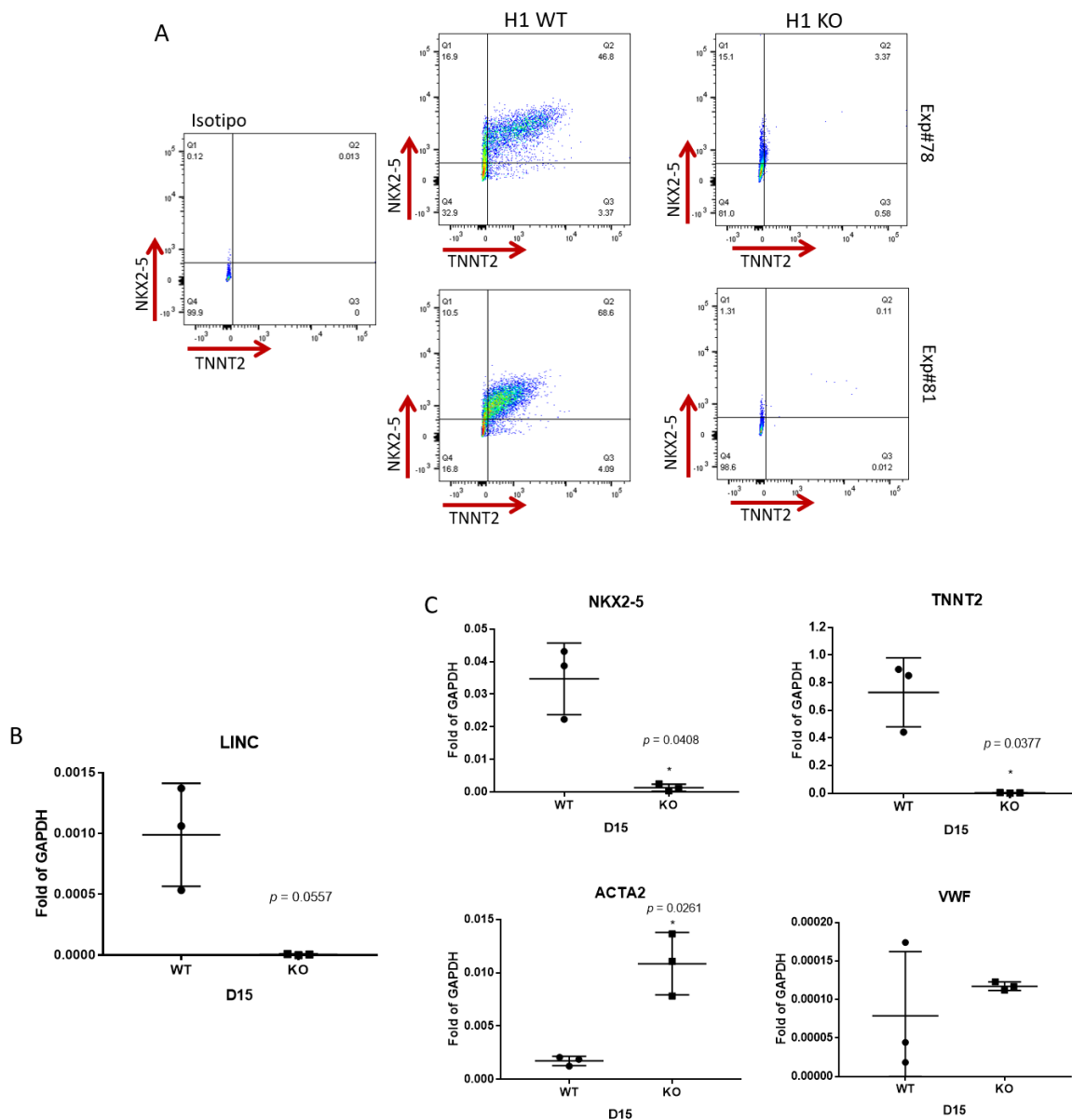


Figura 7.14 Diferenciação cardiomiogênica com a linhagem H1 LINC KO. Células no dia 15 da diferenciação foram avaliadas através da (A) imunomarcagem para NKX2-5 e TNNT2 por citometria de fluxo e qPCR quanto a expressão de (B) *LINC00890* e (C) marcadores de células derivadas de progenitores cardíacos: cardiomiócitos, células musculares lisas e endoteliais. Células parentais H1 (WT) foram usadas como controle. A expressão gênica foi quantificada por qPCR e os valores estão expressos relativos ao gene constitutivo GAPDH. WT, células selvagens parentais. KO, células silenciadas.

Frente a incapacidade de diferenciação de cardiomiócitos das células LINC KO, marcadores de pluripotência, mesoderme e cardíaco foram seguidos ao longo da diferenciação visando comparar células WT e KO. Células LINC KO, tanto da

linhagem H1 quanto da linhagem PLZ, apresentaram níveis de expressão do gene POU5F1 (OCT4) semelhantes às células WT, confirmando seu estado pluripotente (Figura 7.15). Os níveis de expressão do gene T (Brachyury) também se mostraram bem semelhantes entre as células KO e WT, com o pico esperado no D1, indicando que as células KO mantiveram a capacidade de diferenciação em células mesodérmicas. O gene EOMES, marcador de mesoderme cardíaca, apresentou diferenças entre as linhagens H1 LINC KO e PLZ LINC KO. Na linhagem H1, sua expressão nas células KO foi muito mais baixa do que não células WT, indicando que aquelas falharam na formação da mesoderme cardíaca. Por outro lado, as células PLZ KO apresentaram os mesmos níveis de expressão que as WT (Figura 7.15).

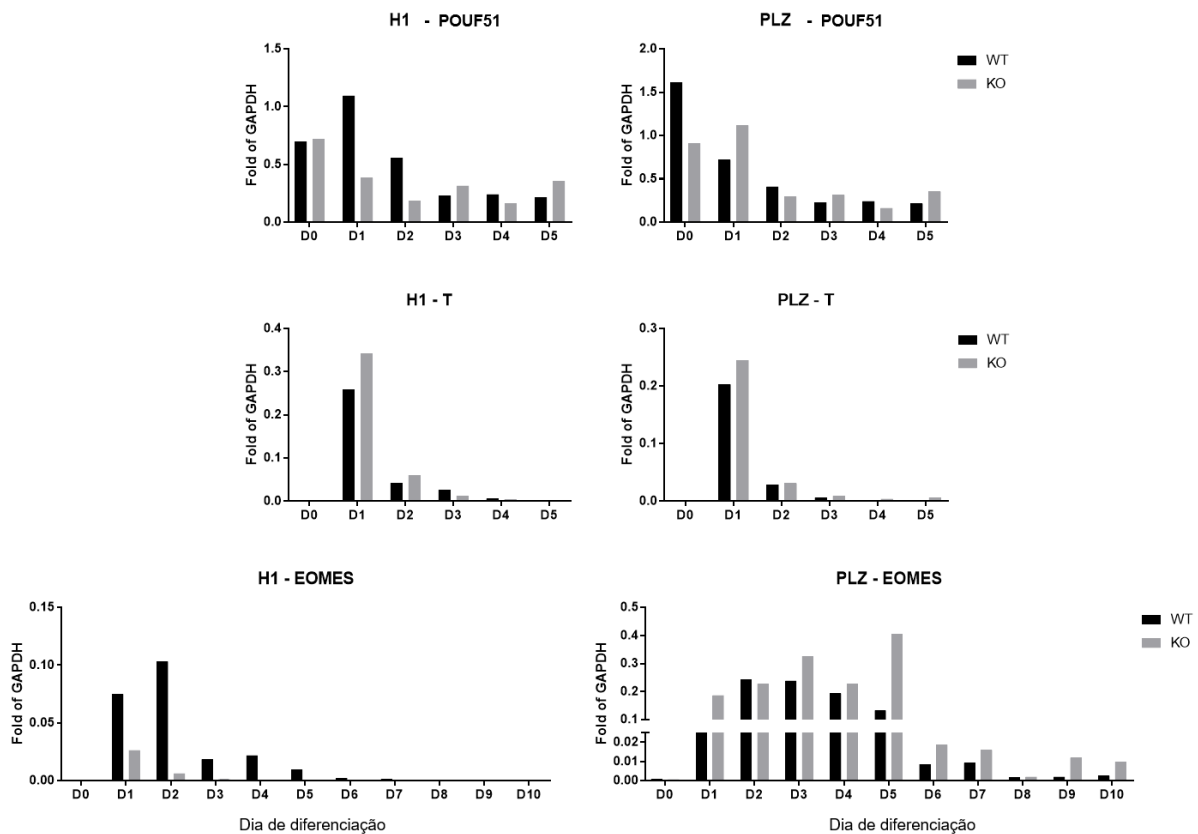


Figura 7.15 Expressão gênica de marcadores de pluripotência e mesoderme durante a diferenciação cardíaca das linhagens H1 LINC KO e PLZ LINC KO. Células parentais H1 ou PLZ (WT) foram usadas como controle. A expressão foi quantificada por qPCR e os valores estão expressos relativos ao gene constitutivo GAPDH. WT, células selvagens parentais. KO, células silenciadas.

Com relação aos marcadores cardíacos, as linhagens H1 LINC KO e PLZ LINC KO também se comportaram de maneira diferente quando comparadas com seus pares WT. A linhagem H1 KO apresentou níveis de expressão muito mais baixos que as células WT nos genes GATA4, NKX2-5 e TNNT2, indicando a incapacidade de

diferenciação em cardiomiócitos (Figura 7.16). Por sua vez, a linhagem PLZ KO apresentou níveis de expressão similares de GATA4 comparado com células WT, e níveis sutilmente menores de NKX2-5 e TNNT2, cuja diferença ficou mais evidente nos dias finais da diferenciação. Apesar disso, ambas as linhagens KO não foram capazes de pulsar espontaneamente em cultura até o dia 15 (Figura 7.16).

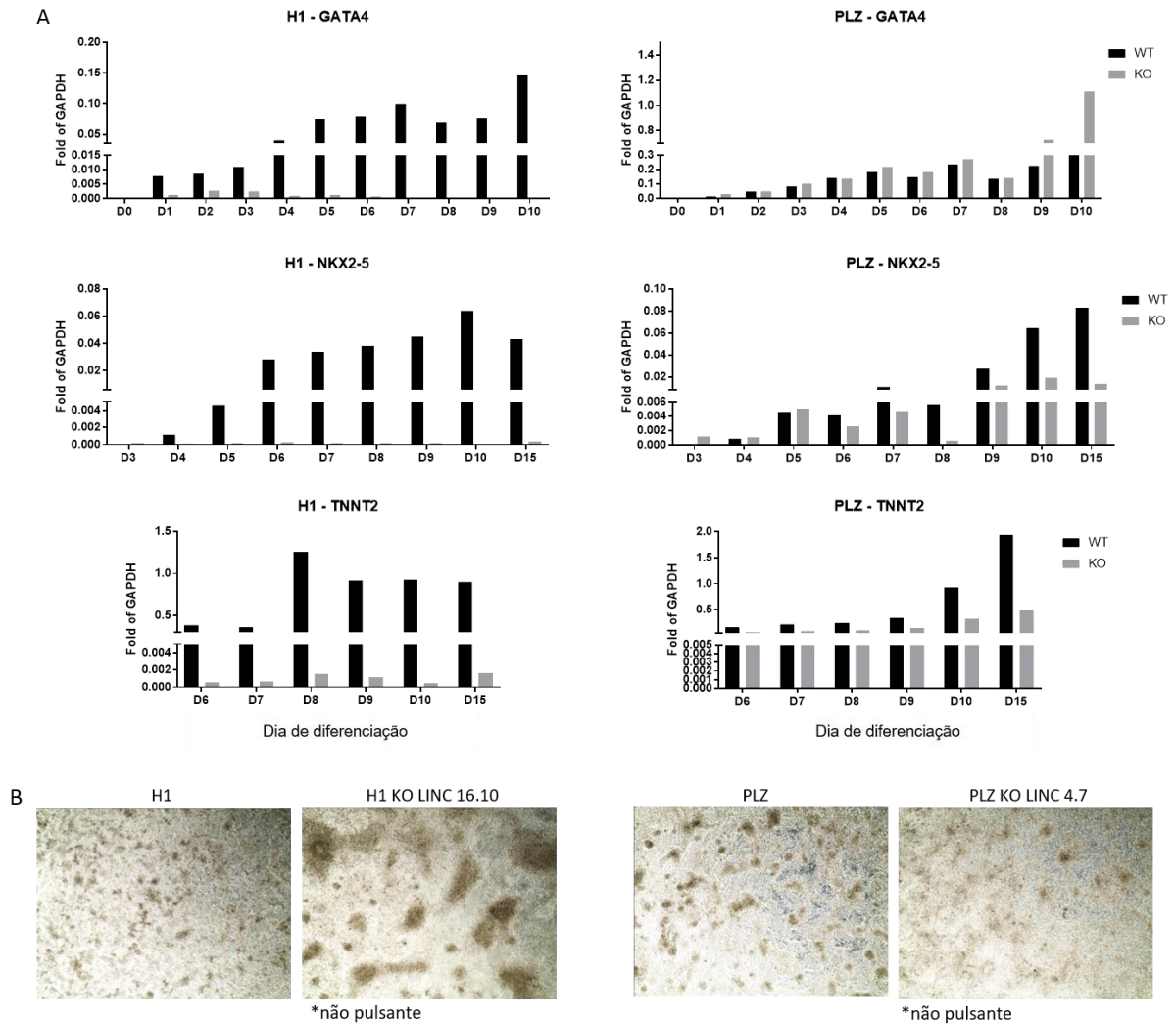


Figura 7.16 Expressão gênica de marcadores cardíacos durante a diferenciação cardíaca das linhagens H1 LINC KO e PLZ LINC KO. Células parentais H1 ou PLZ (WT) foram usadas como controle. (A) A expressão foi quantificada por qPCR e os valores estão expressos relativos ao gene constitutivo GAPDH. WT, células selvagens parentais. KO, células silenciadas. (B) Imagens representativas da diferenciação cardiomiogênica em monocamada das linhagens H1 e H1 KO LINC 8.16.10, e PLZ e PLZ KO LINC 4.7. *não pulsante, células não apresentaram pulsação espontânea.

7.6 Análise *in silico* do gene *LINC00890*

O mapeio dos dados de RNA-seq descritos no Capítulo I foi analisado cuidadosamente na região desse gene, a fim de se identificar um possível enriquecimento de *reads* (leituras) de alguma variante. São anotadas duas variantes de transcritos para esse gene: ENST00000569275, que abrange os éxons 1, 2 e 3; e ENST00000563467, que abrange o éxon 2 e parcialmente os éxons 1 e 3 (Figura 7.19, painel de baixo). A Figura 7.17 traz o mapeio dos dados de RNA-seq, mostrando que a variante ENST00000569275 (variante 1) parece ser a mais representativa.

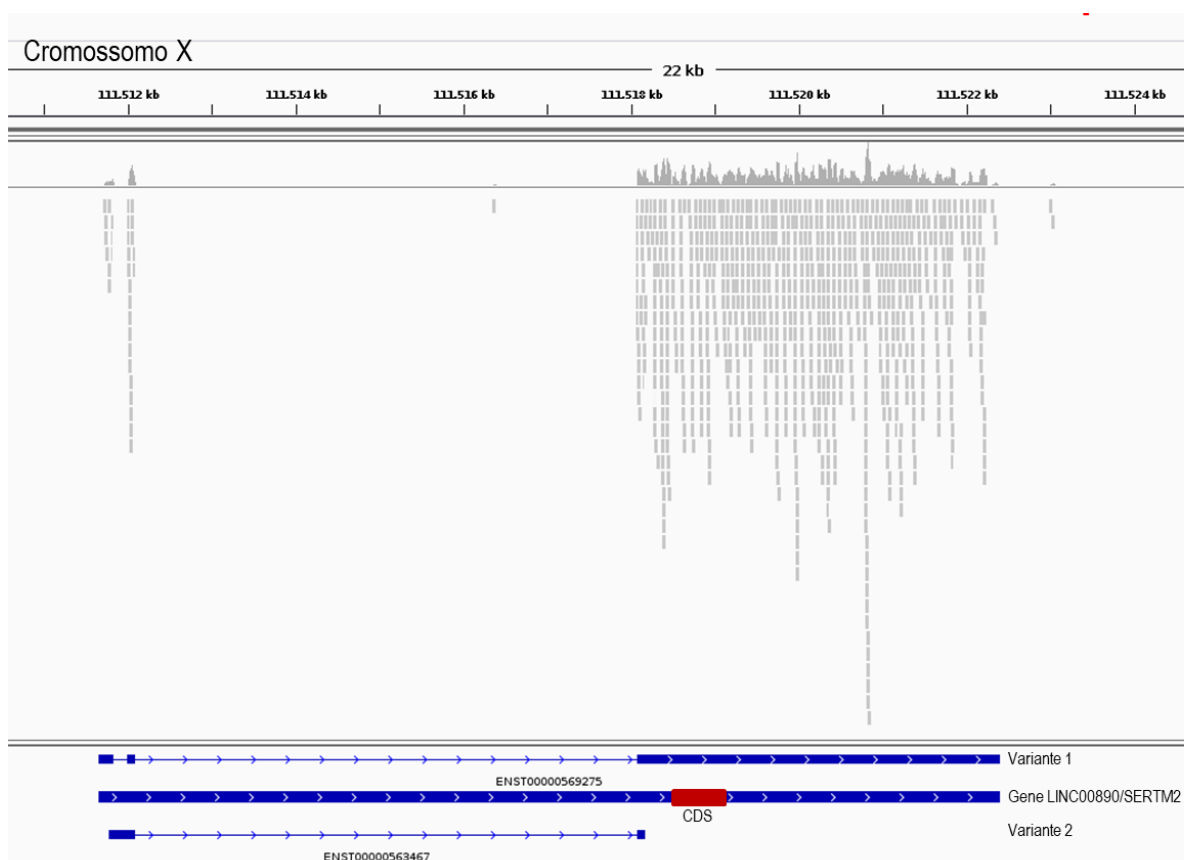


Figura 7.17 Mapeio dos dados de RNA-seq na região do gene *LINC00890/SERTM2*. A CDS predita está indicada em vermelho. Esse gene possui duas variantes de splicing anotadas: ENST00000569275 e ENST00000563467.

Ainda, a análise de genes ortólogos para o gene *LINC00890* humano foi feita, a fim de se identificar similaridades de sequência com outros organismos. Recentemente, o lncRNA *LINC00890* foi anotado como um gene codificante de proteína e renomeado para *SERTM2* (NCBI: NM_001354473.1). Utilizando a ferramenta BLAST (blastn, NCBI) e a sequência nucleotídica de *LINC00890/SERTM2*

(NM_001354473.1) como referência, os alinhamentos que apresentaram maior similaridade estão mostrados na Tabela 7.1. As sequências de vertebrados não-mamíferos (*Python bivittatus* e *Gallus gallus*) e da proteína homóloga SERTM1, conforme Tabela 7.1, foram incluídas na análise filogenética como comparativos. A ferramenta online CLUSTAL Omega (*multiple sequence alignment*) foi usada para gerar a árvore filogenética que mostra a relação entre os organismos, a partir do alinhamento das sequências citadas anteriormente (Figura 7.18).

Observou-se maior identidade e cobertura (QC) em espécies de primatas não-humanos, e maior conservação entre os mamíferos. Os mamíferos de pequeno porte e modelos murinos *Rattus norvegicus* e *Mus musculus* apresentaram baixa cobertura da sequência, com alta identidade, indicando que apenas uma pequena porção do transcrito de RNA humano é similar nesses organismos.

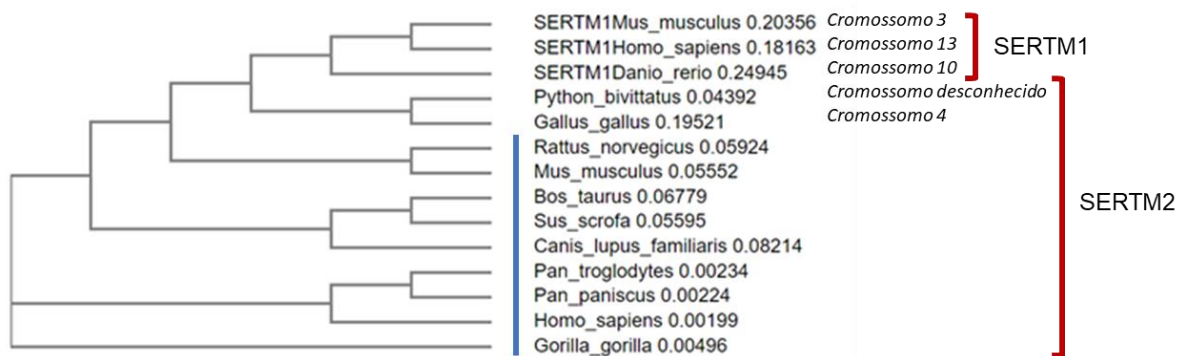


Figura 7.18 Árvore filogenética relacionando os organismos de acordo com a sequência de RNA de *SERTM2*. O alinhamento e geração da árvore filogenética foram feitos pela ferramenta CLUSTAL Ômega, usando as sequências listadas na Tabela 7.1. Árvore filogenética sem correção de distância. Os genes estão localizados no cromossomo X, ou conforme indicado. Sequências da proteína homóloga SERTM1 foram usadas como referências externas para enraizar a árvore. A faixa azul indica os organismos mamíferos e sua relação mais direta quanto à conservação de *SERTM2*.

Tabela 7.1 Análise comparativa de seqüências nucleotídicas de diferentes organismos quanto à similaridade com o gene *LINC00890/SERTM2* humano. QC, query coverage (cobertura da seqüência de referência). Identity, identidade. *A seqüência de *SERTM2* humana foi usada como referência.(Continua na próxima página)

Número de acesso NCBI	Organismo	Nome do gene	QC	Identity
NM_001354473.1	<i>Homo sapiens</i>	Homo sapiens serine rich and transmembrane domain containing 2 (<i>SERTM2</i>), mRNA	*	*
NM_001354555.1	<i>Rattus norvegicus</i>	serine rich and transmembrane domain containing 2 (<i>Sertm2</i>)	8%	88%
NM_001354493.1	<i>Mus musculus</i>	RIKEN cDNA A730046J19 gene (A730046J19Rik)	8%	89%
NM_001354520.1	<i>Bos taurus</i>	serine rich and transmembrane domain containing 2 (<i>SERTM2</i>)	54%	81%
NM_001354500.1	<i>Sus scrofa</i>	serine rich and transmembrane domain containing 2 (<i>SERTM2</i>)	96%	82%
XM_022416152.1	<i>Canis familiaris lupus</i>	PREDICTED:serine rich and transmembrane domain containing 2 (<i>SERTM2</i>)	98%	80%
XM_024353198.1	<i>Pan troglodytes</i>	PREDICTED: serine-rich and transmembrane domain-containing protein 1-like (LOC100608658)	99%	99%
XM_024927345.1	<i>Pan paniscus</i>	PREDICTED: serine rich and transmembrane domain containing 2 (<i>SERTM2</i>)	99%	99%
XR_177245.1	<i>Gorilla gorilla</i>	PREDICTED: uncharacterized LOC101143763 (LOC101143763), ncRNA	99%	99%
XM_025164577.1	<i>Python bivittatus</i>	serine rich and transmembrane domain containing2 (<i>SERTM2</i>), mRNA		
XM_025150255.1	<i>Gallus gallus</i>	serine rich and transmembrane domain containing 2 (<i>SERTM2</i>), mRNA		
XM_001344969.6	<i>Danio rerio</i>	serine rich and transmembrane domain containing 1 (<i>sertm1</i>), mRNA		
NM_177854.4	<i>Mus musculus</i>	serine rich and transmembrane domain containing 1 (<i>Sertm1</i>), mRNA		
NM_203451.2	<i>Homo sapiens</i>	serine rich and transmembrane domain containing 1 (<i>SERTM1</i>), mRNA		

Uma vez que o lncRNA *LINC00890* foi anotado recentemente como um gene codificante de proteína e a fim de verificar a similaridade da porção predita como codificante, o alinhamento foi refeito, considerando a sequência proteica apenas de mamíferos (Figura 7.19). Observou-se alta similaridade entre os organismos analisados, indicando uma alta conservação na sequência da proteína. Ainda, se nota maior identidade com primatas não-humanos. Interessantemente, a identidade com ratos e camundongos foi alta, correspondendo à identidade observada na análise anterior, mas agora com 100% de cobertura. Os apenas 8% de cobertura da sequência nucleotídica observados anteriormente (análise do mRNA) correspondem justamente à essa região codificante, indicando a conservação entre as espécies. A alta identidade observada na sequência codificante é possivelmente o motivo da reanotação do gene *LINC00890/SERTM2* como codificante.

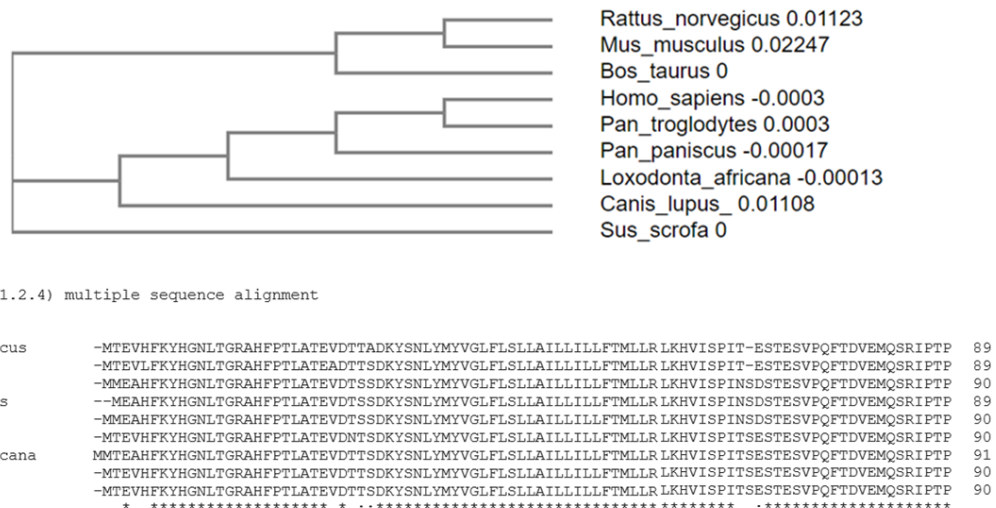


Figura 7.19 Árvore filogenética e alinhamento das sequências proteicas de *SERTM2* em diferentes mamíferos. Ambos foram gerados pela ferramenta CLUSTAL Omega. Árvore filogenética sem correção de distância. Asteriscos (*) representam conservação do aminoácido em todos os organismos analisados.

A plataforma de banco de dados Ensembl traz 33 espécies anotadas que possuem ortólogos de *SERTM2* humano em nível proteico, considerando 100 espécies entre vertebrados e invertebrados. As 33 espécies pertencem aos mamíferos placentários, conforme Tabela 7.2, confirmando a conservação desse gene no nível da Classe dos mamíferos.

Tabela 7.2 Número de ortólogos de *SERTM2*, em nível protéico. Dados da plataforma Ensembl (acesso 17/09/18). Ortólogos 1:1, apenas uma cópia é encontrada em cada espécie.

Espécies	Ortólogos 1:1	Sem ortólogos
Primatas (23 espécies)	17	6
Roedores e espécies afim (24 espécies)	13	11
Laurasiatheria (17 espécies) Carnívoros, ungulados e insetívoros	3	14
Mamíferos placentários (69 espécies)	33	36
Sauropsida (7 espécies) Aves e répteis	0	7
Peixes (11 espécies)	0	11
Todos (100 espécies) Incluindo invertebrados	33	67

Apesar de alguns indícios *in silico* de conservação da possível porção codificante de *LINC00890*, mais dados são necessários para que se possa inferir sobre sua função, e essa investigação está prevista como uma perspectiva futura deste trabalho.

7.7 Perspectivas para a análise do gene *LINC00890*

Ensaio funcionais com as linhagens H1 iLINC, PLZ iLINC e H1 LINC-mut estão previstas para a continuidade deste trabalho. A linhagem iLINC-mut vai ser utilizada em experimentos de diferenciação cardíaca a fim de se estudar a participar da CDS putativa do gene *LINC00890* no fenótipo observado após a indução da expressão desse gene, considerando seu potencial codificante. Além disso, novas linhagens serão construídas usando as células H1 LINC KO e PLZ LINC KO e o sistema de expressão induzida buscando o resgate de fenótipo nas células silenciadas. Análises *in silico* do gene *LINC00890* também serão finalizadas, como análise dos genes vizinhos e sua expressão, e possíveis alvos e redes gênicas durante a diferenciação cardiomiogênica.

8 DISCUSSÃO

Células cardíacas derivadas *in vitro* de células pluripotentes têm o potencial para serem utilizadas na terapia celular, e também como ferramentas na análise dos processos do desenvolvimento (MURRY; KELLER, 2008). Estudos de embriologia utilizando organismos modelo possibilitaram o entendimento sobre a diferenciação linhagem específica de CTP, desvendando vias de sinalização capazes de controlar o desenvolvimento cardiovascular e permitindo a diferenciação *in vitro* de cardiomiócitos a partir de CTP (KATTMAN *et al.*, 2011; KEHAT *et al.*, 2001; LAFLAMME *et al.*, 2007; MURRY; KELLER, 2008; XU *et al.*, 2002).

No presente trabalho, a análise de mRNAs associados aos polissomos – *polysome profiling* seguido do RNA-seq - foi feita em diferentes pontos da diferenciação cardíaca: desde as células indiferenciadas pluripotentes até cardiomiócitos pulsantes. A expressão de marcadores ao longo da diferenciação foi analisada e controles técnicos foram feitos a fim de garantir a qualidade dos dados gerados (Capítulo I). Essa abordagem proporcionou a obtenção de importantes informações com relação a expressão gênica durante a diferenciação cardiomiogênica de CTP, que puderam ser usadas na investigação de possíveis mecanismos pós-transcricionais (Capítulo II). Os RNAs associados aos polissomos mostraram que podem ilustrar a expressão gênica durante o comprometimento à linhagem cardíaca, e a determinação do fenótipo cardiovascular tem influência massiva da regulação pós-transcricional.

Os dados de RNA-seq também possibilitaram a escolha de alguns genes candidatos na participação da diferenciação cardiomiogênica: *RXRG*, *SHISA3*, *CSDC2* e o lncRNA *LINC00890*. Linhagens de super expressão induzida com doxiciclina foram criadas para serem usadas como ferramentas na investigação da atuação de tais genes na diferenciação (Figuras 5.4 e 7.2). Essas linhagens de super expressão foram derivadas das linhagens de hESC H1 e iPSC PLZ e se mantiveram estáveis em cultura, mantendo o estado pluripotente (Figuras 5.6 e 7.3), o que possibilitou sua utilização nos ensaios de diferenciação propostos. Além disso, a indução com doxiciclina e, portanto, a super expressão dos genes em estudo, também foi capaz de manter a pluripotência das células, indicando que tais genes não parecem participar da saída do estado pluripotente.

O gene *RXRG* foi escolhido uma vez que teve a expressão aumentada ao longo da diferenciação. Os receptores de ácido retinóico (RARs e RXRs) atuam como heterodímeros, ligando-se a elementos responsivos no DNA e desencadeando a transcrição de genes (KREZEL *et al.*, 1996; MANGELSDORF *et al.*, 1992). A sinalização por ácido retinóico participa do desenvolvimento embrionário e cardíaco (revisado por STEFANOVIC; ZAFFRAN, 2017) e sua modulação já foi mostrada como participante da indução de células cardíacas atriais (DEVALLA *et al.*, 2015; LEE *et al.*, 2017; XAVIER-NETO *et al.*, 1999; ZHANG *et al.*, 2011). Ainda, os receptores RXRs também podem formar heterodímeros com outros receptores nucleares, permitindo sua atuação em diferentes vias de sinalização, como com os receptores de hormônio de tireóide (TRs), receptores ativados por proliferador de peroxissomos (PPARs) e receptores de vitamina D (VDRs) (TALLAFUSS *et al.*, 2006). A expressão de *RXRG* já foi documentada durante a diferenciação de cardiomiócitos a partir de CTP (LIU *et al.*, 2017b; XU *et al.*, 2009) e também já foi diretamente associada a via de sinalização de PPAR (XU *et al.*, 2009). Essa via é importante reguladora do metabolismo de ácidos graxos em tecidos com alta taxa de oxidação desses compostos, como o do coração. Os heterodímeros PPARa/RXRs regulam a expressão de genes envolvidos na ativação e na oxidação de ácidos graxos, e na internalização de ácidos graxos pela mitocôndria (LOPASCHUK; JASWAL, 2010). Diante disso, hipotetizou-se sobre a atuação da expressão de *RXRG* durante o processo de diferenciação cardíaca e como ela poderia influenciar a identidade dos cardiomiócitos e/ou o metabolismo celular, por exemplo. Dois clones da linhagem criada H1 iRXRG foram testados no protocolo de diferenciação cardíaca, porém não foram capazes de gerar cardiomiócitos eficientemente (Figura 5.7). Por este motivo, a investigação da participação de *RXRG* na diferenciação cardíaca foi interrompida. Alguns indícios nos experimentos realizados levaram a crer que realmente haja algum efeito na expressão de *RXRG* e que a investigação deveria ser continuada em um momento oportuno.

Outro gene selecionado para a investigação durante a diferenciação cardiomiogênica foi o *SHISA3*. Estudos em *Xenopus sp* (YAMAMOTO *et al.*, 2005) e camundongos (FURUSHIMA *et al.*, 2007) identificaram o papel de *SHISA3* como antagonista das vias Wnt e FGF. Mais recentemente, um estudo propôs a atuação de *SHISA3* como supressor tumoral, usando como modelo células humanas tumorais de pulmão (CHEN *et al.*, 2014). *SHISA3* atuaria interagindo com os receptores WNT Frizzled (FZD) no retículo endoplasmático, inibindo seu transporte até a membrana. Desta maneira, regularia negativamente a via Wnt, uma vez que seu receptor não

estaria disponível na superfície da célula. A via de sinalização Wnt é essencial e cuidadosamente controlada durante a diferenciação cardiomiogênica (KWON *et al.*, 2007; LINDSLEY *et al.*, 2006). Por isso, hipotetizou-se que SHISA3 poderia ter um papel na regulação da via Wnt durante a cardiomiogênese e, portanto, ensaios de diferenciação cardíaca com as células da linhagem H1 iSHISA3 foram realizados.

A redução sutil observada na porcentagem de cardiomiócitos nas populações que tiveram a expressão de SHISA3 induzida entre os dias 0-1 da diferenciação (Figura 5.8 C) poderia indicar um efeito desfavorável à diferenciação, possivelmente pela inibição de Wnt no início da diferenciação. Entretanto, mais dados são necessários para confirmar tal efeito. Por outro lado, a inibição de Wnt não foi observada após a indução de SHISA3 no dia 3 da diferenciação, ilustrada pela falha na recuperação da eficiência da diferenciação após a retirada do inibidor XAV939 (Figura 5.8 D e E). O protocolo de diferenciação utilizado nesses experimentos foi o em monocamada, que funciona através da ativação de Wnt no dia 1, com o composto CHIR9921; e da inibição de Wnt no dia 3, com o composto XAV939 (Figura 1.4). A atuação de CHIR9921 se dá pela inibição de GSK3 no complexo de degradação de β -catenina, em um mecanismo independente dos receptores FDZ. Da mesma maneira, a atuação de XAV939 é independente dos receptores FDZ uma vez que atua estabilizando a axina (também componente do complexo de degradação de β -catenina) através do bloqueio da poli-ADP-ribosilação pelas enzimas tanquirase 1 e 2. Por este motivo, o modelo de diferenciação utilizando esses compostos talvez não seja o mais adequado para se investigar o papel de SHISA3, uma vez que a ação altamente específica daqueles se sobreponha a qualquer tipo de interferência que SHISA3 possa fazer na via Wnt.

O gene *CSDC2* se mostrou especificamente expresso nas células já comprometidas à linhagem cardíaca a partir do dia 9 (Figura 5.3). Proteínas da família das *cold shock* (do inglês, choque térmico) estão envolvidas na regulação de RNAs e na regulação pós-transcricional da expressão gênica. As sequências reconhecidas pelo domínio CSD (*cold shock domain*) são muito variáveis, resultando em um amplo repertório de RNA alvos (MIHAILOVICH *et al.*, 2010). Uma proteína da família CSD, a YB-1, parece estar envolvida na patologia da disfunção cardíaca pós-implante. A fibrose perivascular e intersticial são componentes importantes dessa disfunção, e a expressão *de novo* de α -SMA (codificada pelo gene *ACTA2*) em fibroblastos cardíacos é um indicativo da resposta ao estresse mecânico e remodelamento do tecido através da deposição de colágeno. YB-1 se liga ao mRNA de *ACTA2* e regula a eficiência

traducional. Portanto, mudanças na localização sub celular e afinidade de ligação ao RNA das isoformas de YB-1 parecem estar diretamente relacionadas à disfunção cardíaca pós-implante (DAVID *et al.*, 2012).

Frente a grande influência da regulação pós-transcricional durante a diferenciação cardíaca (Capítulo II) e o interesse do grupo de pesquisa na regulação de RNAs, a *CSDC2* foi escolhida para ter seu papel investigado. A *CSDC2*, também conhecida como PIPPin, foi descrita apenas em tecido cerebral de murinos (NASTASI *et al.*, 1999), e já demonstrou-se a importância dos três domínios putativos de ligação ao RNA (PIP1, PIP2 e CSD) para a capacidade de ligação ao RNA (RAIMONDI *et al.*, 2003). A indução de *CSDC2* entre os dias 7 e 10 da diferenciação proporcionou um aumento significativo na expressão de marcadores cardíacos, enquanto os genes marcadores de células musculares lisas e endoteliais (*ACTA2* e *VWF*) não tiveram sua expressão alterada (Figura 5.9). Esses resultados indicam uma possível atuação do gene *CSDC2* em promover a diferenciação cardíaca, contribuindo na diferenciação de cardiomiócitos. Experimentos e análises complementares precisam ser realizados a fim de se inferir sobre os mecanismos que regem a atuação de *CSDC2* na diferenciação de cardiomiócitos. Nesse sentido, a linhagem H1 i*CSDC2*-del foi criada e possibilitará avaliar a participação dos motivos de ligação ao RNA da proteína codificada por esse gene. Os experimentos de diferenciação cardíaca com essas células ainda não foram realizados e estão nas perspectivas do grupo.

Os RNAs não-codificantes longos (lncRNAs), caracterizados por serem maiores que 200 nucleotídeos e não serem traduzidos em proteínas (PIRCHER; GEBETSBERGER; POLACEK, 2014), representam um grupo de transcritos que têm sido amplamente estudados e suas funções têm ganhado destaque em diversos processos biológicos (BUNCH, 2018). Na última década, trabalhos vêm demonstrando o potencial dos lncRNAs em participar da regulação da expressão gênica em diferentes níveis, incluindo a organização da cromatina, a regulação transcricional e também em processos pós-transcricionais (ANGRAND *et al.*, 2015). Além disso, os lncRNAs podem contribuir para a especificação da identidade celular (revisado por LOPEZ-PAJARES, 2016), incluindo a diferenciação cardiomiogênica.

Os dados de RNA-seq descritos no Capítulo I também possibilitaram a análise de lncRNAs expressos nos diferentes estágios da cardiomiogênese de CTP (Capítulo IV). Os resultados mostraram que os lncRNAs possuem um padrão de expressão temporal ao longo da diferenciação cardíaca, tanto os lncRNAs da fração livre quanto os associados aos polissomos. Estes últimos levantam hipóteses interessantes

quanto ao papel que esses transcritos poderiam estar exercendo, uma vez que supostamente não são traduzidos em proteínas. A fim de melhor investigar a influência dos lncRNAs na diferenciação celular e buscar desvendar funções específicas de lncRNAs individuais, o gene *LINC00890* foi escolhido para avaliação funcional na cardiomiogênese.

O lncRNA *LINC00890* se apresentou expresso em progenitores cardíacos (D9) e teve a expressão reduzida em cardiomiócitos (D15) (Figura 7.1). Não há nenhuma informação na literatura com relação a sua função, apenas dados de expressão em diferentes tecidos adultos humanos, os quais mostram a expressão de *LINC00890* em endométrio e próstata (FAGERBERG *et al.*, 2014). O *LINC00890* apareceu associado aos polissomos nos nossos dados de RNA-seq, indicando um possível papel na regulação da tradução de mRNAs ou até mesmo como codificante de peptídeos. Recentemente, uma CDS putativa foi anotada na sequência de *LINC00890* e chamada de *SERTM2* (*serine rich and transmembrane domain containing 2*). A sequência proteica de *SERTM2* possui alta similaridade com sequências de outros organismos (Figura 7.18), principalmente mamíferos, indicando a conservação dessa sequência e um possível papel biológico ainda não descrito.

Os experimentos de diferenciação cardíaca com a linhagem H1 iLINC mostraram que o *LINC00890* parece agir de alguma maneira nesse processo, fato evidenciado tanto pela redução de cardiomiócitos quando o gene foi induzido durante toda a diferenciação (D0-15) (Figura 7.4), quanto pelo aumento de marcadores cardíacos quando foi induzido na etapa de progenitores (D5-10) (Figura 7.5). Essas evidências suportaram a hipótese de que o gene *LINC00890* poderia atuar de diferentes maneiras em momentos distintos da diferenciação: (1) como um inibidor da formação de mesoderme cardíaca no início da diferenciação e (2) promovendo a diferenciação com favorecimento de cardiomiócitos atriais a partir de progenitores cardíacos. A segunda parte da hipótese começou a ser testada e os resultados indicaram um aumento em alguns marcadores cardíacos, aumento na frequência de batimentos e aumento no número de células atriais após a indução da expressão de *LINC00890* entre D5-10. Ensaio moleculares isolados não são suficientes para confirmar a identidade de cardiomiócitos, enquanto os ensaios funcionais representam abordagens mais precisas (MUMMERY *et al.*, 2012). Apenas o marcador de células atriais MYL7 se mostrou significativamente aumentado nas células induzidas (DOX D5-10) (Figura 7.6), entretanto, ensaios funcionais de eletrofisiologia identificaram um maior número de células atriais após a indução com doxiciclina (Figura 7.9).

Cardiomiócitos induzidos DOX D5-10 também apresentaram maior frequência de batimentos (Figura 7.8), corroborando com dados de que cardiomiócitos induzidos com ácido retinóico e que apresentaram diversas características atriais, apresentaram aumento na frequência de batimentos (DEVALLA *et al.*, 2015; LEE *et al.*, 2017). Em conjunto, esses resultados indicam um possível papel do *LINC00890* na determinação da identidade de cardiomiócitos.

As células ventriculares diferenciadas sob indução de doxiciclina D5-10 também apresentaram diferenças em alguns parâmetros eletrofisiológicos. A duração do potencial de ação em 90% de repolarização (APD90), a inclinação máxima do traço ascendente do potencial de ação (dV/dtmax) e a amplitude do potencial de ação (Amp) se mostraram aumentadas nessas células quando comparadas com as células controle sem indução (Figura 7.9). O aumento nesses parâmetros já foi documentado quando cardiomiócitos foram cultivados por mais tempo, indicando a maturação funcional dessas células (BEN-ARI *et al.*, 2016; LUNDY *et al.*, 2013). Esses resultados indicam, então, que os cardiomiócitos ventriculares diferenciados após a indução de *LINC00890* entre os dias 5-10 da diferenciação apresentaram indícios de maior maturação quando comparados com as células controle sem indução e que o *LINC00890* poderia exercer um papel na maturação de cardiomiócitos.

Por sua vez, o silenciamento de *LINC00890* mostrou resultados interessantes quanto à diferenciação cardiomiogênica. Ambas a linhagens silenciadas H1 LINC KO e PLZ LINC KO continuaram com suas características pluripotentes, indicando que *LINC00890* não é essencial na manutenção do estado pluripotente de células-tronco. A estratégia de silenciamento utilizada nesse estudo propôs a deleção completa do gene *LINC00890*. Por este motivo, esperava-se a expressão nula desse gene ao longo de toda a diferenciação cardíaca nas células identificadas como silenciadas. A linhagem PLZ LINC KO apresentou níveis de expressão de *LINC00890* nos dias 5, 10 e 15 da diferenciação, mas que ainda representaram valores muito mais baixos dos expressos pelas células WT. Essa pequena expressão detectada possivelmente é derivada de células WT contaminantes na cultura, visto que o método de seleção clonal não é completamente eficiente. Tais células contaminantes podem interferir na interpretação dos resultados, os quais devem ser analisados com muita cautela. A linhagem PLZ LINC KO, quando diferenciada à cardiomiócitos, mostrou a expressão de marcadores cardíacos precoces, como EOMES e GATA4. Isso indicaria o início do comprometimento dessas células a linhagem cardíaca, com as possibilidades de interferência das células WT ou de não-interferência do silenciamento de *LINC00890*

nessa etapa. Por outro lado, os marcadores cardíacos tardios NKX2-5 e TNNT2 se mostraram reduzidos nas células KO, assim como a ausência de batimentos espontâneos, indicando certo grau de redução de eficiência de diferenciação. Esse experimento foi feito em apenas uma réplica, e necessita de mais dados para maiores conclusões.

Por sua vez, células da linhagem H1 LINC KO, quando submetidas à diferenciação cardiomiogênica, não foram capazes de se diferenciar em cardiomiócitos em três réplicas experimentais (Figura 7.14). Os marcadores cardíacos precoces e tardios apresentaram níveis de expressão muito mais baixos quando comparados aos das células WT. Apesar da diferenciação cardíaca final ter sido comprometida nessas células, a diferenciação até o estágio de mesoderme aconteceu normalmente, indicada pela expressão do marcador de células mesodermiais T, que foi similar entre as células WT e KO. Com relação à diferenciação de outras células derivadas de progenitores cardíacos, avaliamos a expressão do gene ACTA2, marcador de células musculares lisas, o qual apresentou maior expressão nas células KO, indicando que possivelmente o silenciamento de *LINC00890* parece prejudicar especificamente a diferenciação de cardiomiócitos e não de outros derivados dos mesmos progenitores. Interessantemente, nas células da linhagem H1 iLINC quando diferenciadas à cardiomiócitos e com a expressão de *LINC00890* induzida (D5-10), a expressão do gene ACTA2 se mostrou reduzida (Figura 7.5), indicando nesse caso um favorecimento de cardiomiócitos a partir dos progenitores cardíacos. Em conjunto, esses dados apontam para uma possível função de *LINC00890* na diferenciação específica de cardiomiócitos.

Recentemente, a participação do lncRNA chamado de linc1405 foi revelada em células murinas (GUO *et al.*, 2018). O éxon 2 do linc1405 especificamente se liga e forma um complexo com o fator de transcrição Eomes e os fatores de modificação de histonas WDR5 e GCN5. Tal complexo se associa a uma região *enhancer* do gene *Mesp1* controlando sua transcrição e o comprometimento cardíaco. Esse exemplo, juntamente com os lncRNAs *Braveheart* (KLATTENHOFF *et al.*, 2013) e *Fendrr* (GROTE *et al.*, 2013) em murinos, e CARMEM (OUNZAIN *et al.*, 2015) e HBL1 (LIU *et al.*, 2017a) em humanos, demonstram o grande potencial que os lncRNAs têm na regulação dos processos de desenvolvimento cardíaco. A principal atuação desses lncRNAs já descritos envolve a regulação epigenética e interferência na transcrição dos genes alvo. No caso do *LINC00890*, nossa hipótese estaria relacionada a

regulação a nível pós-transcricional, uma vez que ele foi identificado associado aos polissomos.

A associação de lncRNAs com ribossomos já foi documentada (CARLEVARO-FITA *et al.*, 2016; DALLAGIOVANNA *et al.*, 2017; INGOLIA; LAREAU; WEISSMAN, 2011; VAN HEESCH *et al.*, 2014), e pode indicar o papel dessas moléculas outro que o de codificação de proteínas. Poderiam estar atuando na regulação da velocidade e especificidade da maquinaria de tradução ou pareando e regulando especificamente mRNAs, como no *splicing*, no transporte e na degradação (ANGRAND *et al.*, 2015; KARAPETYAN *et al.*, 2013; PIRCHER; GEBETSBERGER; POLACEK, 2014). O lncRNA chamado TINCR (*terminal differentiation-induced ncRNA*) foi identificado como regulador da diferenciação epidermal em células humanas, atuando por mecanismos pós-transcricionais (KRETZ *et al.*, 2013). A deficiência de TINCR impediu a diferenciação terminal da ultraestrutura da epiderme. Foi demonstrado sua interação com diversos mRNAs relacionados a diferenciação, através de um motivo específico na sequência de TINCR. TINCR formava, então, um complexo com a proteína de decaimento de RNA STAU-1 e mediava a estabilização de seus mRNAs alvo (KRETZ *et al.*, 2013).

Os resultados que obtivemos até o momento mostram que o *LINC00890* atua de alguma maneira interferindo na diferenciação de cardiomiócitos. Sua expressão no estágio de progenitores cardíacos parece ajudar na determinação do subtipo de cardiomiócitos e na maturação dos mesmos. Ainda, sua expressão se mostrou essencial durante a diferenciação cardíaca, uma vez que as células silenciadas foram incapazes de diferenciar. Experimentos adicionais precisam ser realizados para que se possa inferir sobre o mecanismo que o *LINC00890* esteja atuando na diferenciação, e se possa esclarecer a importância dele na cardiomiogênese. As perspectivas deste trabalho incluem a avaliação da função da CDS putativa na sequência de *LINC00890*, a caracterização dos mutantes a nível molecular e a inferência sobre as redes gênicas nas quais *LINC00890* poderia estar atuando.

9 CONCLUSÕES

Neste trabalho, o RNA-seq das frações de RNA livres e associados aos polissomos ao longo da diferenciação cardiomiogênica permitiu uma análise global – Capítulos II e IV, e uma análise pontual de genes através da escolha de candidatos a participantes da diferenciação – Capítulos III e V. A técnica de *polysome profiling* e a comparação dos RNAs livres e associados aos polissomos possibilitaram a inferência sobre a crucial participação da regulação pós-transcricional na diferenciação cardiomiogênica, agindo nos genes codificantes e também em genes não-codificantes.

O objetivo inicial deste trabalho foi diferenciar CTE humanas à cardiomiócitos e isolar os RNAs livre e associados aos polissomos. Com o protocolo utilizado, obtivemos células cerca de 60% cTnT positivas, indicando a eficiência da diferenciação em cardiomiócitos. Cinco diferentes pontos durante esse protocolo foram escolhidos para a análise dos RNAs, os quais representam etapas importantes do desenvolvimento cardíaco: pluripotência, sinalização inicial, mesoderme, progenitores cardíacos e cardiomiócitos. Em cada um desses pontos, a técnica de *polysome profiling* foi utilizada e permitiu a identificação das frações de RNAs livre e polissomais e o subsequente isolamento desses RNAs. Com isso, pudemos gerar dados de RNA-seq de alta qualidade, os quais possibilitaram as análises presentes nesta Tese.

Os dados de sequenciamento dos RNAs polissomais mostraram a expressão característica de genes marcadores de desenvolvimento cardíaco, indicando que a análise desses RNAs pode ilustrar a expressão gênica das células durante o comprometimento à linhagem cardíaca. Vias de sinalização e processos biológicos envolvidos com o desenvolvimento embrionário e com a especialização do tecido cardíaco se mostraram enriquecidos quando os dados de RNA-seq foram analisados por Ontologia gênica, demonstrando que o modelo de diferenciação *in vitro* pode mimetizar as etapas do desenvolvimento embrionário e servir como uma ferramenta importante no estudo dos processos moleculares envolvidos.

Também pudemos investigar sobre a importância da regulação pós-transcricional na expressão gênica durante a diferenciação celular, através da análise e comparação dos RNAs livres e associados aos polissomos. Essa regulação parece atuar em um grande número de RNAs, os quais participam de processos biológicos

variados, como o metabolismo celular e o ciclo celular, além de vias de sinalização do desenvolvimento embrionário.

Com a análise dos dados de RNA-seq, pudemos inferir sobre a potencial participação de genes específicos na diferenciação cardíaca. Os genes *RXRG*, *SHISA3* e *CSDC2* foram escolhidos para análise funcional, através da construção de linhagens de superexpressão gênica. Estas linhagens foram derivadas da linhagem de CTE humanas H1 e utilizadas em ensaios de diferenciação cardiomiogênica. Obtivemos resultados promissores com a indução da expressão de *CSDC2* entre os dias 7 e 10 do protocolo de diferenciação, onde os cardiomiócitos finais mostraram uma maior expressão de marcadores cardíacos quando comparados com as células controle. Esse resultado indicou uma possível atuação do gene *CSDC2* em promover a diferenciação cardíaca, contribuindo na diferenciação e/ou maturação de cardiomiócitos.

Ao analisarmos os lncRNAs ao longo da diferenciação cardíaca, observamos que um grande número deles se mostrou expresso diferencialmente entre as etapas da diferenciação. Além disso, também foi possível identificar a associação diferencial de lncRNAs aos polissomos, indicando um potencial papel daqueles como codificantes de proteínas ou atuantes em processos de regulação da tradução. A análise funcional de um lncRNA escolhido, o *LINC00890*, foi iniciada através da construção de linhagens de superexpressão induzida e de silenciamento. Essas células foram submetidas à diferenciação cardíaca e mostraram indícios da atuação deste lncRNA no comprometimento de CTE à cardiomiócitos. Quando o *LINC00890* foi superexpresso entre os dias 5 e 10 do protocolo de diferenciação, mais dos cardiomiócitos derivados mostraram características eletrofisiológicas de células atriais, quando comparados com a condição controle. Ainda, células com o *LINC00890* silenciado não foram capazes de se diferenciar à cardiomiócitos. Esses resultados indicaram um possível papel essencial deste lncRNA na diferenciação de cardiomiócitos e na determinação de seus subtipos.

Em resumo, o presente trabalho mostrou a riqueza dos dados de expressão gênica em larga escala, ilustrando um pouco da complexidade dos processos de regulação durante a diferenciação cardiomiogênica, e investigou sobre o papel dos lncRNAs na diferenciação. Muito ainda há para se explorar nesses dados e continuar contribuindo na elucidação do fascinante processo de diferenciação celular.

10 PERSPECTIVAS GERAIS

A riqueza dos dados gerados neste trabalho permitiu o levantamento de algumas hipóteses que serão investigadas em trabalhos futuros. Por exemplo, experimentos adicionais com o gene *CSDC2* estão planejados em um novo projeto de Doutorado, como mencionado anteriormente. Assim como a continuidade da análise do gene *LINC00890* também está prevista, a fim de complementar os resultados apresentados nesta Tese e, por fim, comporem um manuscrito a respeito da avaliação funcional deste lncRNA. Os dados de RNA-seq continuarão a ser explorados e, sem dúvida, servirão de base para novas hipóteses.

11 REFERÊNCIAS

- ADELI, K. Translational control mechanisms in metabolic regulation: critical role of RNA binding proteins, microRNAs, and cytoplasmic RNA granules. **AJP: Endocrinology and Metabolism**, v. 301, n. 6, p. E1051–E1064, 2011.
- ANGRAND, P. O. *et al.* The role of long non-coding RNAs in genome formatting and expression. **Frontiers in Genetics**, v. 6, n. APR, p. 1–12, 2015.
- ARIF, W.; DATAR, G.; KALSOTRA, A. Intersections of Post-Transcriptional Gene Regulatory Mechanisms with Intermediary Metabolism. **Biochimica et Biophysica Acta (BBA) - Gene Regulatory Mechanisms**, v. 1860, p. 349–362, 2017.
- BARNETT, P.; DEN BOOGAARD, M. VAN; CHRISTOFFELS, V. **Localized and Temporal Gene Regulation in Heart Development**. [s.l.] Elsevier Inc., 2012. v. 100
- BEN-ARI, M. *et al.* Developmental changes in electrophysiological characteristics of human-induced pluripotent stem cell–derived cardiomyocytes. **Heart Rhythm**, v. 13, n. 12, p. 2379–2387, 2016.
- BEN-DAVID, U.; KOPPER, O.; BENVENISTY, N. Expanding the boundaries of embryonic stem cells. **Cell Stem Cell**, v. 10, n. 6, p. 666–677, 2012.
- BEQQALI, A. *et al.* Genome-Wide Transcriptional Profiling of Human Embryonic Stem Cells Differentiating to Cardiomyocytes. **Stem Cells**, v. 24, n. 8, p. 1956–1967, 2006.
- BERGMANN, O. *et al.* Dynamics of Cell Generation and Turnover in the Human Heart. **Cell**, v. 161, n. 7, p. 1566–1575, 2015.
- BHAGAVATI, S. Stem Cell Therapy: Challenges Ahead. **Indian Journal of Pediatrics**, v. 82, n. 3, p. 286–291, 2015.
- BLANCO, S. *et al.* Stem cell function and stress response are controlled by protein synthesis. **Nature**, v. 534, n. 7607, p. 335–340, 2016.
- BOLLINI, S.; SMART, N.; RILEY, P. R. Resident cardiac progenitor cells: At the heart of regeneration. **Journal of Molecular and Cellular Cardiology**, v. 50, n. 2, p. 296–303, 2011.
- BONDUE, A. *et al.* Mesp1 acts as a master regulator of multipotent cardiovascular progenitor specification. **Cell Stem Cell**, v. 3, n. 1, p. 69–84, 2008.
- BOOGERD, C. J. J.; MOORMAN, A. F. M.; BARNETT, P. Protein interactions at the heart of cardiac chamber formation. **Annals of Anatomy**, v. 191, n. 6, p. 505–517, 2009.
- BOYER, L. A. *et al.* Core transcriptional regulatory circuitry in human embryonic stem cells. **Cell**, v. 122, n. 6, p. 947–956, 2005.
- BRADÉ, T. *et al.* Embryonic heart progenitors and cardiogenesis. **Cold Spring Harb Perspect Med**, v. 3, p. 1–18, 2013.
- BRANDENBURG, S. *et al.* The molecular and functional identities of atrial cardiomyocytes in health and disease. **Biochimica et Biophysica Acta - Molecular Cell Research**, v. 1863, n. 7, p. 1882–1893, 2016.
- BRENGUES, M.; TEIXEIRA, D.; PARKER, R. Movement of eukaryotic mRNAs between polysomes and cytoplasmic processing bodies. **Science**, v. 310, n. 5747, p. 486–489, 2005.

- BRUNEAU, B. G. Signaling and Transcriptional Networks in Heart Development and Regeneration. **Cold Spring Harbor Laboratory Perspectives in Biology**, v. 5, 2013.
- BUCHAN, J. R. MRNP granules Assembly, function, and connections with disease. **RNA Biology**, v. 11, n. 8, p. 1019–1030, 2014.
- BUNCH, H. Gene regulation of mammalian long non-coding RNA. **Molecular genetics and genomics : MGG**, v. 293, n. 1, p. 1–15, 2018.
- BURRIDGE, P. W. *et al.* Production of de novo cardiomyocytes: Human pluripotent stem cell differentiation and direct reprogramming. **Cell Stem Cell**, v. 10, n. 1, p. 16–28, 2012.
- CABILI, M. N. *et al.* Integrative annotation of human large intergenic noncoding RNAs reveals global properties and specific subclasses. **Genes & Development**, v. 25, p. 1915–1927, 2011.
- CARLEVARO-FITA, J. *et al.* Cytoplasmic long noncoding RNAs are frequently bound to and degraded at ribosomes in human cells. **Rna**, v. 22, n. 6, p. 867–882, 2016.
- CARTHEW, R. W.; SONTHEIMER, E. J. Origins and Mechanisms of miRNAs and siRNAs. **Cell**, v. 136, n. 4, p. 642–655, 2009.
- CECH, T. R.; STEITZ, J. A. The noncoding RNA revolution - Trashing old rules to forge new ones. **Cell**, v. 157, n. 1, p. 77–94, 2014.
- CHEN, C. C. *et al.* Shisa3 is associated with prolonged survival through promoting β -catenin degradation in lung cancer. **American Journal of Respiratory and Critical Care Medicine**, v. 190, n. 4, p. 433–444, 2014.
- CHORGHADE, S. *et al.* Poly(A) tail length regulates PABPC1 expression to tune translation in the heart. **eLife**, v. 6, p. e24139, 2017.
- CHUNG, S. *et al.* Mitochondrial oxidative metabolism is required for the cardiac differentiation of stem cells. **Nature Clinical Practice Cardiovascular Medicine**, v. 4, n. February, p. S60–S67, 2007.
- COLLIER, A. E.; WEK, R. C.; SPANDAU, D. F. Human keratinocyte differentiation requires translational control by the eIF2 α kinase GCN2. **Journal of Investigative Dermatology**, 2017.
- COSTELLO, I. *et al.* The T-box transcription factor Eomesodermin acts upstream of Mesp1 to specify cardiac mesoderm during mouse gastrulation. **Nat Cell Biol**, v. 13, n. 9, p. 1084–1091, 2015.
- DALLAGIOVANNA, B. *et al.* lncRNAs are associated with polysomes during adipose-derived stem cell differentiation. **Gene**, v. 610, p. 103–111, 2017.
- DAVID, J. J. *et al.* Y-box binding protein-1 implicated in translational control of fetal myocardial gene expression after cardiac transplant. **Experimental Biology and Medicine**, v. 237, n. 5, p. 593–607, 2012.
- DE MIGUEL-BERIAIN, I. The ethics of stem cells revisited. **Advanced Drug Delivery Reviews**, v. 82, p. 176–180, 2015.
- DEN HARTOGH, S. C. *et al.* A comprehensive gene expression analysis at sequential stages of *in vitro* cardiac differentiation from isolated MESP1-expressing-mesoderm progenitors. **Scientific Reports**, v. 6, n. 1, 2016.
- DEUTSCH, M.-A.; STURZU, A.; WU, S. M. At a crossroad: cell therapy for cardiac repair. **Circulation research**, v. 112, n. 6, p. 884–90, mar. 2013.

- DEVALLA, H. D. *et al.* Atrial-like cardiomyocytes from human pluripotent stem cells are a robust preclinical model for assessing atrial-selective pharmacology. **EMBO Molecular Medicine**, v. 7, n. 4, p. 394–410, 2015.
- DJEBALI, S. *et al.* Landscape of transcription in human cells. **Nature**, v. 489, n. 7414, p. 101–108, 2012.
- DUNHAM, I. *et al.* An integrated encyclopedia of DNA elements in the human genome. **Nature**, v. 489, n. 7414, p. 57–74, 2012.
- ELLIOTT, D. A. *et al.* NKX2-5eGFP/w hESCs for isolation of human cardiac progenitors and cardiomyocytes. **Nature Methods**, v. 8, n. 12, p. 1037–1040, 2011.
- EVANS, S. M. *et al.* Myocardial lineage development. **Circulation Research**, v. 107, n. 12, p. 1428–1444, 2010.
- EVSEENKO, D. *et al.* Mapping the first stages of mesoderm commitment during differentiation of human embryonic stem cells. **PNAS**, v. 107, n. 31, p. 13742–13747, 2010.
- FAGERBERG, L. *et al.* Analysis of the Human Tissue-specific Expression by Genome-wide Integration of Transcriptomics and Antibody-based Proteomics. **Molecular & Cellular Proteomics**, v. 13, n. 2, p. 397–406, 2014.
- FAYE, M. D.; GRABER, T. E.; HOLCIK, M. Assessment of Selective mRNA Translation in Mammalian Cells by Polysome Profiling. **Journal of Visualized Experiments**, n. 92, p. 1–8, 2014.
- FUCHS, E.; SEGRE, J. A. Stem cells: a new lease on life. **Cell**, v. 100, n. 1, p. 143–155, 2000.
- FURUSHIMA, K. *et al.* Mouse homologues of Shisa antagonistic to Wnt and Fgf signalings. **Developmental Biology**, v. 306, n. 2, p. 480–492, 2007.
- GAN, L.; SCHWENGBERG, S.; DENECKE, B. Transcriptome analysis in cardiomyocyte-specific differentiation of murine embryonic stem cells reveals transcriptional regulation network. **Gene expression patterns**, v. 16, n. 1, p. 8–22, jul. 2014.
- GARCÍA-RÚA, V. *et al.* Increased expression of fatty-acid and calcium metabolism genes in failing human heart. **PLoS ONE**, v. 7, n. 6, p. 5–7, 2012.
- GEHRING, N. H.; WAHLE, E.; FISCHER, U. Deciphering the mRNP Code: RNA-Bound Determinants of Post-Transcriptional Gene Regulation. **Trends in Biochemical Sciences**, v. 42, n. 5, p. 369–382, 2017.
- GEORGE, V.; COLOMBO, S.; TARGOFF, K. L. An early requirement for nkx2.5 ensures the first and second heart field ventricular identity and cardiac function into adulthood. **Developmental Biology**, v. 400, n. 1, p. 10–22, 2015.
- GERLITZ, G.; JAGUS, R.; ELROY-STEIN, O. Phosphorylation of initiation factor-2 α is required for activation of internal translation initiation during cell differentiation. **European Journal of Biochemistry**, v. 269, n. 11, p. 2810–2819, 2002.
- GESSERT, S.; KÜHL, M. The multiple phases and faces of Wnt signaling during cardiac differentiation and development. **Circulation Research**, v. 107, n. 2, p. 186–199, 2010.
- GROTE, P. *et al.* The Tissue-Specific lncRNA Fendrr Is an Essential Regulator of Heart and Body Wall Development in the Mouse. **Developmental Cell**, v. 24, n. 2, p. 206–214, 2013.

- GUIL, S.; ESTELLER, M. RNA-RNA interactions in gene regulation: The coding and noncoding players. **Trends in Biochemical Sciences**, v. 40, n. 5, p. 248–256, 2015.
- GUO, X. *et al.* A Linc1405/Eomes Complex Promotes Cardiac Mesoderm Specification and Cardiogenesis. **Cell Stem Cell**, v. 22, n. 6, p. 893–908.e6, 2018.
- GUTTMAN, M. *et al.* Ribosome profiling provides evidence that large noncoding RNAs do not encode proteins. **Cell**, v. 154, n. 1, p. 240–251, 2013.
- HANGAUER, M. J.; VAUGHN, I. W.; MCMANUS, M. T. Pervasive Transcription of the Human Genome Produces Thousands of Previously Unidentified Long Intergenic Noncoding RNAs. **PLoS Genetics**, v. 9, n. 6, 2013.
- HELLEN, C. U. T.; SARNOW, P. Internal ribosome entry sites in eukaryotic mRNA molecules. *Internal ribosome entry sites in eukaryotic mRNA molecules*. n. 718, p. 1593–1612, 2001.
- HERSHEY, J. W. B.; SONENBERG, N.; MATHEWS, M. B. Principles of Translational Control: An Overview. **Cold Spring Harbor Perspectives in Biology**, v. 4, n. 12, p. a011528–a011528, 2012.
- HEZRONI, H. *et al.* Principles of Long Noncoding RNA Evolution Derived from Direct Comparison of Transcriptomes in 17 Species. **Cell Reports**, v. 11, n. 7, p. 1110–1122, 2015.
- HOCKEMEYER, D. *et al.* Efficient targeting of expressed and silent genes in human ESCs and iPSCs using zinc-finger nucleases. **Nature Biotechnology**, v. 27, n. 9, p. 851–857, 2009.
- IGLESIAS-PLATAS, I.; MONK, D. Nongenomic regulation of gene expression. **Current Opinion in Pediatrics**, v. 28, n. 4, p. 521–528, 2016.
- INGOLIA, N. T.; LAREAU, L. F.; WEISSMAN, J. S. Ribosome profiling of mouse embryonic stem cells reveals the complexity and dynamics of mammalian proteomes. **Cell**, v. 147, n. 4, p. 789–802, 2011.
- INOUE, H. *et al.* iPS cells: a game changer for future medicine. **EMBO Journal**, v. 33, n. 5, p. 409–417, 2014.
- IVANOV, P. *et al.* Angiogenin-Induced tRNA Fragments Inhibit Translation Initiation. **Molecular Cell**, v. 43, n. 4, p. 613–623, 2011.
- JACKSON, R. J.; HELLEN, C. U. T.; PESTOVA, T. V. The mechanism of eukaryotic translation initiation and principles of its regulation. **Nature reviews. Molecular cell biology**, v. 11, n. 2, p. 113–127, 2010.
- JOHNSSON, P. *et al.* Evolutionary conservation of long non-coding RNAs; Sequence, structure, function. **Biochimica et Biophysica Acta - General Subjects**, v. 1840, n. 3, p. 1063–1071, 2014.
- JONAS, S.; IZAURRALDE, E. Towards a molecular understanding of microRNA-mediated gene silencing. **Nature reviews. Genetics**, v. 16, n. 7, p. 421–433, 2015.
- KARAPETYAN, A. R. *et al.* Regulatory roles for long ncRNA and mRNA. **Cancers**, v. 5, n. 2, p. 462–490, 2013.
- KATTMAN, S. J. *et al.* Stage-specific optimization of activin/nodal and BMP signaling promotes cardiac differentiation of mouse and human pluripotent stem cell lines. **Cell Stem Cell**, v. 8, n. 2, p. 228–240, 2011.
- KEDERSHA, N. *et al.* Stress granules and processing bodies are dynamically linked

- sites of mRNP remodeling. **Journal of Cell Biology**, v. 169, n. 6, p. 871–884, 2005.
- KEENE, J. D. RNA regulons: Coordination of post-transcriptional events. **Nature Reviews Genetics**, v. 8, n. 7, p. 533–543, 2007.
- KEHAT, I. *et al.* Human embryonic stem cells can differentiate into myocytes with structural and functional properties of cardiomyocytes. **Journal of Clinical Investigation**, v. 108, n. 3, p. 407–414, 2001.
- KELEN, K. VAN DER *et al.* Translational control of eukaryotic gene expression. **Critical Reviews in Biochemistry and Molecular Biology**, v. 44, n. 4, p. 143–168, 2009.
- KING, H. A.; GERBER, A. P. Translatome profiling: Methods for genome-scale analysis of mRNA translation. **Briefings in Functional Genomics**, p. 1–10, 2014.
- KLATTENHOFF, C. A. *et al.* Braveheart, a long noncoding RNA required for cardiovascular lineage commitment. **Cell**, v. 152, n. 3, p. 570–583, 2013.
- KOLLE, G. *et al.* Deep-transcriptome and ribonome sequencing redefines the molecular networks of pluripotency and the extracellular space in human embryonic stem cells. **Genome Research**, p. 2013–2025, 2014.
- KOMILI, S.; SILVER, P. A. Coupling and coordination in gene expression processes: A systems biology view. **Nature Reviews Genetics**, v. 9, n. 1, p. 38–48, 2008.
- KRETZ, M. *et al.* Control of somatic tissue differentiation by the long non-coding RNA TINCR. **Nature**, v. 493, n. 7431, p. 231–235, 2013.
- KREZEL, W. *et al.* RXR gamma null mice are apparently normal and compound RXR alpha +/-RXR beta -/-RXR gamma -/- mutant mice are viable. **Proceedings of the National Academy of Sciences of the United States of America**, v. 93, n. August, p. 9010–9014, 1996.
- KRICHEVSKY, A M.; METZER, E.; ROSEN, H. Translational control of specific genes during differentiation of HL-60 cells. **The Journal of biological chemistry**, v. 274, n. 20, p. 14295–14305, 1999.
- KURIAN, L. *et al.* Identification of novel long noncoding RNAs underlying vertebrate cardiovascular development. **Circulation**, v. 131, n. 14, p. 1278–1290, 2015.
- KWON, C. *et al.* Canonical Wnt signaling is a positive regulator of mammalian cardiac progenitors. **Proc Natl Acad Sci U S A**, v. 104, n. 26, p. 10894–10899, 2007.
- LAFLAMME, M. A *et al.* Cardiomyocytes derived from human embryonic stem cells in pro-survival factors enhance function of infarcted rat hearts. **Nature biotechnology**, v. 25, n. 9, p. 1015–1024, 2007.
- LEE, J. H. *et al.* Human Pluripotent Stem Cell-Derived Atrial and Ventricular Cardiomyocytes Develop from Distinct Mesoderm Populations. **Cell Stem Cell**, v. 21, n. 2, p. 179–194.e4, 2017.
- LEWIS, S. E.; KELLY, F. J.; GOLDSPINK, D. F. Pre- and post-natal growth and protein turnover in smooth muscle, heart and slow- and fast-twitch skeletal muscles of the rat. **The Biochemical journal**, v. 217, n. 2, p. 517–526, 1984.
- LI, Y.; LIN, B.; YANG, L. Comparative Transcriptomic Analysis of Multiple Cardiovascular Fates from Embryonic Stem Cells Predicts Novel Regulators in Human Cardiogenesis. **Scientific Reports**, v. 5, n. 1, 2015.
- LIAN, X. *et al.* Directed cardiomyocyte differentiation from human pluripotent stem cells

- by modulating Wnt/ β -catenin signaling under fully defined conditions. **Nature Protocols**, v. 8, n. 1, p. 162–175, 2012.
- LIAO, Y. *et al.* The Cardiomyocyte RNA-Binding Proteome: Links to Intermediary Metabolism and Heart Disease. **Cell Reports**, v. 16, n. 5, p. 1456–1469, 2016.
- LIAO, Y.; SMYTH, G. K.; SHI, W. The Subread aligner: Fast, accurate and scalable read mapping by seed-and-vote. **Nucleic Acids Research**, v. 41, n. 10, 2013.
- LINDSLEY, R. C. *et al.* Canonical Wnt signaling is required for development of embryonic stem cell-derived mesoderm. **Development**, v. 133, n. 19, p. 3787–3796, 2006.
- LIU, J. *et al.* HBL1 Is a Human Long Noncoding RNA that Modulates Cardiomyocyte Development from Pluripotent Stem Cells by Counteracting MIR1. **Developmental Cell**, v. 42, n. 4, p. 333–348.e5, 2017a.
- LIU, Q. *et al.* Genome-Wide Temporal Profiling of Transcriptome and Open Chromatin of Early Cardiomyocyte Differentiation Derived from hiPSCs and hESCs. **Circulation Research**, v. 121, n. 4, p. 376–391, 2017b.
- LOH, K. M. M. *et al.* Mapping the Pairwise Choices Leading from Pluripotency to Human Bone, Heart, and Other Mesoderm Cell Types. **Cell**, v. 166, n. 2, p. 451–468, 2016.
- LOMBARDO, A. *et al.* Site-specific integration and tailoring of cassette design for sustainable gene transfer. **Nature Methods**, v. 8, n. 10, p. 861–869, 2011.
- LOPASCHUK, G. D.; JASWAL, J. S. Energy Metabolic Phenotype of the Cardiomyocyte During Development, Differentiation, and Postnatal Maturation. **Journal of Cardiovascular Pharmacology**, v. 56, n. 2, p. 130–140, 2010.
- LOPEZ-PAJARES, V. Long non-coding RNA regulation of gene expression during differentiation. **Pflugers Archiv European Journal of Physiology**, v. 468, n. 6, p. 971–981, 2016.
- LUI, K. O.; ZANGI, L.; CHIEN, K. R. Cardiovascular regenerative therapeutics via synthetic paracrine factor modified mRNA. **Stem Cell Research**, v. 13, n. 3, p. 693–704, 2014.
- LUNDY, S. D. *et al.* Structural and Functional Maturation of Cardiomyocytes Derived from Human Pluripotent Stem Cells. **Stem Cells and Development**, v. 22, n. 14, p. 1991–2002, 2013.
- LUO, S. *et al.* Divergent lncRNAs regulate gene expression and lineage differentiation in pluripotent cells. **Cell Stem Cell**, v. 18, n. 5, p. 637–652, 2016.
- MANGELSDORF, D. J. *et al.* Characterization of three RXR genes that mediate the action of 9-cis retinoic acid. **Genes and Development**, v. 6, n. 3, p. 329–344, 1992.
- MANSUR, A. DE P.; FAVARATO, D. Mortalidade por Doenças Cardiovasculares no Brasil e na Região Metropolitana de São Paulo : Atualização 2011. **Arq Bras Cardiol**, 2012.
- MARCON, B. H. *et al.* Downregulation of the protein synthesis machinery is a major regulatory event during early adipogenic differentiation of human adipose-derived stromal cells. **Stem Cell Research**, v. 25, p. 191–201, 2017.
- MASEK, T.; VALASEK, L.; POSPISEK, M. Polysome Analysis and RNA Purification from Sucrose Gradients. **Methods in Molecular Biology**, v. 703, p. 293–309, 2011.

- MATA, J.; MARGUERAT, S.; BÄHLER, J. Post-transcriptional control of gene expression: A genome-wide perspective. **Trends in Biochemical Sciences**, v. 30, n. 9, p. 506–514, 2005.
- MATOULKOVA, E. *et al.* The role of the 3' untranslated region in post-transcriptional regulation of protein expression in mammalian cells. **RNA Biology**, v. 9, n. 5, p. 563–576, 2012.
- MATTICK, J. S.; MAKUNIN, I. V. Non-coding RNA. **Human molecular genetics**, v. 15 Spec No, n. 1, p. 17–29, 2006.
- MAYUZUMI, N. *et al.* Effects of ultraviolet B irradiation, proinflammatory cytokines and raised extracellular calcium concentration on the expression of ATP2A2 and ATP2C1. **British Journal of Dermatology**, v. 152, n. 4, p. 697–701, 2005.
- MCMANUS, C. J. *et al.* Ribosome profiling reveals post-transcriptional buffering of divergent gene expression in yeast. **Genome Research**, v. 24, n. 3, p. 422–430, 2013.
- MEYUHAS, O. Synthesis of the translational apparatus is regulated at the translational level. **European Journal of Biochemistry**, v. 267, n. 21, p. 6321–6330, 2000.
- MIHAILOVICH, M. *et al.* Eukaryotic cold shock domain proteins: Highly versatile regulators of gene expression. **BioEssays**, v. 32, n. 2, p. 109–118, 2010.
- MIKULITS, W. *et al.* Isolation of translationally controlled mRNAs by differential screening. **The FASEB Journal**, v. 14, n. 11, p. 1641–1652, 2000.
- MIYATA, M. *et al.* IRES-mediated translational control of AMAP1 expression during differentiation of monocyte U937 cells. **Cell Cycle**, v. 7, n. 20, p. 3273–3281, 2008.
- MUMMERY, C. L. *et al.* Differentiation of human embryonic stem cells and induced pluripotent stem cells to cardiomyocytes: A methods overview. **Circulation Research**, v. 111, n. 3, p. 344–358, 2012.
- MURRY, C. E.; KELLER, G. Differentiation of Embryonic Stem Cells to Clinically Relevant Populations: Lessons from Embryonic Development. **Cell**, v. 132, n. 4, p. 661–680, 2008.
- NASTASI, T. *et al.* PIPPin is a brain-specific protein that contains a cold-shock domain and binds specifically to H1??and H3.3 mRNAs. **Journal of Biological Chemistry**, v. 274, n. 34, p. 24087–24093, 1999.
- NG, S. Y. S. Y.; WONG, C. K. C. K.; TSANG, S. Y. S. Y. Differential gene expressions in atrial and ventricular myocytes: insights into the road of applying embryonic stem cell-derived cardiomyocytes for future therapies. **American Journal of Physiology-Cell Physiology**, v. 299, n. 6, p. C1234–C1249, 2010.
- OBRIG, T. G. *et al.* The mechanism by which cycloheximide and related glutarimide antibiotics inhibit peptide synthesis on reticulocyte ribosomes. **Journal of Biological Chemistry**, v. 246, n. 1, p. 174–181, 1971.
- OCEGUERA-YANEZ, F. *et al.* Engineering the AAVS1 locus for consistent and scalable transgene expression in human iPSCs and their differentiated derivatives. **Methods**, v. 101, p. 43–55, 2016.
- OLSON, E. N. Gene Regulatory Networks in the Evolution and Development of the Heart. **Science**, v. 313, n. 5795, p. 1922–1927, 2006.
- OUNZAIN, S. *et al.* CARMEN, a human super enhancer-associated long noncoding RNA controlling cardiac specification, differentiation and homeostasis. **Journal of Molecular and Cellular Cardiology**, v. 89, p. 98–112, 2015.

- PAIGE, S. L. *et al.* Molecular regulation of cardiomyocyte differentiation. **Circulation Research**, v. 116, n. 2, p. 341–353, 2015.
- PANDA, A. C.; MARTINDALE, J. L.; GOROSPE, M. Polysome Fractionation to Analyze mRNA Distribution Profiles. **Bio Protocols**, v. 5, n. 2, 2017.
- PANG, K. C.; FRITH, M. C.; MATTICK, J. S. Rapid evolution of noncoding RNAs: Lack of conservation does not mean lack of function. **Trends in Genetics**, v. 22, n. 1, p. 1–5, 2006.
- PENNY, G. D. *et al.* Requirement for Xist in X chromosome inactivation. **Nature**, 1996.
- PIRCHER, A.; GEBETSBERGER, J.; POLACEK, N. Ribosome-associated ncRNAs: An emerging class of translation regulators. **RNA biology**, v. 11, n. 11, p. 1335–1339, 2014.
- PORRUA, O.; LIBRI, D. Transcription termination and the control of the transcriptome: Why, where and how to stop. **Nature Reviews Molecular Cell Biology**, v. 16, n. 3, p. 190–202, 2015.
- RAIMONDI, L. *et al.* RNA-binding ability of PIPPin requires the entire protein. **Journal of Cellular and Molecular Medicine**, v. 7, n. 1, p. 35–42, 2003.
- RAJ, A. *et al.* Stochastic mRNA synthesis in mammalian cells. **PLoS Biology**, v. 4, n. 10, p. 1707–1719, 2006.
- RAN, F. A. *et al.* Genome engineering using the CRISPR-Cas9 system. **Nature Protocols**, v. 8, n. 11, p. 2281–2308, 2013.
- RASER, J. M.; O'SHEA, E. K. Noise in gene expression: Origins, consequences, and control. **Science**, v. 309, n. 5743, p. 2010–2013, 2005.
- RIZKI, G.; BOYER, L. A. Lnc ing epigenetic control of transcription to cardiovascular development and disease. **Circulation Research**, v. 117, n. 2, p. 192–206, 2015.
- ROBERTSON, C.; TRAN, D. D.; GEORGE, S. C. Concise review: Maturation phases of human pluripotent stem cell-derived cardiomyocytes. **Stem Cells**, v. 31, n. 5, p. 829–837, 2013.
- ROBINSON, M. D.; MCCARTHY, D. J.; SMYTH, G. K. edgeR: A Bioconductor package for differential expression analysis of digital gene expression data. **Bioinformatics**, v. 26, n. 1, p. 139–140, 2010.
- RODGERS, J. T. *et al.* MTORC1 controls the adaptive transition of quiescent stem cells from G0 to G1. **Nature**, v. 510, n. 7505, p. 393–396, 2014.
- RUIZ-ORERA, J. *et al.* Long non-coding RNAs as a source of new peptides. **eLife**, v. 3, p. 1–24, 2014.
- SAINSBURY, S.; BERNECKY, C.; CRAMER, P. Structural basis of transcription initiation by RNA polymerase II. **Nature Reviews Molecular Cell Biology**, v. 16, n. 3, p. 129–143, 2015.
- SAMPATH, P. *et al.* A Hierarchical Network Controls Protein Translation during Murine Embryonic Stem Cell Self-Renewal and Differentiation. **Cell Stem Cell**, v. 2, n. 5, p. 448–460, 2008.
- SARTIANI, L. *et al.* Developmental Changes in Cardiomyocytes Differentiated from Human Embryonic Stem Cells: A Molecular and Electrophysiological Approach. **Stem Cells**, v. 25, n. 5, p. 1136–1144, 2007.

- SCHEUERMANN, J. C.; BOYER, L. A. Getting to the heart of the matter: long non-coding RNAs in cardiac development and disease. **The EMBO Journal**, v. 32, n. 13, p. 1805–1816, 2013.
- SCHLESINGER, J. *et al.* The cardiac transcription network modulated by *gata4*, *mef2a*, *nkx2.5*, *srf*, histone modifications, and microRNAs. **PLoS Genetics**, v. 7, n. 2, 2011.
- SCHMITZ, S. U.; GROTE, P.; HERRMANN, B. G. Mechanisms of long noncoding RNA function in development and disease. **Cellular and Molecular Life Sciences**, v. 73, n. 13, p. 2491–2509, 2016.
- SCHWANHÄUSSER, B. *et al.* Global quantification of mammalian gene expression control. **Nature**, v. 473, n. 7347, p. 337–342, 2011.
- SENYO, S. E.; LEE, R. T.; KÜHN, B. Cardiac regeneration based on mechanisms of cardiomyocyte proliferation and differentiation. **Stem Cell Research**, v. 13, n. 3, p. 532–541, 2014.
- SIEGEL, R. *et al.* Cancer Statistics , 2014. **CA Cancer Journal Clinical**, v. 64, n. 1, p. 9–29, 2014.
- SIMSEK, D. *et al.* The Mammalian Ribo-interactome Reveals Ribosome Functional Diversity and Heterogeneity. **Cell**, v. 169, n. 6, p. 1051–1065.e18, 2017.
- SINGH, V. K.; KUMAR, N. Mechanism of Induction : Induced Pluripotent Stem Cells (iPSCs). **Journal of Stem Cells**, v. 10, n. September 2016, 2015.
- SKELTON, R. J. P. *et al.* Biomarkers of Human Pluripotent Stem Cell-Derived Cardiac Lineages. **Trends in Molecular Medicine**, v. xx, 2017.
- SONENBERG, N.; HINNEBUSCH, A. G. New Modes of Translational Control in Development, Behavior, and Disease. **Molecular Cell**, v. 28, n. 5, p. 721–729, 2007.
- SPANGENBERG, L. *et al.* Polysome profiling shows extensive posttranscriptional regulation during human adipocyte stem cell differentiation into adipocytes. **Stem cell research**, v. 11, n. 2, p. 902–12, set. 2013.
- SPATER, D. *et al.* How to make a cardiomyocyte. **Development**, v. 141, n. 23, p. 4418–4431, 2014.
- SPITZ, F.; FURLONG, E. E. M. Transcription factors: From enhancer binding to developmental control. **Nature Reviews Genetics**, v. 13, n. 9, p. 613–626, 2012.
- STEFANOVIC, S.; ZAFFRAN, S. Mechanisms of retinoic acid signaling during cardiogenesis. **Mechanisms of Development**, v. 143, p. 9–19, 2017.
- STEFL, R.; SKRISOVSKA, L.; ALLAIN, F. H.-T. RNA sequence- and shape-dependent recognition by proteins in the ribonucleoprotein particle. **EMBO reports**, v. 6, n. 1, p. 33–38, 2005.
- SUNNERHAGEN, P. Cytoplasmatic post-transcriptional regulation and intracellular signalling. **Molecular Genetics and Genomics**, v. 277, n. 4, p. 341–355, 2007.
- SVOBODOVÁ, K. *et al.* Porcine *EEF1A1* and *EEF1A2* genes: genomic structure, polymorphism, mapping and expression. **Molecular Biology Reports**, v. 42, n. 8, p. 1257–1264, 2015.
- SYLVA, M.; VAN DEN HOFF, M. J. B.; MOORMAN, A. F. M. Development of the human heart. **American Journal of Medical Genetics, Part A**, v. 164, n. 6, p. 1347–1371, 2014.

- TAKAHASHI, K. *et al.* Induction of Pluripotent Stem Cells from Adult Human Fibroblasts by Defined Factors. **Cell**, v. 131, n. 5, p. 861–872, 2007.
- TALLAFUSS, A. *et al.* Characterization of retinoid-X receptor genes rxra, rxrba, rxrbb and rxrg during zebrafish development. **Gene Expression Patterns**, v. 6, n. 5, p. 556–565, 2006.
- TEBALDI, T. *et al.* Widespread uncoupling between transcriptome and translome variations after a stimulus in mammalian cells. **BMC genomics**, v. 13, n. 1, p. 220, 2012.
- THOMSON, J. A. ET. AL. Embryonic Stem Cell Lines Derived from Human Blastocysts. **Science**, v. 282, n. 5391, p. 1145–1147, 1998.
- TIAN, Q. *et al.* Integrated Genomic and Proteomic Analyses of Gene Expression in Mammalian Cells. **Molecular & Cellular Proteomics**, p. 960–969, 2004.
- TOHYAMA, S. *et al.* Distinct metabolic flow enables large-scale purification of mouse and human pluripotent stem cell-derived cardiomyocytes. **Cell Stem Cell**, v. 12, n. 1, p. 127–137, 2013.
- TOMPKINS, J. D. *et al.* Mapping Human Pluripotent-to-Cardiomyocyte Differentiation: Methylomes, Transcriptomes, and Exon DNA Methylation “Memories”. **EBioMedicine**, v. 4, p. 74–85, 2016.
- TROUNSON, A.; MCDONALD, C. Stem Cell Therapies in Clinical Trials: Progress and Challenges. **Cell Stem Cell**, v. 17, n. 1, p. 11–22, 2015.
- TUORTO, F. *et al.* RNA cytosine methylation by Dnmt2 and NSun2 promotes tRNA stability and protein synthesis. **Nature Structural and Molecular Biology**, v. 19, n. 9, p. 900–905, 2012.
- ULITSKY, I.; BARTEL, D. P. LincRNAs: Genomics, evolution, and mechanisms. **Cell**, v. 154, n. 1, p. 26–46, 2013.
- VAN HEESCH, S. *et al.* Extensive localization of long noncoding RNAs to the cytosol and mono- and polyribosomal complexes. **Genome biology**, v. 15, n. 1, p. R6, 2014.
- VIDARSSON, H.; HYLLNER, J.; SARTIPY, P. Differentiation of human embryonic stem cells to cardiomyocytes for *in vitro* and *in vivo* applications. **Stem Cell Reviews and Reports**, v. 6, n. 1, p. 108–120, 2010.
- VOLDERS, P. J. *et al.* An update on LNCipedia: A database for annotated human lncRNA sequences. **Nucleic Acids Research**, v. 43, n. D1, p. D174–D180, 2015.
- WAGERS, A. J.; WEISSMAN, I. L. Plasticity of adult stem cells. **Cell**, v. 116, n. 5, p. 639–648, 2004.
- WAMSTAD, J. A. *et al.* Dynamic and coordinated epigenetic regulation of developmental transitions in the cardiac lineage. **Cell**, v. 151, n. 1, p. 206–220, 2012.
- WANG, H.; HAO, J.; HONG, C. C. Cardiac induction of embryonic stem cells by a small molecule inhibitor of Wnt/beta-catenin signaling. **ACS Chem Biol**, v. 6, n. 2, p. 192–197, 2011.
- WEISSMAN, I. L. Stem cells: units of development, units of regeneration, and units in evolution. **Cell**, v. 100, n. 1, p. 157–168, 2000.
- WU, S. PIN *et al.* Atrial Identity Is Determined by a COUP-TFII Regulatory Network. **Developmental Cell**, v. 25, n. 4, p. 417–426, 2013.
- WU, X.; BREWER, G. The regulation of mRNA stability in mammalian cells: 2.0. **Gene**,

v. 500, n. 1, p. 10–21, 2012.

XAVIER-NETO, J. *et al.* A retinoic acid-inducible transgenic marker of sino-atrial development in the mouse heart. **Development (Cambridge, England)**, v. 126, n. 12, p. 2677–87, 1999.

XU, C. *et al.* Characterization and enrichment of cardiomyocytes derived from human embryonic stem cells. **Circulation Research**, v. 91, n. 6, p. 501–508, 2002.

XU, X. Q. *et al.* Global Expression Profile of Highly Enriched Cardiomyocytes Derived from Human Embryonic Stem Cells. **Stem Cells**, v. 27, n. 9, p. 2163–2174, 2009.

YAMAMOTO, A. *et al.* Shisa promotes head formation through the inhibition of receptor protein maturation for the caudalizing factors, Wnt and FGF. **Cell**, v. 120, n. 2, p. 223–235, 2005.

YANG, L. *et al.* Human cardiovascular progenitor cells develop from a KDR+ embryonic-stem-cell-derived population. **Nature**, v. 453, n. 7194, p. 524–528, 2008.

YOU, K. T.; PARK, J.; KIM, V. N. Role of the small subunit processome in the maintenance of pluripotent stem cells. **Genes & development**, p. 2004–2009, 2012.

YOUNG, R. A. Control of the embryonic stem cell state. **Cell**, v. 144, n. 6, p. 940–954, 2011.

ZAMILPA, R. *et al.* Stem cell mechanisms during left ventricular remodeling post-myocardial infarction: Repair and regeneration. **World journal of cardiology**, v. 6, n. 7, p. 610–20, 2014.

ZENG, Y. *et al.* DUSP1 and KCNJ2 mRNA upregulation can serve as a biomarker of mechanical asphyxia-induced death in cardiac tissue. **International Journal of Legal Medicine**, v. 132, n. 3, p. 655–665, 2018.

ZHANG, Q. *et al.* Direct differentiation of atrial and ventricular myocytes from human embryonic stem cells by alternating retinoid signals. **Cell Research**, v. 21, n. 4, p. 579–587, 2011.

ZHU, N. *et al.* NR2F2 regulates bone marrow-derived mesenchymal stem cell-promoted proliferation of Reh cells. **Molecular Medicine Reports**, v. 14, n. 2, p. 1351–1356, 2016.

ZYCH, J. *et al.* Polysome profiling shows the identity of human adipose-derived stromal/stem cells in detail and clearly distinguishes them from dermal fibroblasts. **Stem cells and development**, v. 23, n. 22, p. 2791–802, 2014.

ANEXO I

Tabela 1. Oligos usados na amplificação das sequências para clonagem nos vetores AAVS1. As sequências de homologia/**sítio de restrição** do AAVS1-TRE-FLAG-rtTA estão marcadas em laranja. As sequências attB para clonagem no sistema Gateway com AAVS1-GW-TRE-rtTA estão marcadas em verde.

Gene	Ref GenBank	Nome do oligo	Sequência 5'-3'	Tamanho do amplicon
CSDC2	BC067113.1	CSDC2 flag F	CGATGACAAGATGCAT ACTTCAGAGTCGACGTCA	493 pb
		CSDC2 flag R	TCTGCTCGA ATTAATTA ACTAGGAGCCACGACCTG	
RXRG	BC012063	RXRG flag F	CGATGACAAGATGCAT TATGGAAATTATTCTCAC	1,4 kb
		RXRG flag R	TCTGCTCGA ATTAATTA ATCAGGTGATCTGCAGCGG	
SHISA3	BC127690	SHISA3 flag F	CGATGACAAGATGCAT AGGGCACTGCTGGCGCTT	748 pb
		SHISA3 flag R	TCTGCTCGA ATTAATTA ATCAACTGGAAGTGAAGTC	
LINC00890	BX537697.1	LINC890_attb1F	GGGGACAAGTTTGTACAAAAAAGCAGGCTCCGCATA CTCTAAACTGTTGCC	4,5 kb
		LINC890_attb2R	GGGGACCAC TTTGTACA AGAAAGCTGGGTCATCATC AAACCAGTCATTAT	

Tabela 2. Oligos usados na amplificação das sequências de CSDC2 com deleção e *LINC00890* com inserção para clonagem nos vetores AAVS1. Em laranja, estão destacadas as sequências complementares da estratégia de deleção de CSDC2. Em vermelho, estão destacados os dois nucleotídeos inseridos na sequência de *LINC00890*. Em negrito, está destacado o códon inicial ATG predito para *LINC00890*. WT, *wild type* (do inglês, selvagem).

Gene	Nome do oligo	Sequência 5'-3'	Mutação
CSDC2	CSDC2del_F	AAGCAGTTCTCACGCTCACAG GTGTCTGACATCGAGGGGGAG	deleção
	CSDC2del_R	CTGTGAGCGTGAGA ACTGCTT	
LINC00890	LINCmutAAVS1_F	AATACC ATG GAAATCA AA TCACTAGGCGTGCC	inserção
	LINCmutAAVS1_R	GGCACGCCTAGTGA TT GATTTCCATGGTATT	
	PCR_3LF	GCTGGGCAGGAACCTTACAA	WT
	PCR_3LR	TGTTTGAGCCGAAGGAGCAT	

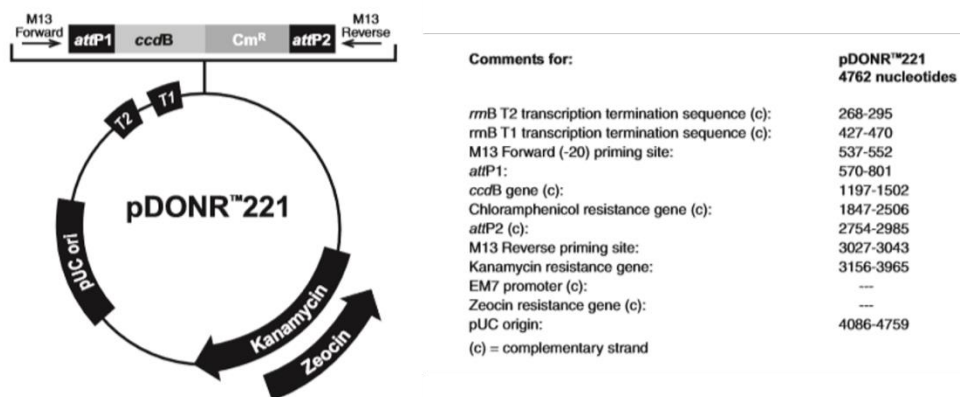


Figura 1. Vetor pDONOR 221 usado para clonagem no sistema Gateway. Retirado de: <https://www.thermofisher.com/br/en/home/life-science/cloning/gateway-cloning/protocols.html>

ANEXO II

Tabela 1. Oligos para construção dos gRNAs de reconhecimento do gene *LINC00890*. Sublinhados estão os nucleotídeos adicionais para a montagem no vetor pSpCas9(BB)-2A-RFP (PX458).

gRNA	Região		Sequência 5'-3'
A	<i>upstream</i>	F	<u>CACCG</u> CATATATTACTGTAAGACCC
		R	AAACGGGTCTTACAGTAATATATGC <u>CG</u>
B	<i>upstream</i>	F	<u>CACCG</u> TAAACTGTTGCCTGCGAAAG
		R	AAACCTTTTCGCAGGCAACAGTTTAC <u>CG</u>
C	<i>downstream</i>	F	<u>CACCG</u> AAGTCAGCATTTCGAATAGA
		R	AAACTCTATTTCGAAATGCTGACTTC <u>CG</u>
D	<i>downstream</i>	F	<u>CACCG</u> GCAGTCGTTACAATTTATAC
		R	AAACGTATAAATTGTAACGACTGCC <u>CG</u>

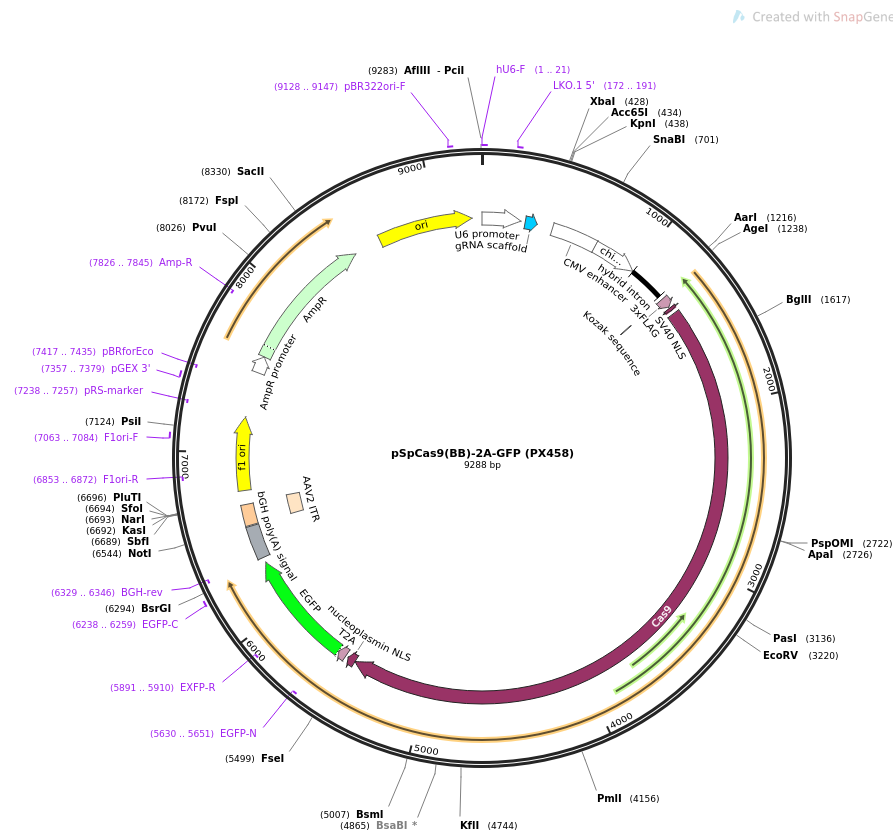


Figura 1. Mapa do vetor pSpCas9(BB)-2A-GFP (PX458). O GFP foi substituído por RFP no laboratório do Dr Michal Kyba e foi usado neste trabalho.

ANEXO III

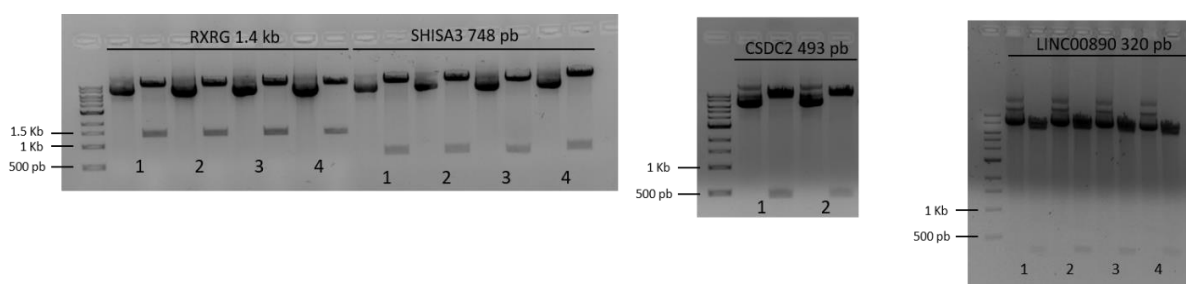


Figura 1. Reações de restrição para confirmação da inserção dos transgenes. O vetor AAVS1-TRE-FLAG-rtTA clonado com RXRG, SHISA3 e CSDC2 foi incubado com NsiI e PaeI, gerando um fragmento com o tamanho correspondente ao amplificado pelos oligo da Tabela 1 (ANEXO I). O vetor AAVS1-TRE-GW-rtTA foi incubado com SspI, reação que gera um fragmento de 320 bp (interno a sequência de *LINC00890*). Imagens de gel de agarose ilustrando os fragmentos gerados pelas reações de restrição. Clones avaliados numerados nas imagens, onde cada par de coluna no gel representa o DNA plasmidial de um clone – sem e com reação restrição, respectivamente. Os clones também foram confirmados por sequenciamento.

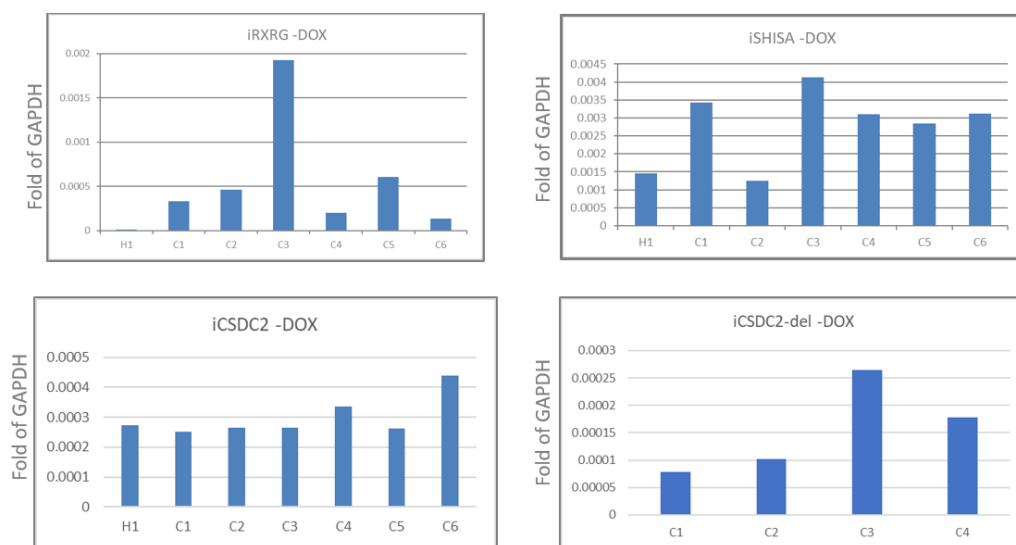


Figura 2. Expressão basal dos transgenes, sem indução de doxíciclina. Valores relativos ao gene constitutivo GAPDH estão mostrados nos gráficos, e foram quantificados por qPCR. Todos os qPCR foram feitos com probes Taqman, exceto iCSDC2-del que foi quantificado usando os oligos listados na Tabela 4.2. H1, expressão nas células parentais sem transfecção.

ANEXO IV

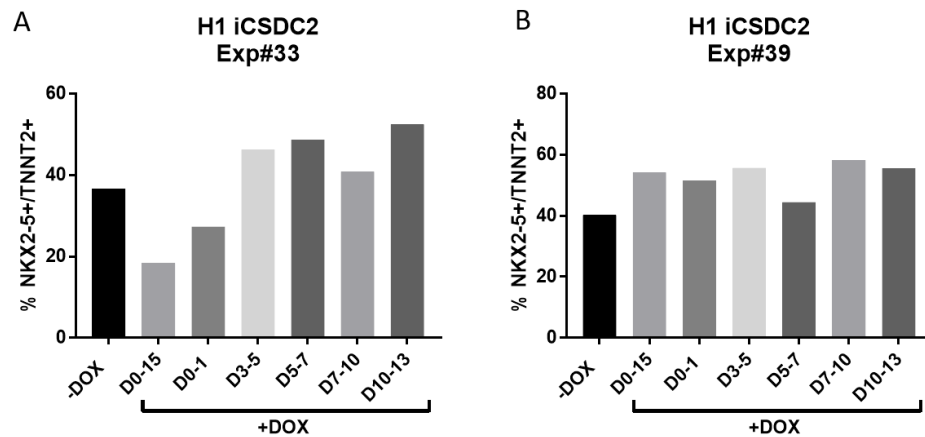


Figura 1. Ensaio funcional de diferenciação cardiomiogênica com a linhagem H1 iCSDC2. Diferentes janelas de indução com doxiciclina (+DOX) foram testadas (por exemplo, D0-15 a indução foi feita do dia 0 ao dia 15 do protocolo de diferenciação). Células sem indução (-DOX) foram usadas como controle. A porcentagem de células duplo positivas NKX2-5+/TNNT2+ está representada nos gráficos em cada um dos experimentos independentes (A – Experimento #33 e B – Experimento #39) e foi medida através da imuno-marcação e leitura por citometria de fluxo.

ANEXO V

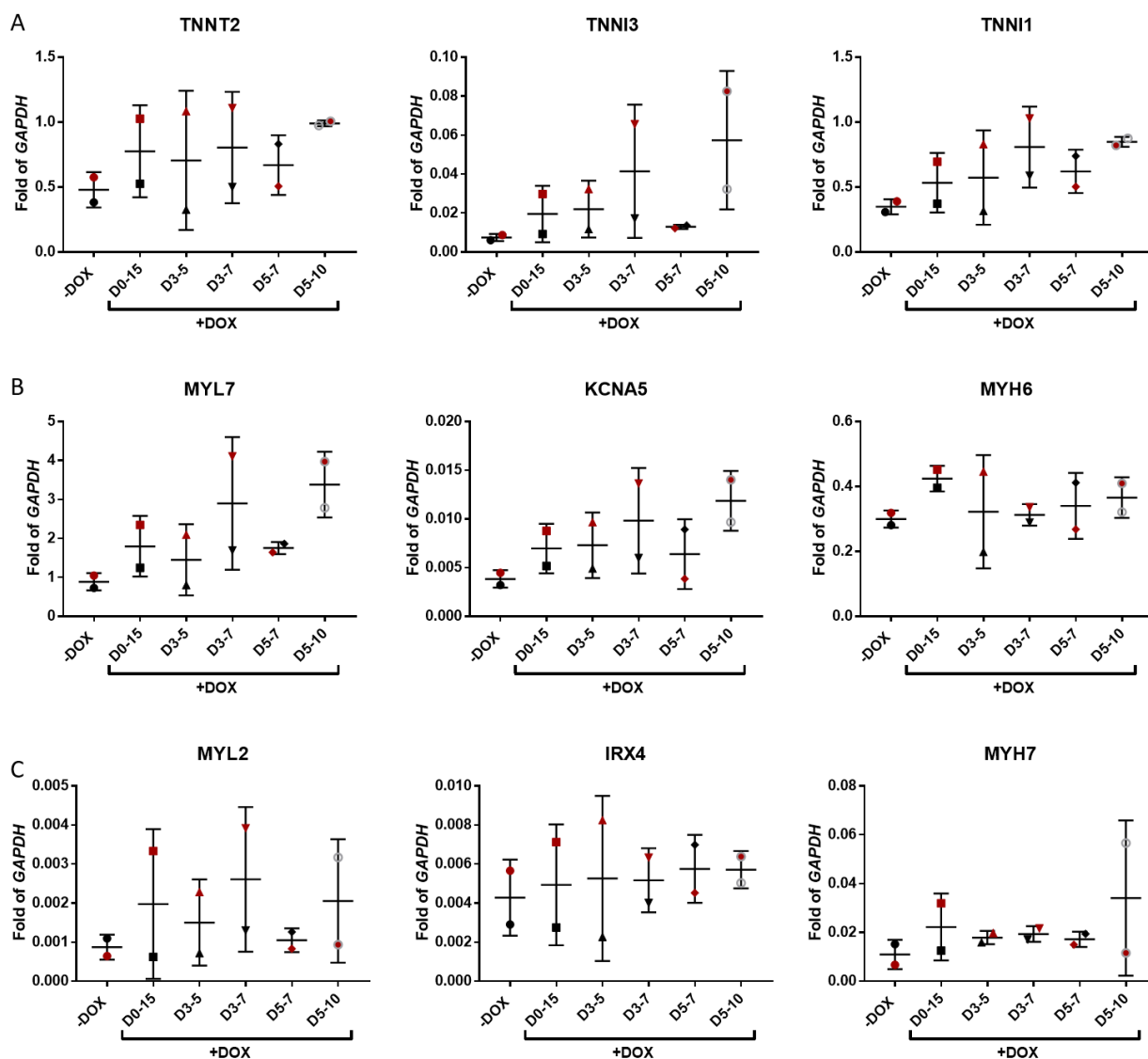


Figura 1. Expressão de marcadores (A) pan-cardíacos, (B) atriais e (C) ventriculares avaliados por qPCR, após a indução da linhagem H1 iLINC com doxiciclina nos tempos indicados. Células sem indução (-DOX) foram usadas como controle. Dois experimentos independentes foram realizados (n=2). Os valores estão expressos como relativos ao gene constitutivo GAPDH.