UNIVERSIDADE DE LISBOA
FACULDADE DE CIÊNCIAS
DEPARTAMENTO DE BIOLOGIA VEGETAL

Population-Level Dynamics of Human Pluripotent Stem Cell Fate Transitions

Joana Estudante Saraiva

Mestrado em Biologia Molecular e Genética

Dissertação orientada por:
Professor Tiago Fernandes
Professora Gabriela Rodrigues

2019

Population-Level Dynamics of Human Pluripotent Stem Cell Fate Transitions

JOANA ESTUDANTE SARAIVA

Acknowledgements

I would like to thank everyone that contributed to this thesis and to all who have supported me during this wonderful time. Particularly, I would like to thank:

Professor Joaquim Cabral for receiving me in SCERG, for the opportunity and attention that has been provided to me, and for his contribution to science.

Professor Tiago Fernandes, my supervisor and to whom I owe my participation in this most interesting project. Thank you for believing in me and sharing your work with me. Thank you for all the support that you gave me in my decisions and in all kinds of subjects related to my thesis and more. Thank you for always listening to me, being prompt to help me, as well as for your constant guidance throughout this project, and for sharing your vast knowledge and experience with me.

Professor Gabriela Rodrigues, also my supervisor, for having inspired me about stem cell research, right before starting my thesis. Thank you also for your kind support in all the matters related with my thesis and my academic goals, always motivating me to go further.

Mariana Branco, my mentor and guardian angel, for everything you have taught me and challenged me in order to improve my skills and work. I have no words to describe all the incessant patience, attention, time, knowledge, kind support, motivation, guidance and inspiration that you have given me, from the first moment I started my thesis until the very last. You have always been so bright in discussing ideas with me and contributing with your critical thinking and knowledge in so many aspects. Also, you have helped me in my most challenging times, solving my doubts and problems with such simplicity, providing me with new opportunities that have been so important to me and my work. Also, thank you for teaching me so much regarding my *in silico* analysis and in the lab, where I had the opportunity to train and learn through your wonderful guidance. I am so grateful to you, that I only hope that at least I have given a little contribution to your work with some of my results, and I look forward to collaborating with you once again in a near future. You have been a great colleague and friend during all this time.

João Cotovio, my second mentor, for teaching so much in the lab, always ready to explain every little detail about stem cell culture and differentiation. Thank you for providing me the opportunity to discover those concepts for the first time and for sharing your entire work with me, which has been very kind and enriching for me. Thank you for answering to all the scientific questions that crossed my mind and for your constant good mood and pleasant company, which made me feel so welcome at SCERG since the beginning.

Ana Rita and Teresa, for answering my questions about neuronal differentiation and guiding me in my thesis. You have been really kind and friendly to me.

Diogo Nogueira, Jaqueline Garcia and Carlos Rodrigues, for our conversations about future options and career perspectives. Thank you for sharing your realistic points of view, experiences and jokes.

I would also like to thank all the other SCERG members, who have been so supportive and cheerful during this time, showing me the best environment that anyone would love to work in.

My friends, from the Arts field (Rui Peixoto, Ana Ferreira Mateus, Fábio Constantino and João Cidade), from Science (Ana Beatriz Varanda, Ricardo Quiteres, Sofia Marques, Leonor Bismarck and many more) for inspiring me and for sharing your good friendship and so many laughter!

My family, mother, father, sister, grandfathers, sister in law… for believing and supporting me at all levels. And a special thanks to my father, for kindly helping me in my decisions, and discussing new ideas with me, in a funny but meaningful way. Your help has been crucial for my personal and academic journey and I would have not been here without you.

My boyfriend, Gonçalo Costa, because I do not know how you can successfully handle me every day! You have believed in me and celebrated my highest moments and also supported me in my lowest ones, giving me whatever I needed to make me feel happy, from a simple foot massage to a sightseeing… Thank you for your huge love, for all you share with me and for being the best partner that anyone could ever had.

So, THANK YOU ALL so much! You have all been really important to me!

Abstract

Stem cells are undifferentiated cells with the ability to self-renew and differentiate into specialized cells, under appropriate conditions. During gastrulation cells rearrange themselves transforming the early embryo into a multi-layered structure containing three germ layers, endoderm, mesoderm and ectoderm. Accordingly, pluripotent stem cells (PSC), *e.g.* embryonic stem cells (ESC) or induced pluripotent stem cells (iPSC), differentiate into cells of all three primitive germ layers. Differentiating stem cells negotiate branching lineage choices, avoiding alternate fates to decisively commit to a single lineage. However, the regulatory mechanisms underlying PSC differentiation into specific lineages remain elusive.

Here, we define a representative roadmap for early human development and reconstruct the differentiation trajectories by which pluripotent cells sequentially elaborate diverse neuroectodermal, cardiac mesodermal and hepatic endodermal progeny. Tracking these three-germ layer-derived lineage-choices *in vitro* uncovers the sequential molecular events that ultimately lead to the establishment of different cell types. Therefore, we provide a comprehensive study through transcriptomic and network modeling approaches, based on RNA-Sequencing data, to investigate the critical changes that may occur during human early cardiogenesis, neurogenesis and hepatogenesis, mimicked by human PSC-derived cardiomyocytes, neurons and hepatocytes. In addition, published datasets were also analysed to cover multiple cellular lineages and construct the developmental trajectories of cells derived from all three germ layers, as well as to identify key lineage specifiers. This paved the way for a deeper understanding of the regulatory mechanisms driving distinct cellular subpopulations present at each differentiation stage. To that aim, we have developed two main approaches, one comprising global gene expression of lineage specifiers and another focusing solely on genes encoding transcription factors, as key regulators of cellular decisions. Interestingly, this transcriptomic analysis allowed us to identify novel key putative regulators of cardiogenesis (*HOXB4, HOXB5*, *HOXB6* and *RARB)* and neurogenesis (*TERF2IP, RFX4* and *ZHX1)*, whose expression results throughout differentiation have been validated by Quantitative Real Time Polymerase Chain Reaction.

Collectively, these data provide a starting point to better understand the mechanisms that control human PSC fate transitions, revealing the cellular landscape of human PSC early differentiation. We expect that this knowledge might one day be exploited for developmental cell biology, regenerative medicine, disease modeling and drug discovery applications.

KEYWORDS

Stem Cell Fate, Transcriptomic Analysis, Human Induced Pluripotent Stem Cells, Cardiac Differentiation, Neuronal Differentiation

Resumo

As células estaminais são células indiferenciadas com a capacidade de auto-renovação e diferenciação em células especializadas, em condições apropriadas. Desempenham funções fundamentais em diversas fases do desenvolvimento, podendo ser totipotentes, pluripotentes, multipotentes e unipotentes.

As células estaminais pluripotentes humanas (*hPSC*), incluindo células estaminais embrionárias humanas (*hESC*) e células estaminais pluripotentes induzidas humanas (*hiPSC*), têm como principais aplicações a descoberta de fármacos, a modelação de doenças e a medicina regenerativa.

O desenvolvimento embrionário humano inicia-se com a fusão dos gâmetas masculino e feminino, gerando-se reprogramações epigenéticas, uma série de divisões mitóticas (clivagem), degradação dos transcriptos maternos e ativação do genoma embrionário. De seguida, o embrião sofre compactação e cavitação, passando pelo estágio de mórula, até formar a blástula. A massa celular interna (*ICM*) do blastocisto diverge entre as células do epiblasto e da endoderme primitiva (hipoblasto), seguindo-se a implantação do blastocisto e gastrulação.

Durante a gastrulação as células rearranjam-se, transformando o embrião numa estrutura contendo três camadas germinativas - ectoderme, mesoderme e endoderme. De forma semelhante, muitos métodos de diferenciação *in vitro* visam o comprometimento de *hPSC* numa linhagem multipotente, que dará origem ao tipo celular de interesse. Deste modo, a diferenciação de *hPSC* numa linhagem celular pretendida requer a regulação complexa de diversas vias de sinalização, através de fatores de crescimento, citocinas e pequenas moléculas.

Quando as *hPSC* são induzidas com Wnt, Activina A ou BMP4, estas células formam uma população semelhante à linha primitiva e subsequentemente, a mesoderme ou endoderme. As células da linha primitiva posterior dão origem à mesoderme, seguindo-se a formação da mesoderme cardíaca e dos progenitores cardíacos, os quais geram células diferenciadas. Alternativamente, as células da linha primitiva anterior podem ser induzidas por Activina/Nodal e deste modo formar endoderme definitiva $FOXA2^{+}$, seguindo-se a formação de hepatoblastos, que irão dar origem a hepatócitos. Caso estas vias não sejam ativas, as *hPSC* irão diferenciar-se em ectoderme. A diferenciação de *hiPSC* em neurónios transita através de diversos estadios intermediários. As *hiPSC* diferenciam-se em progenitores neurais/células estaminais neuroepiteliais que se assemelham às células estaminais neurais (*NSC*), que formam o tubo neural *in vivo*. Durante a neurulação *in vivo*, o tubo neural fecha e formam-se os primeiros tipos de neurónios. O passo correspondente *in vitro* é a formação de rosetas neurais, que irão dar origem a células da glia e neurónios.

A investigação dos mecanismos que regulam as decisões celulares e especificação de sub-populações é fundamental para uma melhor compreensão do desenvolvimento humano embrionário *in vivo* e sua modelação *in vitro*. Este conhecimento poderá vir a ser explorado para a medicina regenerativa, modelação de doenças e descoberta de fármacos. No entanto, abordagens sistemáticas para a cultura e diferenciação de células estaminais permanecem pouco desenvolvidas. Esta necessidade tecnológica surgiu como o grande propósito deste estudo, sendo o seu objetivo principal a análise de processos robustos e reprodutíveis para a diferenciação de *hPSC*.

Com este fim, começámos por expandir uma base de dados génica, de modo a abranger o maior número possível de genes associados à especificação de linhagens celulares derivadas de cada uma das três camadas germinativas. De seguida, analisámos as proteínas codificadas pelos genes dessa base de dados aumentada, de modo a identificar quais os processos biológicos e vias de sinalização enriquecidos associados.

Ainda, foi possível delinear um percurso representativo do desenvolvimento humano, reconstruindo-se as trajetórias de diferenciação pelas quais as h*PSC* elaboram sequencialmente uma diversidade de linhas celulares derivadas da neuroectoderme, mesoderme cardíaca e endoderme

hepática. Esta abordagem revelou os eventos moleculares sequenciais que regulam o estabelecimento dos três diferentes tipos celulares analisados.

Para este estudo desenvolveram-se abordagens de transcriptómica e de modelação de redes, com base nos dados de sequenciação de RNA (*RNA-Seq*) de células diferenciadas. Esta análise permitiu revelar as vias de sinalização e perfis de expressão génica responsáveis pela indução eficiente de cada uma destas linhagens em diversos pontos temporais. Esta estratégia foi desenvolvida ao nível da expressão global de genes e, posteriormente, com enfoque exclusivo em genes que codificam fatores de transcrição, uma vez que estes controlam uma rede regulatória de programas de expressão génica, despoletando diversas respostas celulares. Deste modo, desenvolveu-se uma análise que incluiu a normalização de dados, visualização de dados de transcriptómica por *heatmaps*, *PCAs (Principal Component Analysis)* e diagramas de *Venn*, e investigação dos processos biológicos e principais vias de sinalização associados à expressão génica.

Demonstrou-se que os cardiomiócitos derivados de *hiPSC (hiPSC-CM)* e os neurónios derivados de *hiPSC (hiPSC-Neurons)* exibiram uma diferenciação progressiva robusta, quando comparados a tecidos fetais e adultos cardíacos e cerebrais, respetivamente. Deste modo, em termos de expressão génica, os nossos métodos de diferenciação cardíaca e neural revelaram-se bem sucedidos.

Realizou-se ainda uma comparação entre os protocolos que estabelecemos para a diferenciação cardíaca e neural com protocolos distintos para a obtenção dos mesmos tipos celulares, embora por intermédio de diferentes linhas celulares, fatores de diferenciação e/ou meios de cultura. Apesar destas diferenças, uma análise de *PCA* mostrou que os diversos grupos surgiram por ordem cronológica, claramente distinguindo as amostras de diferentes estadios de diferenciação, permitindo definir as trajetórias de linhagens celulares, validando a nossa eficiência de diferenciação.

Adicionalmente, foi analisada a expressão de genes que apenas codificam fatores de transcrição, regulando a diferenciação de *hiPSC-CM*, *hiPSC-Neurons* e de hepatócitos derivados de *hESC*. A análise de *PCA* evidenciou a existência de diferenças consideráveis entre estas três diferenciações, dada a existência de redes génicas distintas a regular linhagens celulares específicas.

De seguida, procurou-se identificar os padrões de expressão génica que modulam a especificação progressiva da ectoderme e da mesoderme, bem como os seus principais papéis funcionais. Assim, compararam-se as alterações de expressão génica em fases sequenciais da diferenciação neural e cardíaca, juntamente como uma análise dos genes diferencialmente expressos (*DEGs*) (*Differentially Expressed Genes*) no *Gene Ontology (GO).* Esta abordagem foi desenvolvida tanto a um nível global de expressão de genes, como posteriormente direcionada apenas para genes que codificam fatores de transcrição. Ambas demonstraram diferenças consideráveis entre os *DEGs* durante a diferenciação em ectoderme comparativamente à diferenciação em mesoderme, em termos de expressão de genes e suas funções associadas, bem como na modulação de vias de sinalização, os quais foram devidamente identificados. Estes resultados permitiram comparar e discriminar as principais alterações de expressão génica e respetivos processos biológicos enriquecidos, durante a transição de um estado de pluripotência para um comprometimento em ectoderme ou mesoderme.

Adicionalmente, foi feito um *screening* de genes que apenas codificam fatores de transcrição em fases comuns da diferenciação em três *datasets* (*hiPSC-Neurons*, *hiPSC-CM* e hepatócitos derivados de *hESC*). Deste modo, foi possível identificar que fatores de transcrição estão particularmente envolvidos em fases sequenciais da especificação de cada camada germinativa.

Para concluir, realizou-se um estudo do perfil de alterações transcripcionais em fases sequenciais, através da comparação de dois *datasets* de *RNA-Seq* para cada tipo de diferenciação. Esta estratégia foi utilizada para analisar com maior precisão alterações determinadas e suportadas por ambos os *datasets* para cada tipo de diferenciação, validando as nossas observações. Apesar das diferenças entres os protocolos de diferenciação e linhas celulares utilizadas, os dois tipos de neurónios diferenciados continuaram a exibir transcriptomas altamente concordantes durante a diferenciação neural,

evidenciados por picos específicos de expressão génica que ocorreram aproximadamente nos mesmos pontos temporais da diferenciação, apresentando tendências similares de expressão. O mesmo se observou entre os dois *datasets* cardíacos que foram analisados.

Esta investigação permitiu ainda identificar novos reguladores putativos da cardiogénese (*HOXB4, HOXB5, HOXB6* e *RARB)* e neurogénese (*TERF2IP, RFX4* e *ZHX1),* os quais foram subsequentemente validados por *Quantitative Real-Time Polymerase Chain Reaction (qRT-PCR).* Globalmente, os resultados obtidos por *qRT-PCR* foram concordantes com os dados *de RNA-Seq.* Adicionalmente, a eficiência da diferenciação cardíaca foi confirmada por citometria de fluxo, validando assim os resultados obtidos por *qRT-PCR* ao longo da diferenciação.

A caracterização transcriptómica efetuada definiu um mapeamento representativo do desenvolvimento da neuroectoderme e mesoderme cardíaca, revelando picos de expressão temporal quer de marcadores específicos de cada linhagem, quer de novos candidatos envolvidos na diferenciação. Estes resultados evidenciaram uma diferenciação eficiente de *hPSC* nos tipos celulares pretendidos.

Em suma, este estudo pretendeu contribuir para a análise de modelos de cultura celular que possam mimetizar a função *in vivo,* apresentando dados que poderão ser utilizados futuramente na otimização de protocolos de diferenciação. Deste modo, pretendemos fornecer informação importante relativa aos mecanismos que controlam as transições das *hPSC,* com implicações no conhecimento acerca da diferenciação de células estaminais em diferentes linhagens celulares, na biologia celular e do desenvolvimento, e na medicina regenerativa. Esperamos que esta área de investigação se mantenha altamente dinâmica num futuro próximo.

PALAVRAS-CHAVE

Análise Transcriptómica, Células Estaminais Pluripotentes Induzidas Humanas, Diferenciação Cardíaca, Diferenciação Neural, Comprometimento Celular

Table of Contents

List of Tables

List of Figures

List of Abbreviations

2D - two-dimensional
3D - three-dimensional
A-P - anterior-posterior
AV – atrioventricular
AVE - anterior visceral endoderm
bFGF- basic fibroblast growth factor
BMP - bone morphogenetic protein
BSA - bovine serum albumin
cDNA - complementary DNA
CDX2 – caudal type homeobox 2
CK19 – cytokeratin 19
CM – cardiomyocytes
CNS – central nervous system
CTNT - cardiac troponin T
DE - definitive endoderm
DEGs - differentially expressed genes
D-V - dorsal-ventral
EB - embryoid body
EDTA – ethylenediamine tetraacetic acid
EGA - embryonic genome activation
EMT - epithelial-to-mesenchymal transition
EOMES - eomesodermin
EPDCs - epicardium-derived cells
EPI - epiblast
ESC - embryonic stem cells
FB1 - flow cytometry buffer 1
FB2 - flow cytometry buffer 2
FDR - false discovery rate
FGF - fibroblast growth factor
FGF-ERK – fibroblast growth factor / extracellular signal-regulated kinase
FHF - first heart field
FLK-1 – fetal liver kinase-1 (VEGFR-2)
FOXA2 - forkhead box A2
GAPDH - glyceraldehyde-3-phosohate
GATA 4 - GATA binding protein 4
GATA 6 – GATA binding protein 6
GFAP - glial fibrillary acidic protein
GO - gene ontology
GSK3β - glycogen synthase kinase 3 beta
HAND2 - heart and neural crest derivatives expressed 2
HSC - hematopoietic stem cells
hESC - human embryonic stem cell
hESC-CM - human ESC-derived cardiomyocyte
hiPSC - human induced pluripotent stem cell
hiPSC-CM - human iPSC-derived cardiomyocyte
hiPSC-Neurons - human iPSC-derived neurons
HNF4α – hepatocyte nuclear factor 4 alpha
HOXB - homeobox B cluster
hPSC – human pluripotent stem cells
ICM - inner cell mass
iPSC - induced pluripotent stem cell
Isl1 - Islet1
IWP-4 - Wnt production 4
LBB - left bundle branch
LHX6 - LIM homeobox 6
MAP-2 - microtubule-associated protein 2
MAPK – mitogen-activated protein kinase
MCL - markov cluster algorithm
MEF2C - myocyte enhancer factor 2C
mESC - mouse embryonic stem cell
MSC - mesenchymal stromal cells
MESP1- mesoderm posterior 1
MGE - medial ganglionic eminence
miRNA - microRNA
MIXL1 - mix paired-like homeobox
mRNA - messenger RNA
MSX1- msh homeobox 1
MYC - proto-oncogene, bHLH transcription factor
NSC - neuronal stem cells
NKX2.1- NK2 homeobox 1
NKX2-5 - NK2 homeobox 5
NMPs - neuromesodermal progenitors
NPC – neural progenitor cell
OCT4 - octamer-binding transcription factor 4
ORF – open reading frame
OTX2 - orthodenticle homeobox 2
PAX6 - paired box 6
PBS - phosphate-buffered saline
PC1 – principal component 1
PC2 – principal component 2
PCA - principal component analysis
PE - primitive endoderm
PF - purkinje fibers

PFA – paraformaldehyde
PI3K - phosphoinositide 3-kinase
PSC - pluripotent stem cells
qRT-PCR - quantitative real-time polymerase chain reaction
RA - retinoic acid
RARB - retinoic Acid Receptor Beta
RBB - right bundle branch
RFX4 - regulatory factor X4
SAN - sinoatrial node
SHF - second heart field
SHH - sonic hedgehog
siRNA - small interfering RNA
SOX17 – SRY-box 17
SOX2 - SRY-box 2
T - brachyury
TAP - transient amplifying progenitor
TBX20 - T-box 20
TBX5 - T-box 5
TBXT - T-box transcription factor T
TE - extraembryonic trophectoderm
TERF2IP - telomeric repeat-binding factor 2-interacting protein 1
TFs – transcription factors
TGFβ - transforming growth factor-beta
THF - third heart field
TUJ1- neuron-specific class III beta-tubulin
Wnt - wingless integrated
ZHX1- zinc fingers and homeoboxes 1
β-Tubulin III – class III beta-tubulin

1 Introduction

1.1 Stem Cells

Stem cells are undifferentiated cells with the ability to self-renew (generate undifferentiated daughter cells to preserve stem cell populations) and to differentiate into specialized cells, under appropriate conditions. They play essential roles at diverse stages of development and present distinct potential fates - Totipotent, Pluripotent, Multipotent, and Unipotent.[1-3].

Therefore, the zygote and early blastomeres form all the embryonic tissues and the extraembryonic ones (yolk sac, amnion and placenta), a capacity termed totipotency. However, the zygote is hardly considered a stem cell in mammals, since it cleaves into blastomeres of equal developmental potency for at most three cell divisions, and subsequently displays very limited self-renewal potential[2].

Pluripotent stem cells (PSC), *e.g.* embryonic stem cells (ESC) or induced pluripotent stem cells (iPSC) differentiate into cells of all three primitive embryonic germ layers – ectoderm, mesoderm and endoderm[1], although they exhibit limited contribution to extraembryonic tissues[4]. These cells are isolated from the inner cell mass (ICM) of blastocysts and cultured as immortal cell lines (ESC),[5] or they can be induced from somatic cells by direct reprogramming with defined transcription factors (iPSC)[6]. PSC are also able to contribute to chimera formation/blastocyst complementation and to form teratomas, while iPSC reactivate telomerase gene expression[7,1]. The terms 'naive' and 'primed' describe the early and late phases of epiblast (EPI) ontogeny, pre and post-implantation, respectively, and its respective derivatives. Thereby, human pluripotent stem cells (hPSC) are classified as 'primed', being enriched in repressive chromatin features [5,8].

Multipotent stem cells differentiate into multiple cell types restricted to one lineage, *e.g.* hematopoietic stem cells (HSC), which form blood cells and immune cells; mesenchymal stromal cells (MSC), which form osteoblasts, chondrocytes and fat cells; neuronal stem cells (NSC), which form neurons, astrocytes and oligodendrocytes; and intestinal stem cells[1,2,9].

Unipotent stem cells, such as satellite, epidermal and spermatogonial stem cells, differentiate into only one cell type, *e.g.* to skeletal muscle, keratinocytes, or sperm, respectively[8]. Moreover, adult stem cells are either multipotent or unipotent and mainly reside in specialized niches[1].

1.1.2 Applications of Stem Cells

Recently, there has been an increasing interest in developing cell therapies to potentially replace diseased or damaged tissues[10]. To this aim, the choice of stem cell type and application is conditioned by the accessibility of the cell type, its potential benefits and risks, and the likelihood of its translation to clinical trials[11].

hPSC, including human ESC (hESC) and human induced pluripotent stem cells (hiPSC), have three main applications – disease modeling, drug screening and regenerative medicine[12].

hiPSC derived from patients and healthy individuals combined with efficient gene modification technologies have led to novel opportunities to model human diseases, in order to understand their mechanisms, and finding ways to delay or reverse them *in vivo*. However, many diseases have multicellular contributions. Therefore, the next generation of disease models is increasingly based on combinations of cell types, sometimes in ''organ-on-chip'' formats, microfluidic devices that integrate several cell types of various developmental lineages in chips[13]. Alternatively, we could use ''organoids,'' three-dimensional (3D) *in vitro* culture systems that recapitulate the *in vivo* architecture, physiology and genetic information of original tissues, resembling an organ[14]. Models based on hiPSC have many advantages in comparison with hESC, since they can provide rapid readouts of disease pathology and be employed as drug screening platforms for the development and validation of therapeutic compounds, either as target-based screening or phenotypic screening[12,15]. Moreover, hiPSCs represent attractive cell sources since they are isolated from accessible cell types, easily expanded *ex*

vivo, and able to differentiate into nearly any cell type, providing high quantities of cells. The major limitations that still need further optimization include the lack of mature structural and functional phenotypes in hPSC derivatives, low reproducibility, and the degree of complexity that can be achieved with *in vitro* stem cell models[12,13].

In addition, hPSC can potentially be used for regenerative medicine, to promote endogenous regenerative processes or replace damaged tissues after cellular transplantation[12]. To this aim, "3D patch-based platforms" have been extensively studied for regeneration. They are composed of biomaterials either with or without cells and once transplanted, they provide biological therapeutics and growth factors[16]. The major concerns before these type of approaches reach clinical trials comprise the reprogramming efficiency to generate hiPSC, which currently has low yield, and the risk of malignant transformation due to the presence of the oncogene *C-MYC* used in the reprogramming procedure[17]. Although differentiated cells derived from iPSC do not generate teratomas, it is critical to ensure that the final product does not contain undifferentiated cells that present the risk of tumorigenesis. Moreover, there is also a risk of acquiring somatic mutations and chromosomal rearrangements during the induction of pluripotency, and there have been reports of genetic and epigenetic variation between iPSC lines that were derived using the same methods[12]. Therefore, improved protocols to differentiate hPSC into desired cell types with precise identity and cellular functions are required, which could be achieved with techniques to sort the iPSC-derived cells before transplantation, using magnetic or fluorescence-activated cell sorting (MACS or FACS). Also, the risk of tumorigenicity can be tested in animal models prior transplantation[12]. Another concern arises when cells are allogeneic, potentially leading to immunological rejection unless immunosuppression or other strategies are applied[10,18], as reported elsewhere[19]. Hence, to circumvent this latter issue, patient-specific hiPSC seems to be a promising strategy for cell replacement therapy[10,17].

1.2 Early Human Development: Cleavage, Gastrulation, and Axis Formation

Human development begins with the fusion of the egg and sperm pronuclei, encompassing epigenetic reprogramming and modification, a series of cleavage divisions leading to an extensive wave of degradation of maternal transcripts, and embryonic genome activation (EGA). EGA occurs between the 4- and 8-cell stages, consisting of newly synthesized transcripts by the embryo[20,21]. Cleavage is a series of mitotic divisions that immediately follow fertilization, in which the net size of the embryo remains the same, resulting in numerous smaller cells called blastomeres. In humans, there are three cleavage divisions from 1 cell to 2 cells, 2 cells to 4 cells and 4 cells to 8 cells[20,22]. Following EGA, the embryo subsequently undergoes compaction, when blastomeres adhere to each other to form the morula, and then cavitation. Cavitation leads to the formation of a blastocyst, that comprises a fluid-filled cavity and an ICM surrounded by trophectoderm (TE) cells. TE consists of extra-embryonic cells that, upon implantation, give rise to the placental cytotrophoblast and syncytiotrophoblast. Furthermore, ICM diverges into early EPI and primitive endoderm (hypoblast) cells (Fig. 1.1), followed by implantation at approximately day 7 of development[20].

The blastocyst comprises three cell lineages: pluripotent EPI cells that form the embryo; and extraembryonic TE and primitive endoderm (PE) cells, that contribute to the placenta and yolk sac, respectively[23] (Fig. 1.1). During compaction in mouse, the first cell fate decision, which segregates the ICM and TE, involves differential Hippo signaling (active in the ICM and blocked in TE), due to mechanisms of cell-cell contact. Therefore, in the first decision, *OCT4* mutually represses *CDX2* expression, enabling some cells to become pluripotent in the ICM and other cells to originate the TE, respectively[24]. In the second cell fate decision, the segregation of the EPI and PE lineages within the ICM involves differential fibroblast growth factor (FGF) signaling, where cells receiving higher levels of FGF become PE[25]. Therefore, ICM cells express key transcription factors (TFs), such as *NANOG* and

SOX2 (EPI) and *GATA6, GATA4, SOX17* (PE)[23,26,27]. However, there has been evidence suggesting that the formation of human PE does not depend on FGF signaling, conversely to the mouse[27].

Furthermore, during gastrulation cells rearrange themselves transforming the early embryo into a multi-layered structure containing three germ layers, endoderm, mesoderm and ectoderm[28] (Fig. 1.1). At the primitive streak, EPI cells undergo an epithelial to mesenchymal transition (EMT), whereby epithelial cells are transformed into migratory mesenchymal cells[29], and then move away from the streak to become incorporated into either the definitive endoderm or the extraembryonic mesoderm and embryonic mesoderm[30]. Specific fates are related with the proximal-distal axis of the embryo where cells move through and away from the streak. These cells appear to be capable of both synthesizing and responding to FGFs, particularly signaling via FGF8 and/or FGF4. Conversely, cells that never traverse the streak are fated to form neuroectoderm and surface ectoderm[31].

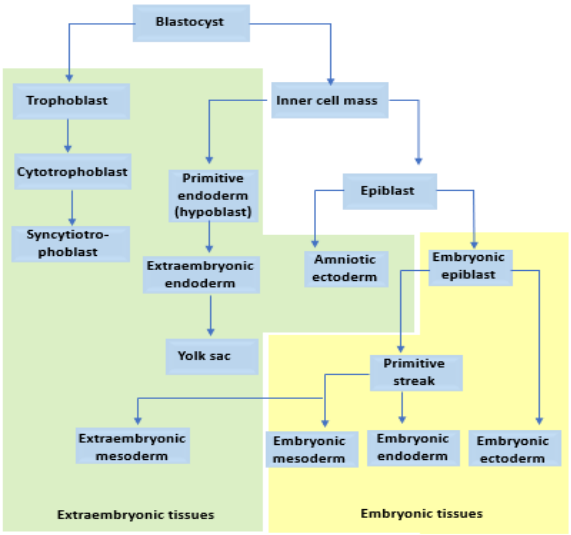


Fig. 1.1 Lineage specification and cell fate decisions.

After the implantation of the blastocyst, the trophoblast develops through various stages, generating the embryonically derived portion of the placenta. The inner cell mass (ICM) originates the epiblast (EPI) and the hypoblast (primitive endoderm). The EPI gives rise to the embryo and amnion, while the hypoblast generates the yolk sac cells. (Adapted from Gilbert and Barresi, 2016).

Additionally, embryos must develop three definitive axes: the anterior-posterior (A-P) axis, the dorsal-ventral (D-V) axis, and the right-left axis. The mammalian embryo has two signaling centers: one in the node (responsible for neural induction and for the patterning of most of the A-P axis), and one in the anterior visceral endoderm (AVE) (crucial for positioning the primitive streak)[32]. The primitive streak and other posterior tissues are the sources of wingless integrated (Wnt), bone morphogenetic proteins (BMP) and FGF proteins. Conversely, the AVE produces antagonists of these paracrine factors on the anterior side of the embryo[33].

1.3 Ectoderm Development

1.3.1 The Vertebrate Nervous System and Epidermis

Ectoderm, the outer germ layer covering late-stage gastrula, has three major roles: one part of the ectoderm will become the neural plate, which will involute and form the neural tube, the precursor of the central nervous system (CNS). The neural tube gives rise to the brain at the anterior, and spinal cord at the posterior end[34]. At the neural plate and neural tube stages, secondary organizers regulate the antero-posterior specification of distinct neural regions[35]. Another part of the ectoderm will give rise to the epidermis, the outer layer of the skin, on the ventral side. Between the compartments forming the epidermis and the CNS lies the presumptive neural crest, where the ventral and dorsal boundaries meet.

This progenitor cell population delaminates and migrates away to form, among other things, the peripheral nervous system (all the nerves and neurons lying outside the CNS), cartilage and bone of the face and pigment cells (melanocytes)[34,36]. The processes by which the three ectodermal regions are formed is called neurulation, and an embryo undergoing these processes is called a neurula.

The cells of the neural plate are characterized by expression of the Sox family of TFs (*SOX1*, *2*, and *3*), which activate the genes that specify cells to become neural plate and inhibit the formation of epidermis and neural crest by blocking the transcription and signaling of BMPs[37]. Evidence has shown that neural induction involves suppression of an epidermal fate rather than induction of a neural fate, therefore the default state of the naive ectoderm is neural, not epidermal[38].

1.3.2 Neural Tube Formation and Patterning

The neural tube is polarized along a D-V axis. The Sonic hedgehog (Shh) protein derives from the notochord and ventral floor plate and promotes ventral identity by repressing dorsal fate, whereas members of both Wnt and BMP families are secreted from the roof plate and surface ectoderm, favouring dorsal identities. Additional activities have also been reported to non-BMP members of the transforming growth factor beta (TGFβ) superfamily, as well as retinoic acid (RA)[39]. The dorsal region is associated with sensory circuits, due to BMP activity, whereas the ventral region comprises the motor neurons, favoured by Shh activity. In the middle, there are numerous interneurons[40].

Cells in the posterior regions of the neural tube begin as neuromesodermal progenitors (NMPs), that can give rise to either neural or somitic cell types from the paraxial mesoderm. NMPs are born in the caudal lateral EPI and are positively maintained by FGF and Wnt signals during tailbud elongation. Conversely, RA is expressed by somitic mesoderm and inhibits FGF signaling. Therefore, if an NMP cell enters the neural mesenchyme, it becomes a preneural progenitor cell with the ability to respond to either Shh or BMP signals. Then, preneural progenitors become proneural progenitors when exposed to lower gradients of FGF/Wnt and moderate levels of RA[41].

NSC generate new neurons or glial cells in the wall of the neural tube, in the ventricular zone. Newborn neurons can migrate away from the ventricular zone and form a new layer, the mantle or intermediate zone (gray matter), and can migrate towards the marginal zone. This process forms the cortical layers of the brain[42]. The adult brain has two NSC pools located in the sub-ventricular zone and the sub-granular zone of the dentate gyrus of the hippocampus. The majority of NSC are quiescent (type B1q cells) and generate proliferative, activated NSC (type B1a cells), which can, in turn, produce neural progenitor cells (NPC or transient amplifying progenitor [TAPs] or type C cells), a proliferative cell population that expresses markers of early neuronal differentiation. Finally, the NPC give rise to neuroblasts (type A cells), which migrate to the olfactory bulb where they become primarily interneurons[43]. The brain forms three primary vesicles, the prosencephalon (forebrain), mesencephalon (midbrain), and rhombencephalon (hindbrain). The prosencephalon and rhombencephalon become further subdivided[44].

1.4 Mesoderm Development

1.4.1 Mesodermal Subtypes

The formation of mesodermal tissues in the embryo occurs synchronously with the neural tube development. The notochord extends beneath the neural tube, from the posterior region of the forebrain into the tail. On either side of the neural tube lie mesodermal cells, that become further divided into mesodermal subtypes. Mesoderm development begins with the formation of the primitive streak, which then segregates into mesodermal subtypes - chordamesoderm, paraxial, intermediate, and lateral plate mesoderm[45-47].

Chordamesoderm, or axial mesoderm, is the central region of trunk mesoderm and generates the notochord, a transient tissue responsible for the induction and patterning of the neural tube and body axis formation[48].

Paraxial mesoderm is derived from the anterior primitive streak and segments into somites, which are epithelial building blocks on either side of the neural tube[46]. The anterior-most paraxial mesoderm does not segment, forming the head mesoderm, providing the mesenchyme and musculature of the head[49]. In the segmented paraxial mesoderm, somites are patterned along the D-V axis, where the ventral somite (sclerotome) gives rise to bone and cartilage, while the dorsal somite (dermomyotome) generates brown fat, dermis and skeletal muscle[46].

The intermediate mesoderm is adjacent to the paraxial mesoderm and forms the urogenital system - kidneys, gonads, and their associated ducts[50].

Farthest away from the notochord, the lateral mesoderm derives from the middle primitive streak and resides on either side of the two bands of intermediate mesoderm. The lateral plate mesoderm furtherly splits into the somatic and splanchnic layers, giving rise to limb bud mesoderm[51] and cardiac mesoderm[46], respectively.

1.4.2 Cardiac Development

The heart is the first organ to function during embryogenesis and is specified through the interaction between inductive and inhibitory cues from the adjacent endoderm and ectoderm. The next step in cardiogenesis is the specification and differentiation of cardiac cells, orchestrated by distinct multipotent progenitor populations, the first heart field (FHF), and the second heart field (SHF)[52-54]. The FHF forms the cardiac crescent and leads to the formation of a linear heart tube, giving rise primarily to the left ventricle and a fraction of the atria. Progenitors from the anterior SHF contribute to the right ventricle and the outflow tract, whereas the posterior SHF cells generate the atria and the inflow tract. Lineage-tracing studies indicate that both heart fields are marked by expression of *Flk-1* and the TF *NKX2-5*, while the TF *ISL1* selectively marks the SHF[55-57]. In addition, there is also a contribution from the proepicardial organ and cardiac neural crest cells to the heart development. The proepicardium is a transient structure that eventually starts to give rise to the epicardium, which, in turn, can contribute to several lineages within the heart[58]. The cardiac neural crest cells contribute to the septation of the outflow tract, to the parasympathetic innervation, and to the formation of heart valves[53]. Subsequently, the linear heart tube undergoes rightward looping and remodeling events of chamber formation, septation, and valve development lead to the formation of the four-chambered heart[53,59], where TF encoding genes such as *TBX5* and *TBX20*, in combination with *NKX2-5* and *GATA4*, have shown to play a regulatory role[60].

The early embryonic heart tube comprises the endocardium, the inner endothelial layer; myocardium, muscular middle layer; and the epicardium, outer epithelial layer, derived from the proepicardium[61]. The heart is composed by different cells – cardiac fibroblasts; atrial and ventricular myocytes; smooth muscle that generate the venous and arterial musculature; the endothelial lining of the heart and valves; and the epicardium[61,62].

1.5 Endoderm Development

1.5.1 Tubes and Organs for Digestion and Respiration

The endoderm has three major embryonic functions – to induce the formation of several mesodermal organs, such as notochord, heart and blood vessels, which, in turn, help specify the endoderm; to form the respiratory and gastrointestinal tracts, and all of their associated organs; and to generate the epithelium of several glands[63,64].

The endoderm arises from two sources. The main source is the set of cells that migrate through the primitive streak during gastrulation and segregate from bipotential mesendoderm (transient common

precursor cell population that gives rise to both mesoderm and endoderm), to form definitive endoderm (DE). This source of cells replaces part of the extraembryonic visceral endoderm that gives rise to the yolk sac[63]. The DE is a monolayer of cells on the ventral side of the developing embryo, which then forms a tube as the embryo rotates along the A-P axis and is patterned into three progenitor domains, that comprise the foregut, midgut and hindgut[65]. Nodal signaling drives the expression of a conserved network of TFs within the endodermal lineage, including MIX-like proteins: *FOXA2*, *SOX17*, *EOMES* and *GATA4-6*. These factors activate then a cascade of genes that function to segregate endoderm and mesoderm lineages, specify cells to an endodermal fate, and integrate signaling events that pattern the nascent endoderm. Therefore, the fate choice between endoderm and mesoderm appears to depend on the concentration of Nodal secreted from the visceral endoderm, where high levels of Nodal induce the endodermal network of TFs, while BMPs and FGFs act against Nodal and specify migrating cells to become mesoderm[66].

The DE of the gut is then defined along the A-P axis. This axis in vertebrates is specified by gradients of Wnts, FGFs, and BMPs, each of which has highest concentrations posteriorly. The endoderm near the head will form the anterior foregut cells, which will generate the precursors of the lung and thyroid glands. The endoderm in the posterior becomes a collection of midgut-hindgut precursor cells and will form the intestinal progenitor cells. The region between them, in the area of moderate BMPs, FGFs, and Wnts, becomes the posterior foregut precursors, generating the pancreas and the liver[65].

1.5.2 Hepatic Development

The hepatic diverticulum (or liver bud) thickens and transitions from a monolayer of cuboidal endoderm cells into a multilayer of pseudostratified cells called hepatoblasts, which delaminate, proliferate and migrate to the surrounding septum transversum, forming the liver bud[67]. Then, the diverticulum extends away from the foregut into the surrounding mesenchyme, which stimulates the endoderm to proliferate, branch, and form the glandular epithelium of the liver[66].

The mature human liver has four lobes containing different cell types, such as hepatocytes, the main parenchymal cells of the liver, constituting ~80% of its mass and regulating body metabolism; cholangiocytes, epithelial cells that line the bile ducts; sinusoidal endothelial cells[65], creating the blood channels in the liver[68]; stellate cells, the main mesenchymal component of the liver; and Kupffer cells, liver macrophages located in the sinusoids[65]. Hepatoblasts are marked by the expression of alpha-fetoprotein and albumin, as well as markers for both hepatocytes (albumin, *HNF4α*, keratin) and cholangiocytes (*CK19*). Hepatoblasts proliferate and differentiate into hepatocytes and cholangiocytes[65].

1.6 Pluripotent Stem Cell Signalling throughout In Vitro *Differentiation*

Derivation of hPSC into a specific lineage or cell type requires the regulation of distinct signaling pathways (Fig. 1.2), through specific growth factors, cytokines and small molecules[69]. Embryology has provided important insights into key pathways guiding hPSC fates, resulting in the efficient induction of ectoderm, mesoderm, and endoderm and many of their derivatives. The next challenge will be to address the functionality of these cells, both *in vitro* and in preclinical models[70].

Activin-Nodal and FGF signaling cooperate together to maintain human pluripotency[71,72]. Expression of *BRACHYURY (T)* is typically used to monitor the formation of a primitive streak-like population and the subsequent mesoderm induction. BMP4 induces the formation of a BRACHYURY$^+$ primitive streak population as well as the later development of FLK-1$^+$ mesoderm[73]. Conversely, blocking Wnt signaling at early stages of differentiation inhibits the formation of cells expressing *BRACHYURY* and mesoderm development, which demonstrates that this pathway is involved in the generation of the primitive streak and mesoderm[74]. However, other studies have reported that Wnt signaling is associated with self-renewal[75]. Activin-Nodal signaling induces a primitive streak population and the subsequent formation of mesoderm and endoderm, according to the strength of the

signal[76]. In contrast, inhibition of BMP4, Wnt and Activin signaling leads to an anterior neural-fate acquisition in the dorsal ectoderm, providing the basis of the 'default' model of neural induction. In the ventral ectoderm, where the signaling ligands escape the inhibitors, an epidermal fate is induced[34]. Following inhibition, if cells are subsequently induced by FGF, they will generate neurons, since FGF-ERK signaling has been shown to inhibit BMP signaling, thereby promoting neural differentiation[34].

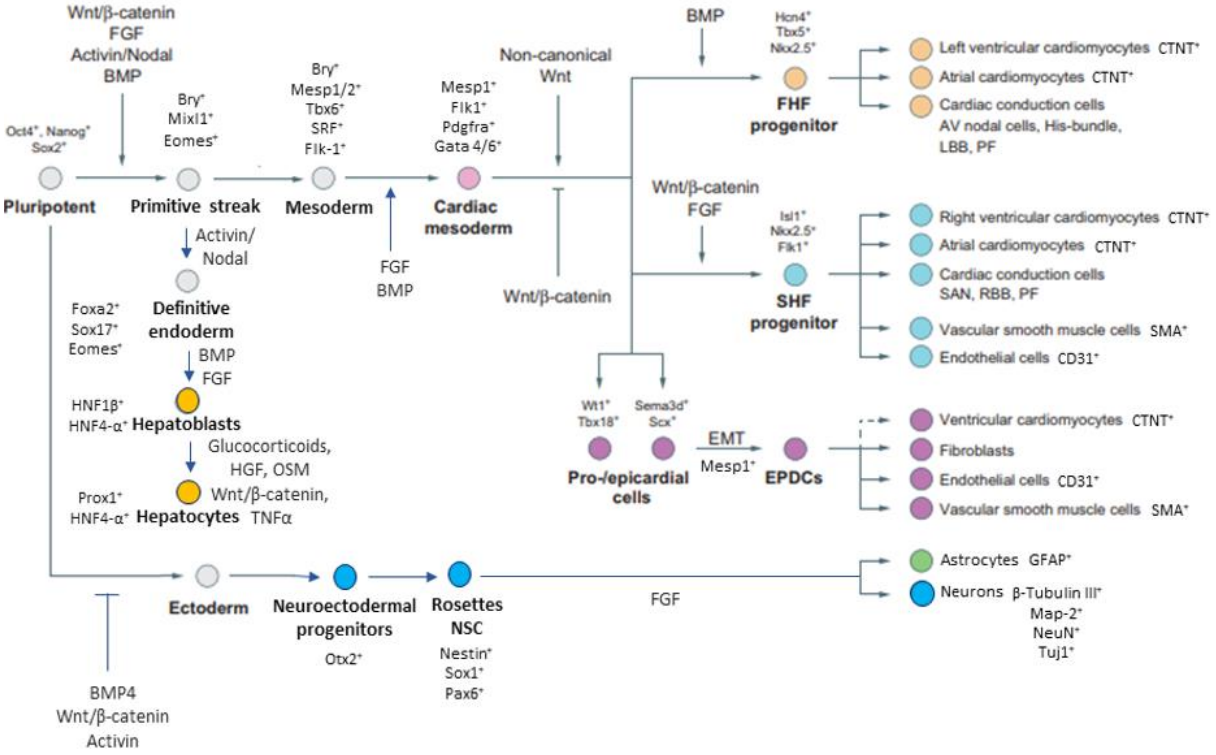


Fig. 1.2 Inducing pluripotent stem cell differentiation into cardiac mesoderm, hepatic endoderm, and neuroectoderm derivatives.

Schematic representation demonstrating the main relationships between cardiac, hepatic and neural cell lineages, sublineages, marker expression, and signaling pathways governing each cell type. When PSC are induced with Wnt, Activin A or BMP4, these cells will form a primitive streak-like population, and subsequently mesoderm or endoderm. If these pathways are not activated, PSC will differentiate into ectoderm. The differentiation of hiPSC into neurons transit through intermediate stages that resemble the neural stem cells (NSC) *in vivo*. hiPSC differentiate into neural progenitor/neuroepithelial stem cells that correspond to the NSC that form the neural plate *in vivo*. During *in vivo* neurulation, the neural tube closes, patterning along the developmental axes occurs, and the first types of neurons are generated. The correspondent step *in vitro* is the formation of neural rosettes that will give rise to neurons and glia. From primitive streak induction, the posterior primitive streak cells generate FLK-1$^+$ mesoderm, followed by the formation of cardiac mesoderm and cardiac progenitors that will give rise to differentiated cells. Alternatively, the anterior primitive streak cells may be induced by Activin/Nodal and form FOXA2$^+$ definitive endoderm, followed by the generation of hepatoblasts and hepatocytes. Abbreviations: FHF, first heart field; SHF, second heart field; EPDCs, epicardium-derived cells; EMT, endothelial-to-mesenchymal transition; SAN, sinoatrial node; RBB, right bundle branch; LBB, left bundle branch; PF, purkinje fibers; AV, atrioventricular. (Adapted from Gordillo *et al*., 2015; Später *et al*., 2014; Murry and Keller, 2008).

1.6.1 Commitment Towards Neuroectoderm

One of the most efficient ways to differentiate hPSCs towards neural commitment is by dual SMAD inhibition of Activin/Nodal and BMP signaling, which belong to the TFG-β superfamily, obtaining neural progenitors (*e.g.* PAX6$^+$ and OTX2$^+$ cells), capable of originating neurons (*e.g.* β-Tubulin III$^+$, MAP-2$^+$, NeuN$^+$ and TUJ1$^+$ cells), and astrocytes (GFAP$^+$ cells)[77]. The TGF-β family is therefore crucial for the maintenance of hiPSCs pluripotency and self-renewal, as well as for the differentiation into mesendoderm[78]. Moreover, Wnt signaling is associated with the promotion of neural crest formation in embryonic development and its continuous activation by GSK3β inhibition and Activin A signaling inhibition focus neuroectoderm commitment into neural crest cell derivation[79].

1.6.2 Commitment Towards Mesendoderm

Activin, BMP and Wnt signaling are associated with the induction of mesendoderm regulators, like BRACHYURY, EOMES and MIXL1[80]. These precursors can further differentiate into FLK1$^+$ mesoderm progenitors or FOXA2$^+$ endoderm progenitors[81]. At this point, a thin balance between mesoderm and endoderm exists, and Activin A stimulation leads to a FOXA2$^+$ population of definitive endoderm, especially if PI3K signaling is suppressed[82,83]. Conversely, BMP inhibits paraxial mesoderm formation and promotes lateral plate mesoderm specification[84]. Furthermore, Lian *et al* have demonstrated that the modulation of Wnt/β-catenin signaling directs cardiomyocyte (CM) differentiation from hPSC. Thereby, the authors reported that using B27 without insulin and inducing the sequential activation of canonical Wnt signaling by GSK3 inhibitor treatment, followed by inhibition of Wnt signaling by small molecules (IWPs) (GiWi protocol), is sufficient to drive multiple hPSC lines to CMs[85]. Therefore, specification of nascent BRACHYURY$^+$ mesodermal progenitors towards a cardiogenic fate requires inhibition of canonical Wnt/β-catenin signaling and activation of noncanonical Wnt signaling, promoting also the differentiation of those cardiac progenitors[86]. Moreover, evidence has shown that insulin, present in B27 supplement, strongly inhibits CM yield during the first 5 days of differentiation[85].

Overall, these observations show that the usual regulatory pathways control self-renewal and multiple stages of commitment, in a time, dose and cell-dependent context, modulating the stimulation or inhibition of lineages specification.

1.7 Transcriptomic Analysis by RNA-Sequencing

The transcriptome is the complete set of transcripts present in a cell, and their quantity, for a specific physiological condition or developmental stage. Recently, the development of novel high-throughput sequencing methods, namely RNA-Sequencing (RNA-Seq) has enabled both mapping and quantifying transcriptomes. The main objectives of this approach are: to catalogue all types of transcripts, including mRNAs, non-coding RNAs and small RNAs; to analyse the transcriptional structure of genes, in terms of their start sites, 5′ and 3′ ends, splicing patterns and other post-transcriptional modifications; and to quantify the differential expression levels of each transcript during development and under different physiologic conditions[87]. In addition, RNA-Seq can be implemented without prior knowledge of the reference or sequence of interest and provides multiple applications such as: ‘de novo’ reconstruction of the transcriptome (without a reference genome). Also, this technology presents some advantages over the cDNA microarrays, such as the high level of data reproducibility and reducing the number of technical replicates for the experiments[88].

A typical RNA-Seq experiment is summarised in Fig. 1.3. Generally, RNA is extracted and a population of RNA, total or fractionated, such as poly(A)$^+$, is converted to a library of cDNA fragments with adaptors attached to one or both ends. Each molecule, with or without PCR amplification, is then sequenced in a high-throughput sequencing technology to obtain short sequences from one end (single-end sequencing) or both ends (pair-end sequencing). The reads are usually 30–400 bp, depending on the sequencing technology used[87]. Theoretically, any high-throughput sequencing technology can be used for RNA-Seq, such as the Illumina IG[89], Applied Biosystems SOLiD[90] and Roche 454 Life Science[91] systems, which have already been performed for this purpose.

Global transcriptomic analysis has provided new ways to explore heterogeneous and rare subpopulations, including samples with a low number of cells, circumventing limitations associated with investigating early human development and stem cell biology[92]. This method provides us with new insights about developmental biology, and particularly, stem cell fate decisions.

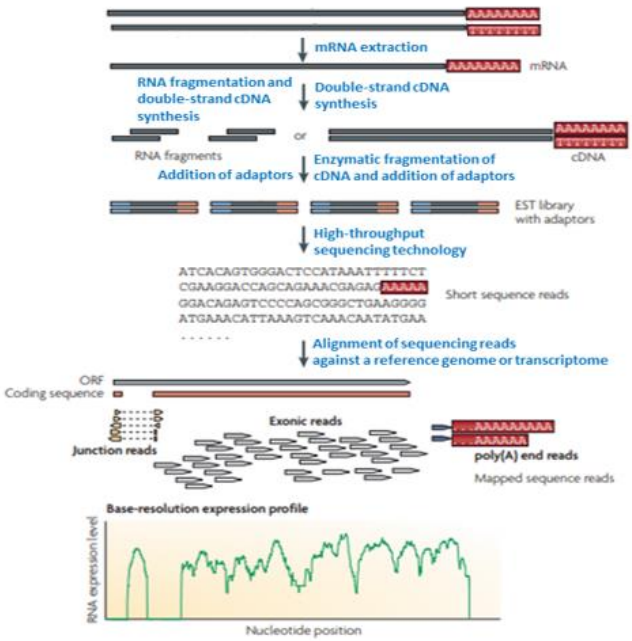


Fig. 1.3 A typical RNA-Seq experiment.
Briefly, mRNA is extracted and converted into a library of cDNA fragments through either RNA fragmentation or DNA fragmentation. Sequencing adaptors are subsequently added to each cDNA fragment and a short sequence is obtained from each cDNA using high-throughput sequencing technology. The resulting sequence reads are aligned against the reference genome or transcriptome and classified as three types: exonic reads, junction reads and poly(A) end-reads. These three types are used to produce a base-resolution expression profile for each gene, as illustrated at the bottom; a yeast ORF with one intron is shown. (Adapted from Wang *et al*., 2009).

2 Aim of Studies

Investigating the regulatory mechanisms underlying cell fate decision and subpopulation specification is fundamental for a better understanding of early human development, which cannot be interrogated *in vivo* due to restrictions on human embryo studies. This knowledge might one day be exploited for regenerative medicine, disease modeling and drug discovery fields. However, the regulatory mechanisms underlying hPSC differentiation into specific lineages remain elusive. Such need represents the motivation for initiating this study and its overall goal is to reproduce and analyse robust processes for the differentiation of hPSC.

Hence, this work aims to use hiPSC as *in vitro* model systems to analyse their controlled differentiation into specific homogeneous cell types. We aim to delineate a representative roadmap for early human development and reconstruct the differentiation trajectories by which pluripotent cells sequentially elaborate a diversity of neuroectodermal, cardiac mesodermal and hepatic endodermal progeny. Tracking these three-germ layer-derived divergences uncovers the sequential molecular events that ultimately lead to the establishment of different cell types.

Furthermore, we aim to provide a comprehensive study through transcriptomic and network modeling approaches, based on RNA-Seq data, to unveil the key signaling pathways and gene expression profiles that efficiently induce each of these lineages at sequential stages. We expect to perform this at a global gene expression level, and then focusing solely on genes encoding TFs, since these modulate a broad regulatory network orchestrating gene expression programs that elicit different cell responses. It is also expected to possibly predict novel key developmental regulators with few or no prior literature reports.

Overall, this study aims to contribute to the analysis of well-established cell culture models that better reflect *in vivo* function, presenting new insights that can be harnessed for optimization of differentiation protocols. Hence, we expect to provide important information regarding the mechanisms

that control hPSC fate transitions, with implications for understanding trans-lineage differentiation for stem cells, developmental cell biology and regenerative medicine.

3 Materials and Methods

3.1 Data Analysis Workflow

The data analysis pipeline used in this study is shown in Fig. 3.1. Thereby in the first task (prediction of lineage specifiers), we intended to expand our original gene database, containing genes sorted by their involvement in each specific germ-layer commitment. To that aim, we performed a database review through the resource Genome (NCBI) to identify human genes previously reported in the literature as being involved specifically in each germ layer commitment. For further validation, we found that these three distinct groups of selected genes from NCBI were enriched for the expected Gene Ontology (GO) terms. Therefore, ectoderm-associated gene results were predicted as being significantly overexpressed in ectoderm development (false discovery rate (FDR) of $4.16e^{-26}$); mesoderm-associated gene results were significantly enriched for mesoderm development (FDR of $5.04e^{-177}$); and endoderm-associated gene results were significantly related to endoderm development (FDR of $1.9e^{-57}$). Subsequently, these three groups of validated genes from NCBI were added to our previous/original gene database, in order to complete it with additional elements. In this way, we created a literature-based curation list of ectoderm, mesoderm and endoderm-associated gene clusters, which allowed us to expand our gene database to cover a higher number of specific germ layer-related genes (Supplementary Table 7.1, Supplementary Information).

In the second task, we elaborated a clustering analysis of this resulting expanded gene database, through STRING platform, which predicts functional protein association networks, to analyse their associated biological processes and signaling pathways enrichment (Fig. 4.1 and Supplementary Table 7.2).

The third task consisted of RNA-Seq data processing and network analysis of hPSC differentiation into specific cell types, particularly into neurons, CM and hepatocytes. This task was furtherly split into two sub-tasks: a first analysis that was based on global gene expression, and a second analysis focused solely on genes encoding TFs. The first sub-task comprised data normalization, clustering techniques such as heatmaps and PCAs (Principal Component Analysis), and Venn diagrams, together with GO enrichment of RNA-Seq gene expression. The second sub-task allowed us to identify genes encoding TFs guiding hPSC towards ectodermal, mesodermal and endodermal specification (Supplementary Table 7.3 and Supplementary Fig. 7.2).

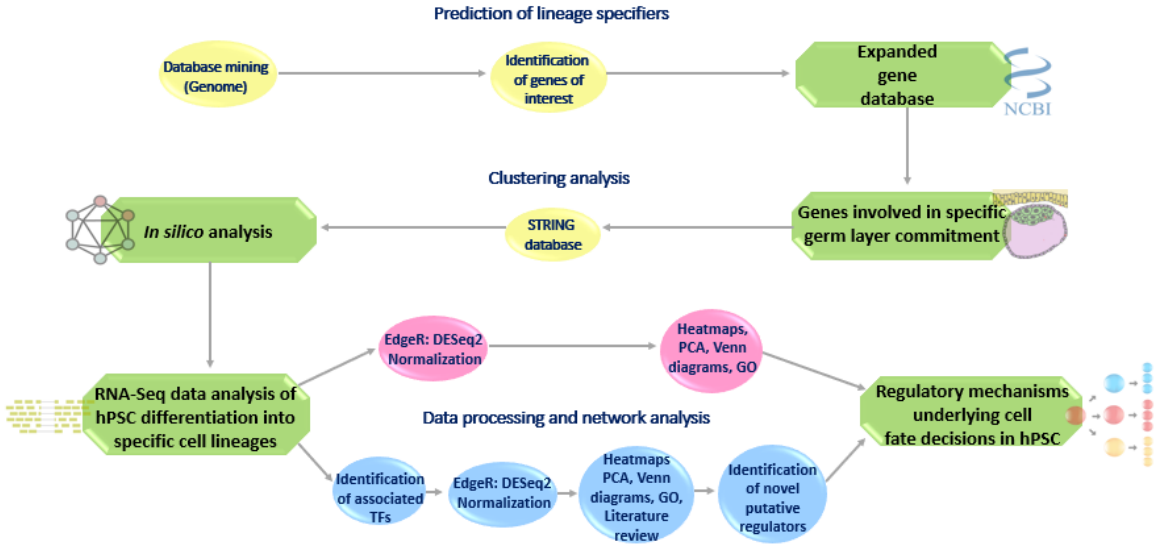


Fig. 3.1 Data analysis workflow.

Data analysis pipeline, from the gene prediction to clustering analysis, data processing and network analysis. Abbreviations: RNA-Seq, RNA-Sequencing; hPSC, human pluripotent stem cells; PCA, principal component analysis; GO, gene ontology; TFs, transcription factors.

Then, data normalization, clustering techniques (Heatmaps, PCAs), Venn diagrams and literature review of the gene functions, together with GO enrichment of RNA-Seq gene expression were performed. This led us to identify novel putative regulators of cardiogenesis and neurogenesis, whose expression was validated by Quantitative Real-Time Polymerase Chain Reaction (qRT-PCR).

Overall, this pipeline contributed to a better understanding of the regulatory mechanisms underlying cell fate decisions in hPSC.

3.2 Expansion of Human Pluripotent Stem Cells

3.2.1 Cell Lines

One of the hiPSC lines used in this work was WT- F002.1A.13 (TCLab - Tecnologias Celulares para Aplicação Médica, Unipessoal, Lda.), reprogrammed from fibroblasts obtained from a skin biopsy on an adult female. By using a retroviral system, the hiPSCs have been generated through ectopic expression of a defined set of reprogramming factors, Oct4, Sox2, Klf4 and c-Myc. The cell line has been tested in terms of differentiation potential towards the three germ layers, while the analysis revealed a normal karyotype.

Another cell line used in this study was iPS-DF6-9-9T.B, provided by WiCell Bank (Wisconsin, USA). This cell line is vector free and was reprogramed from foreskin fibroblasts with a karyotype 46, XY that were collected from healthy donors using defined factors in the Laboratory of Dr. James Thomson, at University of Wisconsin.

When referred, published datasets were used to complement our analysis, with the following cell lines: H9 and HES3 hESC lines (Meganathan *et al*., 2017)[93]; WA01 and HuES6 hESC lines (Frank *et al*., 2019)[94]; and Human H1 ES cells (Li *et al*., 2017)[95].

3.2.2 Cell Passaging

As adhesion substrate, Matrigel® matrix (Corning®) was used. Aliquots were thawed on ice and diluted 1:100 (v/v) in cold Dulbecco's Modified Eagle Medium: Nutrient Mixture F-12 (DMEM/F12) (Gibco™/ Thermo Fisher Scientific). This formulation was used to coat culture plates (Corning®), that were left at room temperature at least for two hours or stored at 4ºC for later use up to 2 weeks.

The WT- F002.1A.13 cell passaging was performed using an enzyme-free cell passaging with EDTA solution (Invitrogen™/ Thermo Fisher Scientific) diluted in phosphate-buffered saline (PBS) at a concentration of 0.5 mM. This method allows the detachment of hiPSCs colonies into small aggregates. After culture medium removal, cells were washed twice and then incubated with EDTA at room temperature for 5 min. After EDTA removal, cells were flushed with culture medium (essential 8™ medium) (E8) (Gibco®) and collected to a centrifuge tube followed by seeding onto new Matrigel® coated culture plates. Splits 1:3 were performed.

The iPS-DF6-9-9T.B cell passaging was performed by the same method, excluding the culture medium (mTeSR1™ medium) (STEMCELL Technologies™). In contrast with the mTeSR1 medium, which has 18 components added to DMEM/F12 base medium, E8 contains only 8 essential components and is a completely defined medium, without bovine serum albumin (BSA), but presenting a similar cell expression pattern when compared to mTeSR1 medium[96].

When referred, published datasets were used to complement our analysis and cells were expanded as previously described[93-95].

3.3 Differentiation Protocols

To obtain samples for RNA-Seq, neuronal differentiation of hiPSC was performed by a PhD student from our group[97], based on dual-SMAD inhibition protocol[77]. Briefly, human neural progenitor cultures were passaged at day 12 by using EDTA dissociation buffer (0.5 mM), and were re-plated in a split ratio of 1:1 into poly-L-ornithine (15 µg/mL; Sigma)-treated and Laminin (20 µg/mL, Sigma) coated plates. At day 14, when structures like neural rosettes were observable, N2B27 medium was supplemented with bFGF-Basic fibroblast growth factor (10 ng/mL, Peprotech) during 48h. At day 16, cells were again passaged by using EDTA, into new laminin-coated wells, in a split ration of 1:3. The medium was changed daily, without the addition of any small molecule or factor. At day 28, cells were split with accutase and plated into laminin-coated wells at a density of 100.000 cells/cm^2. The N2B27 medium was replaced every two days until day 70. When referred, a published dataset by Meganathan *et al*[93] was used to complement our analysis and accordingly, neuronal differentiation of hESC was performed as previously described[93]. Thereby, cortical excitatory and inhibitory neuronal differentiation was performed by modification of prior protocols[98,99].

To obtain samples for RNA-Seq, hiPSC-CM differentiation was performed by a PhD student from our group[100], based on the Wnt signaling modulation protocol previously described by Lian *et al*[85], using 6 µM of the GSK3b inhibitor CHIR99021 (Stemgent) at day 0 and 5 µM of the Wnt signaling inhibitor IWP4 (Stemgent) at day 3. Cells were seeded onto Matrigel® coated 12-well tissue culture plates (Corning®) at a cell density of 4x10^5 cells/well for a final volume of 1 mL/well (1x10^5 cells/cm^2). Culture medium was changed daily until a confluence of 90%-95% was achieved.

In addition, cardiac differentiation was also performed after RNA-Seq, for subsequent downstream analysis (by qRT-PCR and flow cytometry), as explained in sections 3.5 and 3.6, respectively. When referred, a published dataset by Frank *et al* was used to complement our analysis, whereby hESC-CM differentiation was performed as previously described[94,101].

When referred, a published dataset by Li *et al* was used to complement our study, providing us with additional RNA-Seq data that we analysed, whereby hESC-hepatocyte differentiation was achieved as previously described[95].

3.4 Transcriptomic Analysis by RNA-Sequencing

3.4.1 Sample Collection and RNA Extraction

Prior RNA isolation, cell samples from sequential stages of cardiac and neuronal differentiation were collected by students from our group, in the context of their PhD thesis, respectively[97,100].

RNA samples were isolated in triplicates from each of the selected time points. Total RNA from the samples was extracted using High Pure RNA Isolation Kit (Roche, Cat. 11828665001), according to manufacturer´s instructions.

3.4.2 RNA-Seq Sample Preparation and Sequencing

RNA libraries were prepared for sequencing using Lexogen QuantSeq 3'mRNA-Seq Library Prep Kit FWD for Illumina using standard protocols. Briefly, 500 ng of total RNA were primed with the oligo dT primer containing Illumina-compatible linker sequences. After first strand synthesis, the RNA was removed, and the second strand synthesized with Illumina-compatible random primers. After magnetic bead-based purification, the libraries were PCR amplified introducing the sequences required for cluster generation. Sequencing was performed using HiSeq (50 cycles protocol) or NextSeq (75 cycles protocol) platforms. Sequencer Software HiSeq Control Software 2.2.58 was used for base calling of samples processed on Illumina HiSeq 2000. Base calling of samples processed in NextSeq Sequencer was performed with the Real-Time Analysis (RTA) v2.

3.4.3 RNA-Seq Data Analysis

Sample read quality, reads mapping and counting were performed by a standard protocol from BlueBee Genomics Platform (http://www.bluebee.com/).

3.4.4 Datasets

The datasets used in this study were derived from RNA-Seq of sequential stages of CM and neuronal differentiation (as described in 3.2-3.4 sections). In addition, when referred, published RNA-Seq datasets were used to complement our analysis (from neural, cardiac, and hepatic fetal and adult cell types), as summarized in Supplementary Table 7.4; and (from neural, cardiac and hepatic differentiations) in Supplementary Table 7.5.

3.4.5 Differential Gene Expression Analysis

With the RNA-Seq read counts matrix, we then used the DESeq2 package of R (version 3.5.1) to perform data normalization and differentially expressed genes (DEGs) analysis. Information about DESeq2 package is available online at https://bioconductor.org/packages/release/bioc/html/DESeq2.html.

3.4.6 Bioinformatic Analysis of Transcriptomes

In order to analyse both global gene expression and solely genes encoding TFs, several approaches were performed. PCA was used to determine whether samples display greater variability between experimental conditions than between replicates of the same treatment. Therefore, plots were used to visualize gene expression values in two dimensions, representing the two principal components, that is, the two orthogonal directions of the data with the highest variance[102].

Expression heatmaps were designed as a visualization tool of patterns and relationships among high dimensional transcriptomic data. Both PCA and expression heatmaps were performed using ClustVis (http://biit.cs.ut.ee/clustvis), a web tool for visualizing clustering of multivariate data (BETA)[103]. These plots were created from read counts normalization to counts per million (CPM), or fold-change > 2.

Venn diagrams were created manually and used as a method to visualize complex genetic set relations, to compare multiple transcriptomic datasets and to identify the number of DEGs in each condition. The size of circles and areas was designed proportionally to the number of elements included.

To identify TFs guiding hPSC towards ectodermal, mesodermal and endodermal specification, genes encoding TFs were screened, according to their log_2(CPM), for each of the selected common timepoints between three types of differentiation datasets (at days 0, 1, 3 and 9). Genes encoding TFs were then identified in scatter plots (Supplementary Fig. 7.2), to compare the correlation/proximity between the expression of these TFs in each germ layer commitment. Subsequently, we performed a literature review of their previously described gene functions to confirm their role in the determined differentiation process, or to ascertain if some of them had no or few prior literature reports (data not shown).

For the whole transcriptomic analysis by RNA-Seq, significantly DEGs were selected based on the following cutoff: adjusted p-value < 0.05 and fold-change > 2.

3.4.7 Clustering Analysis - STRING Protein-Protein Associations and Gene Ontology Enrichment Analysis

GO terms were identified using the PANTHER (protein annotation through evolutionary relationships) classification system (version 13.1)[104]. GO terms were identified by analysing DEGs using the following settings: GO Biological Process, test type FISHER, reference list Homo Sapiens and FDR<0.05. Non-redundant GO enriched terms were selected and plotted using Corrplot package from R.

The STRING database (https://string-db.org/) was used to extract well-curated interactions between protein encoding genes that were identified in our transcriptomic analysis. This allowed us to construct

a regulatory network of known and predicted protein–protein interactions and to assess their enriched biological processes and signaling pathways. The network was clustered into groups of genes sharing similar profiles and cluster positions were determined by an algorithm based on global confidence binding score. Proteins were then clustered applying the Markov Cluster Algorithm (MCL), with an inflation parameter = 3 (which controls the granularity of clustering). Additionally, when referred, 10 more proteins were predicted for the networks.

3.5 Quantitative Real-Time Polymerase Chain Reaction

Total RNA from cell samples of sequential stages of CM differentiation was extracted using High Pure RNA Isolation Kit following the provided instructions. RNA was quantified using a nanodrop and 1 µg of RNA was converted into cDNA with High Capacity cDNA Reverse Transcription Kit (Applied Biosystems™/ Thermo Fisher Scientific) also following the provided instructions. PCR reactions were run using SYBR Green PCR master mix. The list of primers used is in Supplementary Table 7.6. Reactions were run in triplicate using ViiA™ 7 Real-Time PCR Systems (Applied Biosystems™/ Thermo Fisher Scientific) and data was analysed using QuantStudio™ Real-Time PCR Software (Applied Biosystems™/ Thermo Fisher Scientific). The analysis was performed using the $^{\Delta\Delta}$Ct method, therefore expression is always represented as relative to day 0 and normalized against the expression of the housekeeping gene glyceraldehyde-3-phosphate dehydrogenase (GAPDH).

For neuronal differentiation, we used previously harvested cDNA at sequential stages of differentiation and we directly performed qRT-PCR using those samples.

3.6 Flow Cytometry

For flow cytometry, differentiated CM on day 15 were collected. After culture medium removal, cells were washed with PBS and then incubated with 0.25% (v/v) trypsin-EDTA (Gibco™/ Thermo Fisher Scientific) in PBS, at 37ºC for 7 min. Then, to inactivate enzymatic activity, culture medium was added and cells were collected and centrifuged at 200 xG for 3 min. Cell pellet was washed with PBS and centrifuged again and after supernatant removal, cell pellet was fixed with 2% (v/v) paraformaldehyde (PFA) reagent (Sigma-Aldrich®/ Merck) in PBS, and stored at 4ºC until flow cytometry.

3.6.1 Cardiac Troponin T

Fixed cells were centrifuged at 200 xG for 3 min. Cell pellet was resuspended and incubated with 90% (v/v) cold methanol in Milli-Q® water at 4ºC for 15 min. Cells were then washed 3 times with flow cytometry buffer 1 (FB1), constituted by 0.5% (v/v) bovine serum albumin (BSA) solution (Sigma-Aldrich®/ Merck) in PBS, and centrifuged at 200 xG for 3 min, each time. Cell pellet was resuspended and incubated in the primary antibody mouse IgG anti-cTnT (Invitrogen™/ Thermo Fisher Scientific) diluted 1:400 (v/v) in FB2, constituted by 0.1% (v/v) Triton X-100 (Sigma-Aldrich®/ Merck) in FB1, at room temperature for 1 h. After incubation, cells were washed with FB2 and cell pellet was resuspended and incubated in the secondary antibody Alexa Fluor™ 488 goat anti-mouse IgG (Invitrogen™/ Thermo Fisher Scientific) diluted 1:1000 (v/v) in FB2, at room temperature for 30 min in the dark. Cells were washed twice with FB2 and centrifuged at 200 xG for 3 min, each time. Cell pellet was resuspended in FB1 for a final volume of 300 µL/FACS tube. Flow cytometry was performed using a FACSCalibur™ flow cytometer (Becton Dickinson) and data analysis using Flowing Software 2.0.

4 Results and Discussion

4.1 In Silico *Analysis – Prediction of Lineage Specifiers*

In order to assess hPSC fate transitions, a systematic *in silico* analysis was performed to predict which genes (Supplementary Table 7.1) and associated biological processes and/or signaling pathways have

been reported as playing an active role during specific early differentiation events. Therefore, we expanded our original gene database and created a literature-based curation list of ectoderm, mesoderm and endoderm-associated gene clusters. This approach was used to map a cellular landscape for early lineages and decisions, exhibiting heterogeneous patterns of differentiation connected to multiple biological responses (Fig. 4.1 and Supplementary Table 7.2). These responses alternated from generalized developmental processes involved in whichever fate, to germ layer-particular events, such as forebrain and telencephalon regionalization, in ectodermal-related genes; or heart development, in mesodermal-related genes. This suggests the existence of gene clusters commonly activated throughout any continuing differentiation process, while others correspond to lineage-specific regulated events, that may be induced in a transient state during commitment of pluripotent cells to a particular lineage, which is consistent with previous evidence[105]. Moreover, the size of each sub-cluster gives us indications about the number of predicted protein-protein interactions enriched expression.

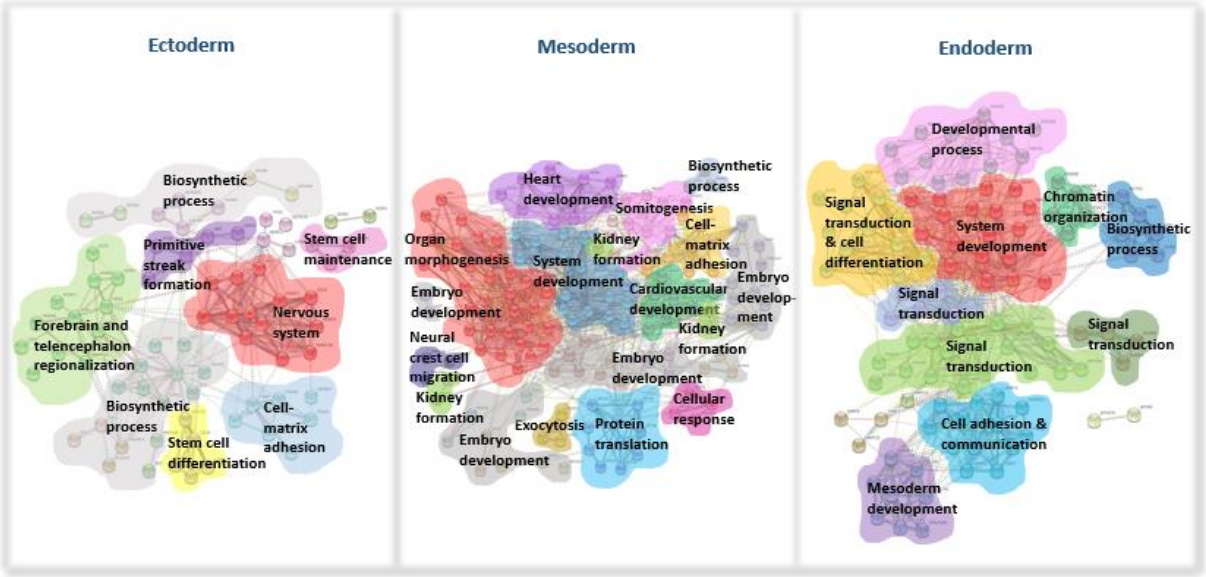


Fig. 4.1 Gene network clustering analysis.
Each cluster color represents a protein-protein interaction network prediction based on protein encoding genes curated from literature, known to be involved in human ectodermal, mesodermal and endodermal specification. These interaction networks were obtained by STRING database, that predicts associations based on *in vivo* and *in vitro* experimental assays. Proteins (nodes) were clustered using the Markov Cluster Algorithm (MLC) with an inflation parameter = 3. 10 more proteins were predicted for the network. This figure highlights the biological processes (FDR < 0.05) associated to the analysed proteins. Additional biological processes were predicted for each sub-cluster of proteins, but were here excluded for simplification purposes.

4.2 RNA-Seq Data Analysis of hPSC Differentiation into Specific Cell Types

To perform a transcriptomic analysis, we first sought to discriminate different stages of differentiation and to compare the level of *in vitro* differentiation derived from our protocols with *in vivo* human development. Therefore, we compared the RNA-Seq gene expression profiles of hiPSC-derived neurons (hiPSC-Neurons) with the transcriptomic profiles of human fetal and adult brain tissues; and of hiPSC-derived CM (hiPSC-CM) with human fetal and adult heart tissues (datasets information provided in Supplementary Table 7.4). Additionally, fetal and adult liver gene expression is shown, however until now, we have not yet sequenced our hiPSC-derived hepatocytes, which would therefore be useful for a broader comparative analysis. For that reason, our whole subsequent transcriptomic analysis will be mainly particularized to the ectodermal and mesodermal lineages.

As shown in Fig. 4.2, Principal Component Analysis (PCA) clearly distinguished the samples from different stages, with fetal, adult and differentiated cells forming distinct groups in all three differentiations (neural, cardiac and hepatic). Principal Component 1 (PC1) relates to differentiation progression, with the differentiated cells presenting the lowest degree of differentiation, the adult tissues

exhibiting the highest, and the fetal tissues with intermediate levels, consistent with previous data[106-108]. This indicated a robust directional differentiation. The only exception was between the fetal and adult heart (Fig. 4.2 C), which could be explained by cellular heterogeneity due to different cell types specified among datasets. Moreover, most differentiated cells, fetal, and adult samples clustered in well-defined "groups", although a subpopulation of cells from fetal liver exceptionally presented divergencies between replicates. Importantly, all three replicates from either differentiated CM (Fig. 4.2 C) or differentiated neurons (Fig. 4.2 A) clustered together, demonstrating no or few divergencies of our gene expression results, further validating our cardiac and neuronal differentiation methods.

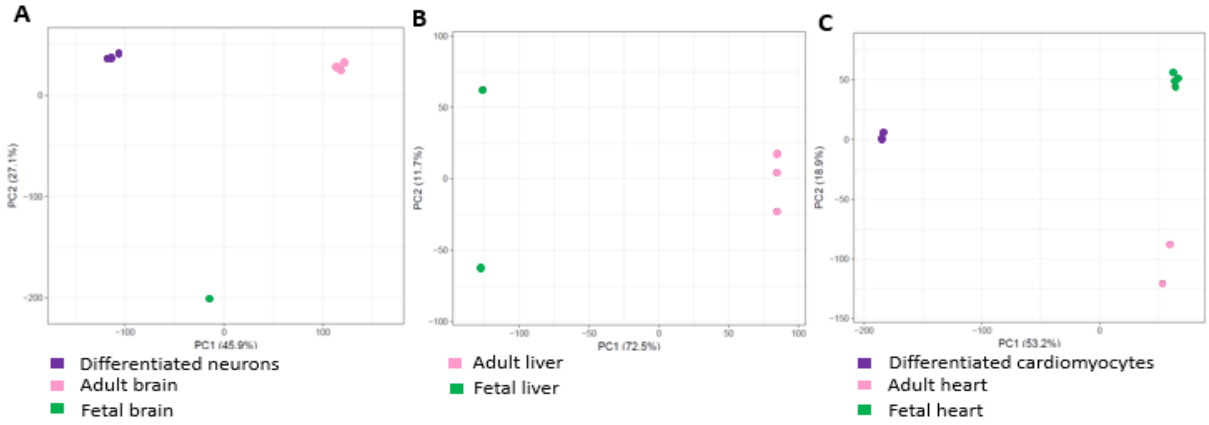


Fig. 4.2 Comparison of *in vitro* and *in vivo* expression profiling.
Principal component analysis for RNA-Seq gene expression, from reads normalization in counts per million. X and Y axis show principal component 1 (PC1) and principal component 2 (PC2) that explain (A) 45.9% and 27.1%, (B) 72.5% and 11.7%, and (C) 53.2% and 18.9% of the data set total variance, respectively. For fetal brain, n=1 and for all others, n=3.

Furthermore, we sought to compare our neuronal and cardiac differentiation protocols with distinct protocols to obtain the same desired cell types, although using different cell lines, differentiation factors and/or culture mediums to obtain them. Schematic of the protocols and stages of differentiation is depicted in Fig. 4.3A, B, D and E. In addition, a dataset from a hepatic differentiation protocol (Fig. 4.3G) and the resulting PCA (Fig. 4.3H) is shown. These variations in the protocols implied a distinct modulation of the signaling pathways and gene expression patterns that regulated stem cell fate decisions. We found that besides those differences, the PCA plot showed that distinct groups appeared in chronological order, clearly distinguishing the samples from different stages and we were able to define the lineage trajectories (Fig. 4.3C, F and H). In both neuronal and cardiac differentiation, Principal Component 2 (PC2) relates to differentiation progression, whereas in the hepatic differentiation, PC1 relates to it, indicating a robust directional differentiation. The arc connecting these experimental groups can therefore be regarded as a representative description of the differentiation track by gene expression changes. This allowed us to successfully validate our differentiation efficiency with the comparative gene expression analysis with other *in vivo* and *in vitro* types of differentiation.

To conclude this analysis, we present an overview of the intersecting DEGs between those states, at sequential stages of both differentiations (Supplementary Fig. 7.1).

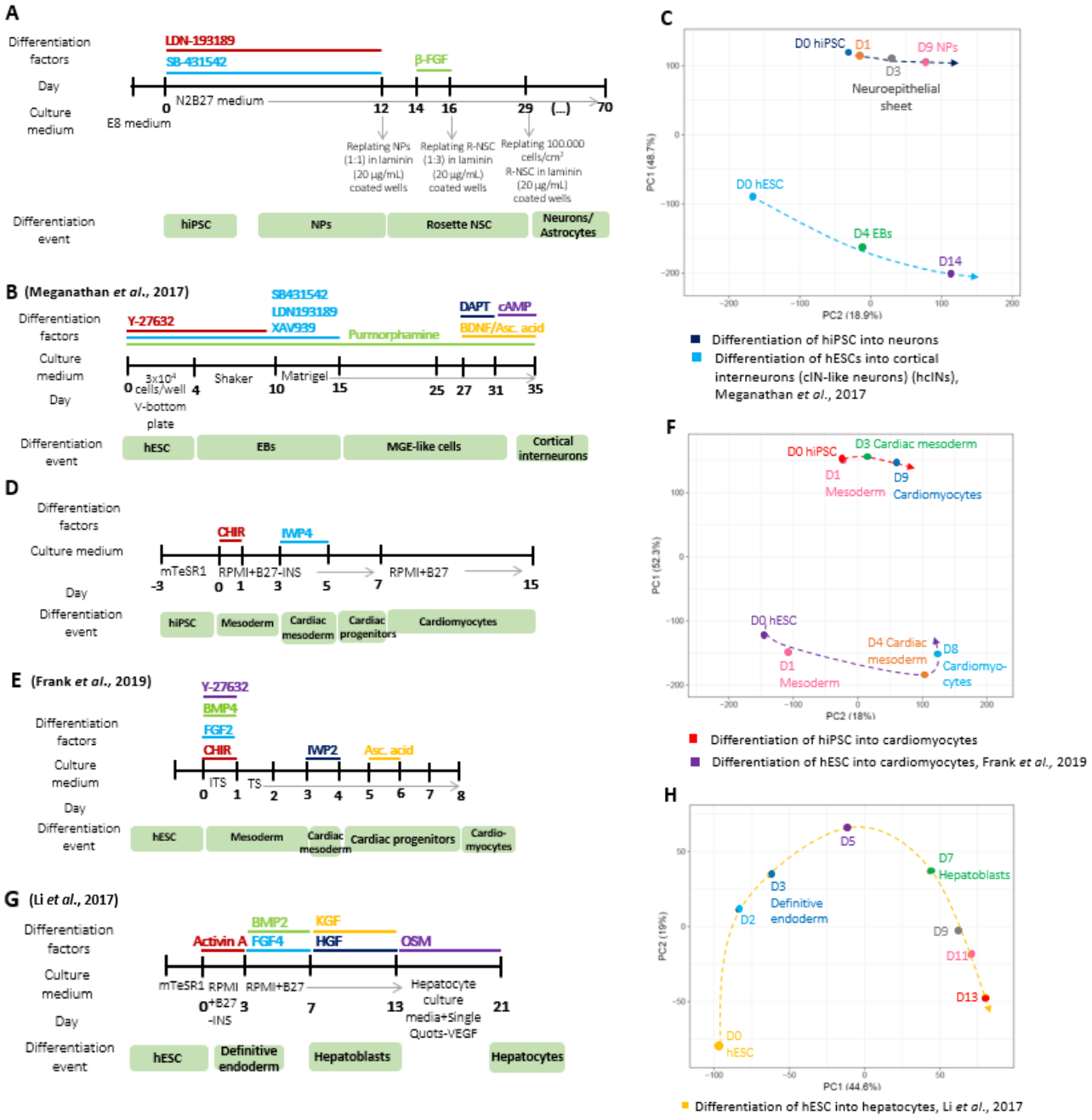


Fig. 4.3 Global gene expression profiling of human pluripotent stem cell fate transitions.

(A) Schematic of the protocol and stages of hiPSC-Neurons by our group. (B) Schematic of the protocol and stages of hESC-derived cortical interneurons, by Meganathan *et al.*, 2017. (C) Principal component analysis (PCA) indicates the transitions that occur during neuronal differentiation. X and Y axis show principal component 2 (PC2) and principal component 1 (PC1), that explain 18.9% and 48.7% of the data set total variance, respectively. (D) Schematic of the protocol and stages of hiPSC-CM, by our group and (E) of hESC-derived CM by Frank *et al.*, 2017, respectively. (F) PCA indicates the transitions that occur during cardiac differentiation. X and Y axis show PC2 and PC1, that explain 18% and 52.3% of the data set total variance, respectively (G) Schematic of the protocol and stages of hESC-derived hepatocytes, by Li *et al.*, 2017. (H) PCA indicates the transitions that occur during hepatic differentiation. X and Y axis show PC1 and PC2, that explain 44.6% and 19% of the data set total variance, respectively. PCAs were generated from reads normalization in counts per million. For all cardiac and neuronal differentiations, n=3. For hepatic differentiation, n=1. Abbreviations: hiPSC, human induced pluripotent stem cells; NPs, neural progenitors; NSC, neural stem cells; hESC, human embryonic stem cells; hiPSC-CM, hiPSC-derived cardiomyocytes; EBs, embryoid bodies; MGE, medial ganglionic eminence; Asc. acid, ascorbic acid; OSM, oncostatin M.

4.3 Key Transcription Factors in the Differentiation of hPSC

The differentiation of hPSC into specific cell types is controlled by several cytokines, growth factors, extracellular matrix molecules and TFs[69]. The activation of TFs may have important effects, such as the concomitant upregulation of genes and signaling pathways responsible for the induction and progression of cell type-specific differentiation[109]. Therefore, in this section, we intended to identify master TFs that regulate the expression of other genes, to critically discuss and evaluate their roles in the differentiation of hPSC toward neurons, CM and hepatocytes.

In order to assess hPSC fate transitions, first, we performed a systematic RNA-Seq expression analysis to identify which statistically significant upregulated genes encoding TFs played an active role guiding hPSC towards mesodermal and ectodermal specification, at sequential stages of differentiation (Supplementary Table 7.3).

Then, we sought to analyse the expression of key TFs governing different stages of differentiation. Therefore, we compared the RNA-Seq gene expression profiling of three germ layer-derived cell types, namely hiPSC-CM, hiPSC-Neurons, and hESC-derived hepatocytes, at common sequential stages of differentiation. To that aim, we selected all genes coding TFs and generated a PCA to visualize the differentiation trajectories of hPSC (Fig. 4.4).

PCA clearly distinguished the neuroectodermal, cardiac mesodermal and hepatic endodermal fates, as revealed by well-defined lineage divergences that emerged early in the differentiation process, soon after exiting pluripotency (day 0). Notably, at day 0, the expression results of ectoderm differentiation were closer to those of mesoderm, than to those of endoderm differentiation. This could be due to differences among the cell lines, especially from those employed in the external published hepatic dataset analysed in this study (using hESC instead of hiPSC), although all represent pluripotent cells at that state. Overall, PCA indicates that there are considerable differences between all types of differentiation, due to distinct networks of TFs governing specific cell types. Therefore, from pluripotency, cells navigate through germ layer specification, comprising three transcriptionally distinct

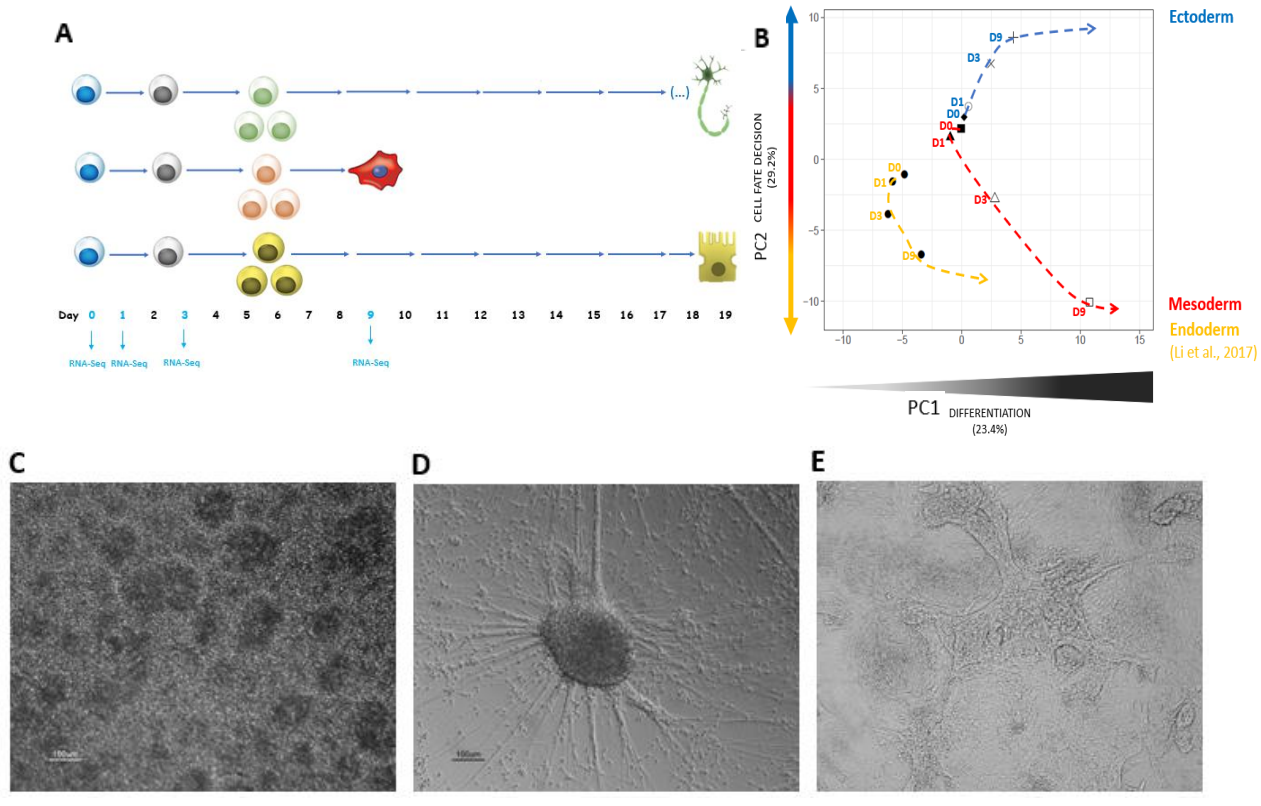


Fig. 4.4 Human pluripotent stem cell fate transitions.

(A) Schematic of the neuronal, cardiac and hepatic *in vitro* differentiation. RNA-Seq gene expression from highlighted days of differentiation (light blue) was analysed. (B) Principal component analysis for RNA-Seq gene expression, from

reads normalization in counts per million. X and Y axis show principal component 1 (PC1) and principal component 2 (PC2), that explain 23.4% and 29.2% of the data set total variance, respectively. PC1 corresponds to differentiation progression and PC2 relates to cell fate decision. For ectoderm and mesoderm, n=3; for endoderm, n=1 (published dataset from Li *et al.*, 2017). (C) Bright-field image of neural rosettes on day 15 of differentiation (D) Bright-field image on day 34 of neuronal differentiation. (E) Bright-field image of cardiomyocytes on day 15 of differentiation. Abbreviations: D, day.

subpopulations. PC1 relates to differentiation progression, which increases along time; and PC2 relates to cell fate decision, either ectodermal, mesodermal or endodermal. As stated before, the arc connecting these experimental groups can therefore be regarded as a representative description of the differentiation track by gene expression changes.

Subsequently, we focused our transcriptomic analysis particularly on ectodermal and mesodermal specification, in order to identify genes encoding TFs that were specifically up/down-regulated in each one of these two-germ layer differentiation, or commonly up/down-regulated between both germ layers commitment. Furthermore, we complemented this analysis with a GO analysis of DEGs (Fig. 4.5) for each type of differentiation. Expression heatmaps yielded four clusters of genes that were synchronously upregulated in both ectoderm and mesoderm differentiation (Fig. 4.5A), and then selected for GO analysis (Fig. 4.5B); and three clusters of genes that were simultaneously downregulated in both differentiations (Fig. 4.5C), and then selected for GO analysis (Fig. 4.5D). In addition, heatmaps depicting genes that were mutually exclusive upregulated (Fig. 4.5E) or downregulated (Fig. 4.5G) in either differentiation is shown, with the respective GO analysis of DEGs at sequential stages (Fig. 4.5F and H, respectively).

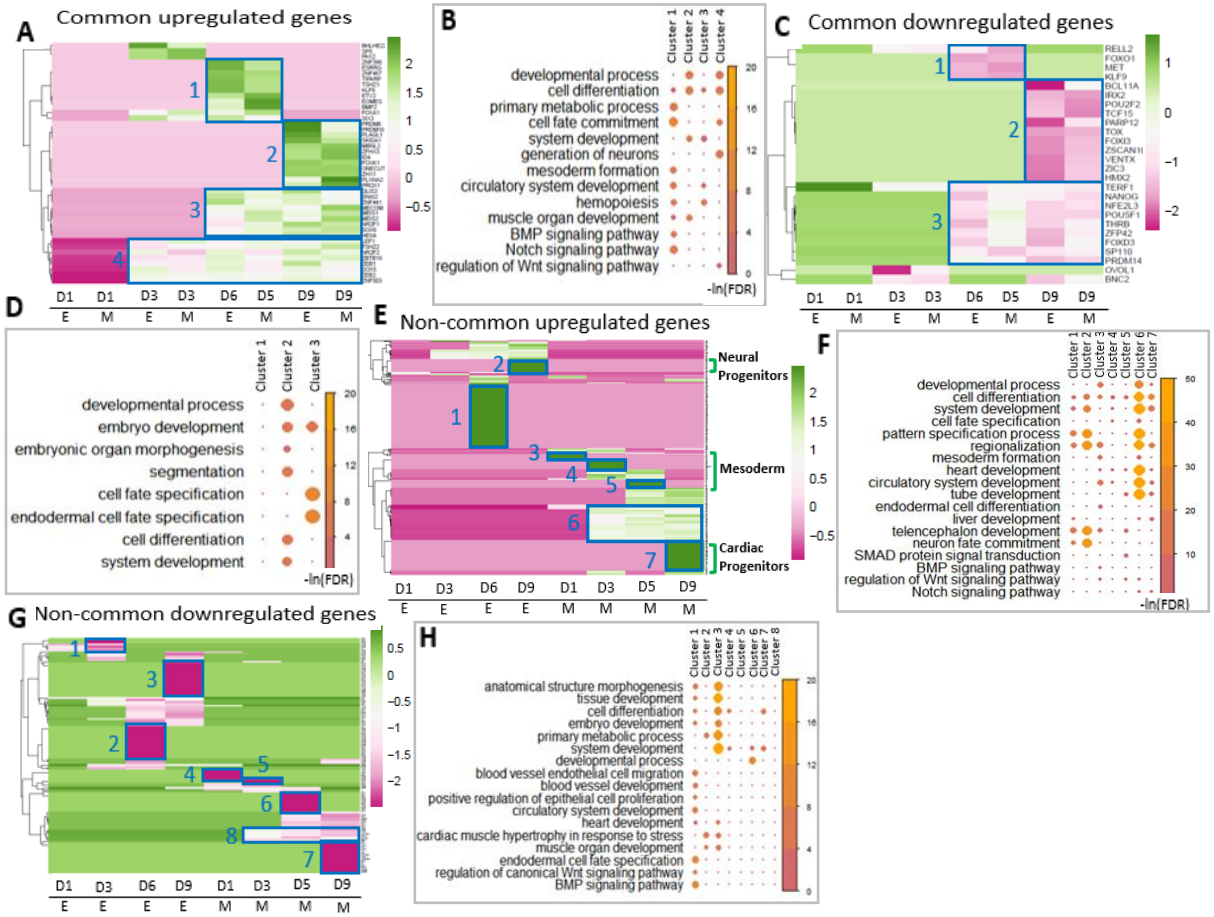


Fig. 4.5 Expression profile of transcription factors encoding genes at sequential stages of neuroectoderm and cardiac mesoderm differentiation.

(A) Heatmap comparing the expression of common upregulated genes between neuroectoderm and cardiac mesoderm differentiation, (B) and their respective clusters (highlighted in blue) were selected for Gene Ontology (GO) enrichment analysis. (C) Heatmap comparing the expression of common downregulated genes between neuroectoderm and cardiac

mesoderm differentiation, (D) and their respective clusters (highlighted in blue) were selected for GO enrichment analysis. (E) Heatmap comparing the expression of non-common upregulated genes between neuroectoderm and cardiac mesoderm differentiation, (F) and their respective GO enrichment analysis. (G) Heatmap comparing the expression of non-common downregulated genes between neuroectoderm and cardiac mesoderm differentiation, (H) and their respective GO enrichment analysis. In heatmaps, genes are displayed on the vertical axis, with the different time points on the horizontal axis. High expression levels are in dark green and low expression levels are in dark magenta. Heatmap shows average values from n=3 experiments for each time point. The color in the heatmap represents the log_2 (fold-change) of normalized expression values. For GO analysis, specific days or gene clusters without statistically significant results were not shown. Day 6 of ectoderm differentiation was compared to day 5 of mesoderm differentiation in order to obtain more early time points to analyse. Significantly DEGs were selected based on the following cutoff: adjusted *p*-value < 0.05 and fold-change > 2. Abbreviations: E, ectoderm; M, mesoderm; D, day.

Dynamic gene expression changes could be observed from day 3 until day 9 of either differentiation, generalized to all heatmaps. In particular, there were clusters of genes that were co-upregulated at the same time points in both differentiations, *e.g.* cluster 2 of Fig. 4.5A, upregulated at days 9 of both ectoderm and mesoderm differentiation. Thereby, from day 3 to day 9 of both differentiations (cluster 4), there were sustained gene networks that led to an enrichment of GO terms related to development, cell fate commitment, cell differentiation, generation of neurons and regulation of Wnt signaling (Fig. 4.5B). In a smaller temporal scale, particularly from day 5/6 to day 9 (cluster 3), there was an investment especially in hemopoiesis and circulatory system development-related processes. Specifically, at day 9 (cluster 2), there was an upregulation of genes that led to an enrichment of GO terms related to development and differentiation. At day 5/6 (cluster 1) they appear related to a wide range of already stated biological processes, including specific signaling pathways, like BMP, Notch and Wnt as well.

In addition, there were clusters of genes that were co-downregulated at the same time points for both differentiations (Fig. 4.5C), with GO terms related to endodermal cell fate specification from day 5/6 to day 9 (cluster 3, Fig. 4.5D), which is consistent with the establishment of ectodermal and mesodermal cell fate specification; and in cell differentiation, segmentation, morphogenesis and development, at day 9 (cluster 2).

Moreover, there were clusters of genes mutually exclusive upregulated at sequential stages of both differentiations (Fig. 4.5E). There was an enrichment of GO terms related to cell differentiation, system development, regionalization, telencephalon development, neuron fate commitment, etc. at days 6 and 9 of ectoderm differentiation (cluster 1 and 2, Fig. 4.5F); while processes such as tube, circulatory system, and heart development, regionalization, etc. were enriched during mesoderm differentiation (clusters 3-7). As expected, cardiac-related processes such as mesoderm formation, heart and circulatory system development appeared much more enriched during cardiac differentiation than during neuronal differentiation, especially in cluster 6 (from day 3 to 9). Similarly, neural-related processes such as neuron fate commitment and telencephalon development were more enriched during neuronal than during cardiac differentiation, especially in cluster 2, the later ectoderm analysed time point. These data indicate an efficient discrimination of GO terms in each differentiation type. During CM differentiation, there are known markers of mesoderm, such as *T, TBX6, MESP1/2*[110-112], that were upregulated from day 1 to day 5 (clusters 3-5). Coding genes for cardiac progenitors markers such as *Islet-1, NKX2−5, TBX5* and *GATA-4*[113-115], were also upregulated at day 9 of cardiac mesoderm differentiation (cluster 7). Similarly, during neuronal differentiation, coding genes for neural progenitors markers such as *LHX5, PAX6* and *SOX1*[116,117] were upregulated at day 9 of neuroectoderm differentiation (cluster 2).

Finally, there were clusters of genes that were mutually exclusive downregulated at sequential stages of both differentiations (Fig. 4.5G), for GO terms related to developmental process and cell differentiation, especially at later time points, suggesting that there is a greater investment in these processes at earlier time points of differentiation (Fig. 4.5H). As expected, cardiac-related processes such as cardiac muscle hypertrophy in response to stress, circulatory system and heart development were downregulated during neuroectoderm differentiation; and BMP and Wnt signaling pathways were

downregulated at day 3 of neuroectoderm differentiation, supporting the 'default' model of neural induction[34].

After comparing common and mutually exclusive gene clusters of either differentiation, we sought to summarize our analysis from a broader perspective. Hence, we were interested in distinguishing which biological processes were differentially enriched exclusively in each type of differentiation, or simultaneously in both, at sequential time points. To that aim, we selected a group of non-redundant processes, some of them generally associated with development, whereas others are particularly characteristic of each germ layer specification, to compare their differential enrichment throughout differentiation. Additionally, we applied the same method to ascertain which signaling pathways were regulating these differentiation processes (Fig. 4.6).

Generally, there is an upregulation of gene networks that lead to an enrichment of GO terms related to developmental process and cell differentiation along time, which is more gradual during ectoderm than mesoderm differentiation. In particular, mesoderm development increased especially from day 1 to 3 of mesoderm differentiation. Then, from day 3, cardiac mesoderm-related processes, such as heart development, appeared much more enriched, which is consistent with the addition of the small molecule IWP4 (Wnt inhibitor) at day 3 of differentiation, thus inducing cardiac mesoderm and blocking other mesodermal fates. In addition, liver and endoderm development was only associated with mesoderm differentiation, not ectoderm. This observation could be explained by an upregulation of mesendodermal regulators, since there is a transient common precursor cell population that gives rise to both mesoderm and endoderm (mesendoderm)[63]. In corroboration, evidence has shown that endoderm induces the formation of several mesodermal organs, such as notochord, heart and blood vessels. Accordingly, some mesodermal organs help specify the endoderm[64], further supporting the observed correlations.

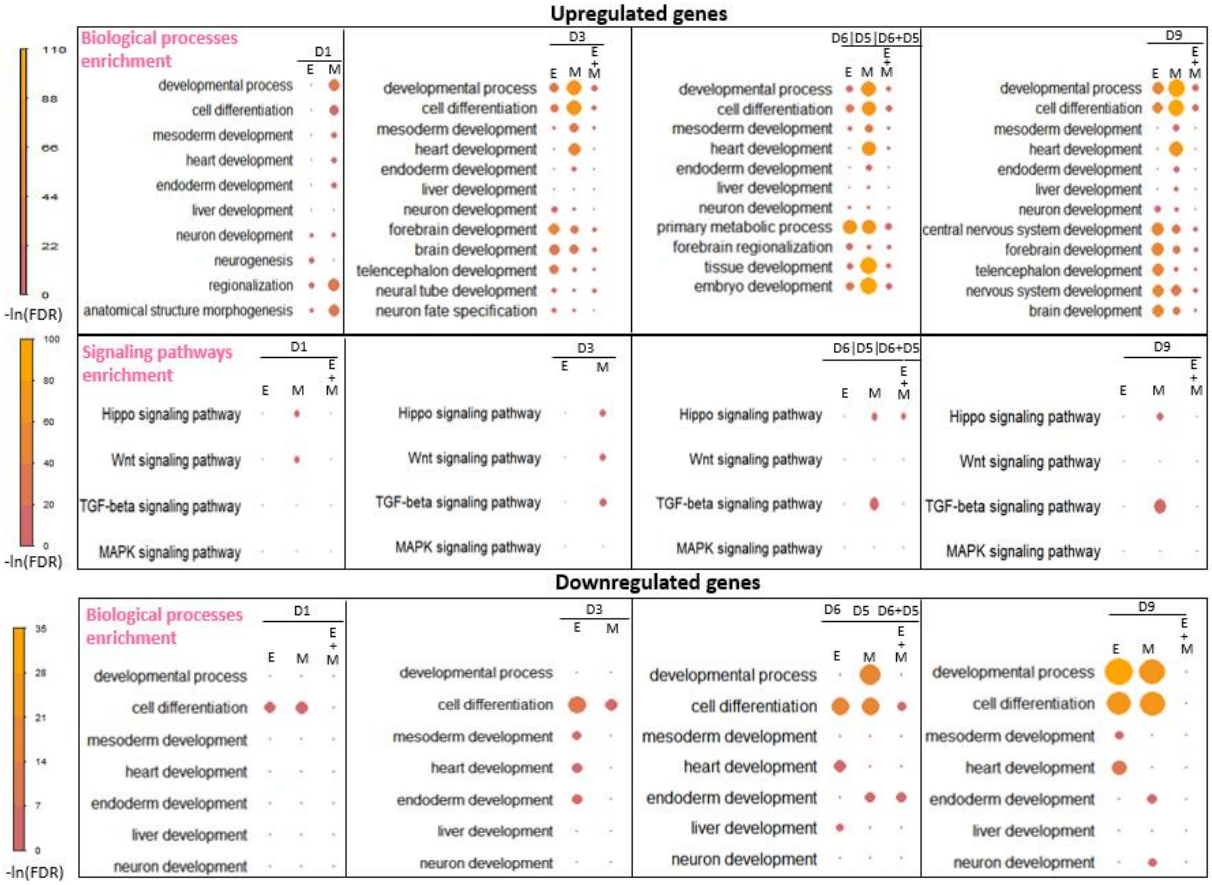


Fig. 4.6 Gene ontology analysis of key regulators of ectoderm and mesoderm differentiation.
Differentially expressed genes (DEGs) encoding transcription factors (TFs) in ectoderm and mesoderm differentiation were grouped into Gene Ontology (GO) categories (*p*-value < 0.05). These graphs represent the statistically significant

results for biological processes and signaling pathways at days 1, 3, 6 and 9 of ectoderm differentiation, and days 1, 3, 5 and 9 of mesoderm differentiation. Day 6 of ectoderm differentiation was compared to day 5 of mesoderm differentiation in order to obtain more early time points to analyse. Non-statistically significant results from GO on the analysed time points were not shown. High enrichment levels are in orange and low enrichment levels are in pink. Significantly genes were selected based on the following cutoff: adjusted *p*-value < 0.05 and fold-change > 2. Abbreviations: D, day; E, ectoderm; M, mesoderm; E+M, common between ectoderm and mesoderm.

Similarly, liver development was only enriched from day 5 of mesoderm differentiation, probably due to the same reasons. Conversely, several neuroectoderm-related processes appeared much more enriched during ectoderm differentiation, at distinct time points of differentiation. However, they were particularly enriched at later time points (*i.e.* day 9), when a significant enrichment is generally observed among multiple neural-related processes, such as the development of the central nervous system, forebrain, telencephalon, neuron, brain, etc.

Some genes were downregulated at distinct time points between both differentiations with GO terms related to developmental process and cell differentiation. As stated before, this downregulation was especially observed at later time points. As expected, genes associated with mesoderm and cardiac-related processes, such as mesoderm and heart development, were downregulated from day 3 to 9 of ectoderm differentiation. Similarly, genes involved in neural-related processes, such as neuron development, were downregulated at day 9 of mesoderm differentiation, further corroborating our analysis. Interestingly, genes related to endoderm development were downregulated as soon as day 3 of ectoderm differentiation; and from day 5 of mesoderm differentiation, probably in order to induce and maintain mesoderm instead of endoderm, from the mesendoderm state.

Regarding the regulation of signaling pathways throughout differentiation, Hippo signaling was active in a sustained way during mesoderm differentiation. Indeed, this pathway has been shown to play important roles in skeletal and cardiac muscles, both derived from mesoderm differentiation[118]. Additionally, this pathway has been proven to regulate CM proliferation and heart size[119]. TGF-β was gradually active along mesoderm differentiation, from day 3 to 9, as expected, since mesoderm commitment can be achieved by Activin/NODAL and BMP signaling pathways[120]. Wnt signaling pathway was active at days 1 and 3 of mesoderm differentiation and was subsequently downregulated after the latter time point. This is consistent with our employed protocol to derive CM, using the GSK3b inhibitor CHIR99021, that stimulates Wnt activation at day 0, and then the Wnt signaling inhibitor IWP4 at day 3 of differentiation[121]. Genes coding for TFs involved in MAPK signaling pathway were not upregulated. MAPK/ERK signaling has been shown to be involved in CM survival[122]. Remarkably, none signaling pathway played an active role during neuroectoderm differentiation, which corroborates the 'default' model of neural induction. According to this model, inhibition of BMP4, Wnt and Activin signaling leads to a neural-fate acquisition[34].

To conclude this analysis, we present an overview of the intersecting DEGs encoding TFs, modulated at sequential stages of both differentiations (Supplementary Fig. 7.3). Overall, it is possible to assess the differential expression of key regulators governing the controlled differentiation of hPSC into specific cell types, particularly into CM and neurons.

Taken together, these results demonstrate that we were able to compare and define the major gene-expression changes of the key regulators accompanying the transition from a pluripotent to an ectodermal versus mesodermal state and to discriminate both populations in terms of biological process enrichment.

After confirming that distinct networks of TFs regulate cell fate decisions, we sought to assess which particular TFs were mainly governing each of these specific lineage choices. Those TFs were then identified in scatter plots, to compare the correlation/proximity between the expression of these TFs in each germ layer commitment (Supplementary Fig. 7.2). Subsequently, we performed a literature review of their previously described gene functions to confirm their role in the determined differentiation process, or to ascertain if some of them had no or few prior literature reports (data not shown). We found

that some of them have already been reported as key lineage regulators, which our analysis has confirmed, whereas others have not, thus requiring further investigation. In sum, this strategy allowed us to efficiently discriminate TFs acting as key regulators of hPSC fate transitions, through the comparative analysis between ectodermal, mesodermal and endodermal-derived cell lineages.

4.4 Prediction of Novel Regulators

In order to fill the knowledge gap regarding the transcriptional regulation during early CM and neuron differentiation, we profiled dynamic transcriptional changes at sequential stages. Two high-throughput sequencing datasets were compared for each type of differentiation: one derived from differentiated cells by our group; and another from a reported work (detailed information in Supplementary Table 7.5). This strategy was used as a proof of concept to accurately investigate temporal changes in transcription, that have been determined and thus supported by both datasets, further validating our observations.

We first sought to perform a high-throughput screen for selected lineage-specific markers, to compare their expression at similar time points between both datasets: between our dataset and one obtained in the literature, for both neuronal and cardiac differentiation (Fig. 4.7A and 4.8A, respectively).

Regarding neuronal differentiation, we selected the genes *OTX2*, *SATB2*, *NESTIN*, *PAX6*, *MAP2*, *SOX1*, *VGAT* and *NKX2.1*, since they are generally associated with early neural differentiation[123]. Their positive expression was assessed to validate each type of differentiation, since these markers must be expressed independently of differences among protocols. Although their significant expression has been successfully confirmed, some slightly different patterns of expression have arisen, probably due to differences in the differentiation protocols and cell lines used (Fig. 4.7A). Indeed, to differentiate hESC into cortical interneurons, the authors used ventral NPC medium to generate embryoid bodies with ventral telencephalic characteristics. Also, they used distinct differentiation factors, which modulated TGF-β, Wnt, Hedgehog and Notch signaling pathways. Conversely, to differentiate hiPSC into neurons, we used N2B27 medium and we added small molecules to modulate TGF-β signaling pathways, based on dual-SMAD inhibition protocol[77]. *NKX2.1* has been reported to be expressed in the medial ganglionic eminence (MGE)[124], which is consistent with a higher expression of this TF in the MGE-derived interneurons, in comparison with the hiPSC-Neurons. Also, at the neural tube stage, *PAX6* is downregulated in the region that will become the ventral telencephalon, concomitant with the upregulation of *NKX2.1* in this region[125], which is corroborated by a lower expression of *PAX6* and a higher expression of *NKX2.1* in the hESC-derived cortical interneurons, compared with the hiPSC-Neurons. Moreover, *SATB2* expression was slightly higher in the hESC-derived cortical interneurons, as expected, since it represents a cortical tissue patterning marker[126]. Furthermore, neural progenitor cell markers, such as *SOX1*, *NESTIN* and *PAX6*[123]*;* neuronal markers, such as *MAP2*[123] and *OTX2*[127]*;* and GABAergic neuronal markers, such as *VGAT*[128] were expressed in both differentiation datasets. Therefore, these results suggest that both hiPSC-Neurons and hESC-derived cortical interneurons expressed lineage-specific gene markers, thus indicating an efficient neuronal differentiation and validating both protocols.

After comparing the expression of lineage-specific gene markers between these two distinct neuronal differentiation protocols and assessing their major differences, we developed a method to accurately analyse common dynamic transcriptional changes in both datasets, that could be specifically related to neurogenesis. Our employed strategy is summarised in Fig. 4.7B. Therefore, we present a comparison of temporal expression changes between both datasets, regarding the initial list of candidates that likely are involved in neuronal differentiation. For simplification purposes, only the main peaks of gene expression are depicted here (Fig. 4.7C). Despite differences between the differentiation protocols and cell lines, both differentiated neurons and cortical interneurons still exhibited high concordant transcriptomes during neuronal differentiation, as highlighted by specific peaks of gene expression occurring at approximately the same time points of differentiation, thereby displaying similar trends of

expression (Fig. 4.7C). Moreover, we have identified the final candidates likely associated with neuron differentiation, revealing three novel putative regulators (*TERF2IP*, *RFX4* and *ZHX1)* (Fig. 4.7B), and their respective RNA-Seq gene expression (Fig. 4.7D).

Overall, this transcriptome characterization defines a roadmap for neuroectoderm development, revealing the temporal peaks of expression of both novel candidates and lineage-specific gene markers and highlighting an efficient differentiation of stem cells into neural derivatives.

Subsequently, the same approach developed to analyse both neuronal datasets was also applied to investigate both cardiac differentiation datasets, comparing their transcriptional changes (Fig. 4.8).

Regarding cardiac differentiation, we selected the genes *CDX2*, *MIXL1*, *MSX1*, *MESP1*, *TBXT*, *NKX2.5*, *ISL1*, *HAND2* and *MEF2C*, since they are generally associated with early CM

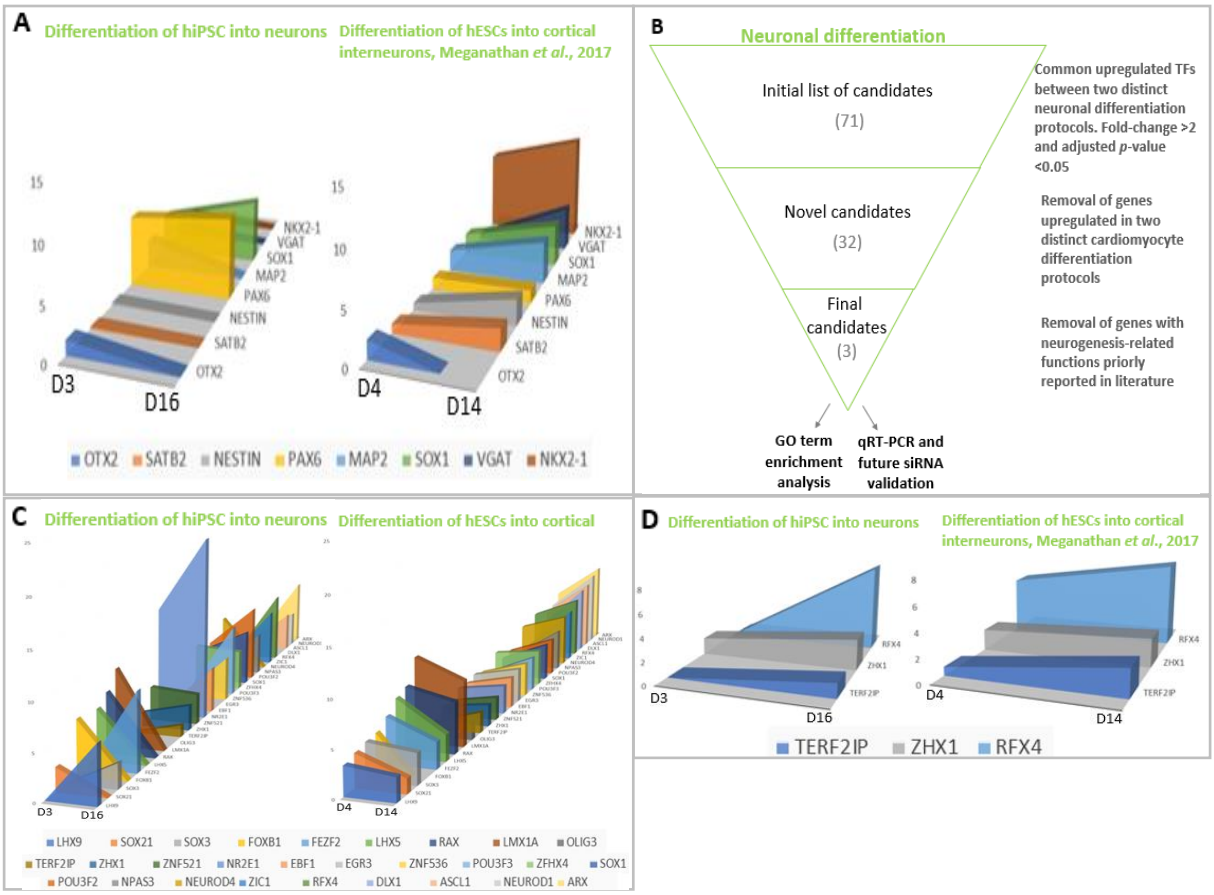


Fig. 4.7 Transcriptome characterization in early neuronal differentiation.
(A) High-throughput screen for selected lineage-specific neuroectoderm gene markers. Dynamic comparison between two datasets, yielding hiPSC-Neurons and hESC-derived cortical interneurons. (B) Schematic view of the proposed method for predicting neural lineage specifiers. As a way to establish gene profiles associated with neuronal differentiation, candidate lineage specifiers were identified in three steps. We first compared two datasets of results to identify common upregulated TFs between two distinct neuronal differentiation protocols. Then, we performed a subtractive analysis of gene profiles that were upregulated in two distinct cardiomyocyte differentiation protocols. This strategy has allowed us to eliminate large numbers of genes that were overexpressed in either dataset of results and to uniquely associate different gene networks with neural commitment. Finally, we subtracted genes with neurogenesis-related functions priorly reported in literature, in order to identify putative novel regulators of neurogenesis. (C) Comparison of dynamic expression of the initial list of candidates between two datasets, identified in our proposed predictive method. (D) Comparison of dynamic expression of the final list of candidates between two datasets, identified in our proposed predictive method. In all RNA-Seq expression graphs, X axis shows the differentiation time points, y axis shows log_2 (fold-change), and z axis shows genes encoding transcription factors.

differentiation[58,121]. During cardiac differentiation, both differentiated cells expressed lineage-specific gene markers, revealing an efficient cardiac differentiation from both protocols. To differentiate hESC

into CM, the authors used differentiation factors, which modulated Wnt, FGF and BMP signaling pathways. In contrast, to differentiate hiPSC into CM, we used small molecules to modulate the Wnt signaling pathway. Despite these differences in the cell lines and differentiation protocols, both hiPSC-CM and hESC-derived CM (hESC-CM) still exhibited similar patterns of expression, with specific genes being up/down regulated at similar time points of differentiation (Fig. 4.8A). Indeed, in both datasets results, some genes display a higher expression in the first stage of differentiation, such as: *TBXT*, primitive streak marker, and *MIXL1*, mesendoderm marker[121,129]. Then, followed a group of genes that exhibited their highest expression approximately at day 3 of differentiation, such as: *MSX1*, mesendoderm marker, and *MESP1*, mesoderm and cardiac mesoderm marker[121,130]. *CDX2* expression was high at the beginning of the differentiation and after day 3, it diminished, consistent with an early

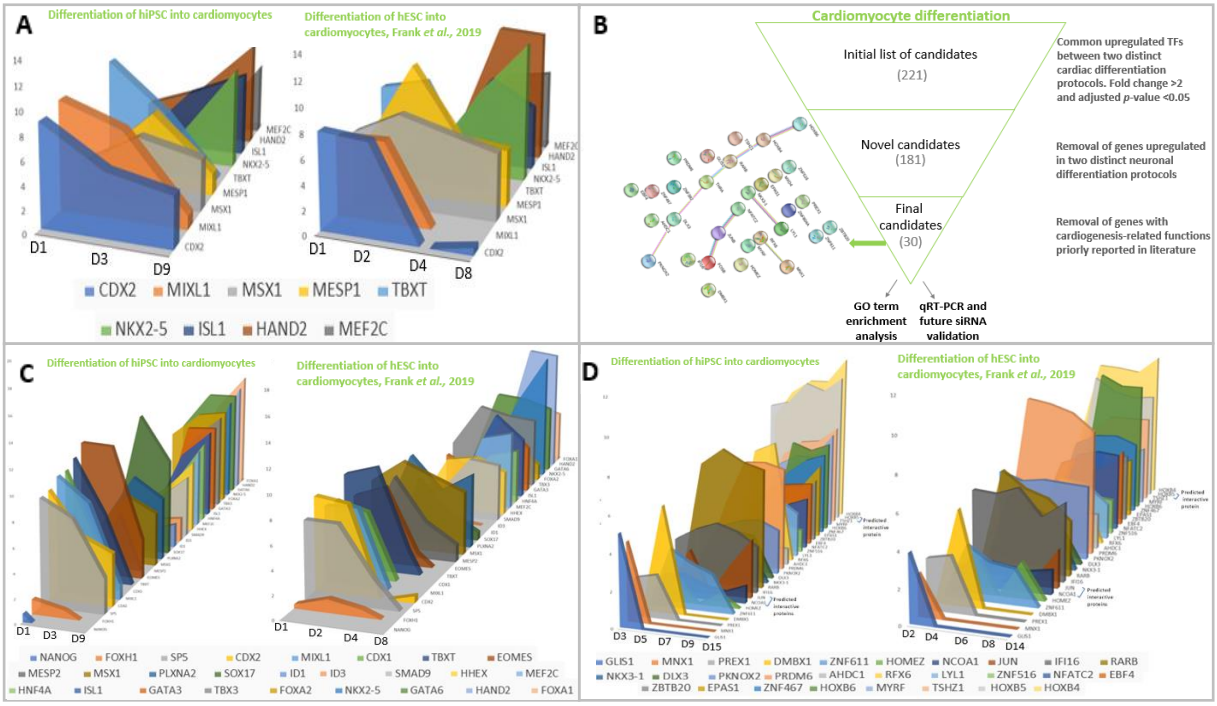


Fig. 4.8 Transcriptome characterization in early cardiac differentiation.
(A) High-throughput screen for selected lineage-specific cardiac mesoderm gene markers. Dynamic comparison between two datasets, yielding hiPSC and hESC-derived CM. (B) Schematic view of the proposed method for predicting cardiac lineage specifiers. As a way to establish gene profiles associated with cardiac differentiation, candidate lineage specifiers were identified in three steps. We first compared two datasets of results to identify common upregulated TFs between two distinct cardiac differentiation protocols. Then, we performed a subtractive analysis of gene profiles that were upregulated in two distinct neuronal differentiation protocols. This strategy has allowed us to eliminate large numbers of genes that were overexpressed in either dataset of results and to uniquely associate different gene networks with cardiac commitment. Finally, we subtracted genes with cardiogenesis-related functions priorly reported in literature, in order to identify putative novel regulators of cardiogenesis. It is also shown the final candidates' gene network clustering analysis, obtained by STRING database, that predicts associations based on *in vivo* and *in vitro* experimental assays. (C) Comparison of dynamic expression of the initial list of candidates between two datasets, identified in our proposed predictive method. (D) Comparison of dynamic expression of the final list of candidates between two datasets, identified in our proposed predictive method. The expression results of interactive proteins predicted in STRING database (JUN, NCOA1 and HOXB5) were added to this graph. In all RNA-Seq expression graphs, X axis shows the differentiation time points, y axis shows log_2 (fold-change), and z axis shows genes encoding transcription factors.

Wnt activation stimulating mesoderm induction, and later Wnt inhibition (from day 3) to induce cardiac mesoderm[131]. At the final stages of differentiation, cardiac mesoderm and cardiac progenitors gene markers, such as *NKX2-5*, *ISL1*, *MEF2C* and *HAND2*[121,129], reached their highest levels of expression. In sum, these results confirm an efficient CM differentiation in both datasets expression results.

After comparing the expression of lineage-specific gene markers between these two distinct cardiac differentiation protocols, we adapted our previously developed method to accurately analyse common

dynamic transcriptional changes in both datasets, that could be now specifically related to cardiogenesis. Our employed strategy is summarised in Fig. 4.8B. Therefore, we present a comparison of temporal expression changes between both datasets, comprising some initial candidates that likely are involved in cardiac differentiation. For simplification purposes, only the main peaks of gene expression are depicted here (Fig. 4.8C). It is also presented the protein association networks of the final candidates likely associated with cardiac differentiation, revealing thirty novel putative regulators (Fig. 4.8B), and their respective RNA-Seq gene expression (Fig. 4.8D). Despite differences between the differentiation protocols and cell lines, both hiPSC-CM and hESC-CM exhibited high concordant transcriptomes during cardiac differentiation, as highlighted by specific peaks of gene expression occurring at approximately the same time points of differentiation, thereby displaying similar trends of expression, either regarding the initial or the final list of candidates expression (Fig. 4.8C and D, respectively).

Taken together, our results provide a comprehensive view of expression changes during neurogenesis and cardiogenesis that extend previous studies, identifying key TFs and defining a roadmap for neuroectoderm and cardiac mesoderm development. Furthermore, we developed a method to screen for putative novel regulators of neurogenesis and cardiogenesis, that will be further validated *in vitro.*

4.5 Validation Experiments

To validate the expression of some of the final candidate genes identified through transcriptomic analysis, at different stages of differentiation, we performed qRT-PCR. Therefore, RNA from cell samples of sequential stages of CM differentiation was extracted, converted to cDNA and amplified. The analysis was performed using the $^{\Delta\Delta}$Ct method, therefore expression is always represented as relative to day 0 and normalized against the expression of the housekeeping gene GAPDH. Moreover, in order to evaluate the efficiency of our cardiac differentiation, the presence of the specific marker of CMs (CTNT) was assessed by flow cytometry on differentiated cells collected at day 15. Samples presented approximately 30% CTNT^{+} cells (Fig. 4.9).

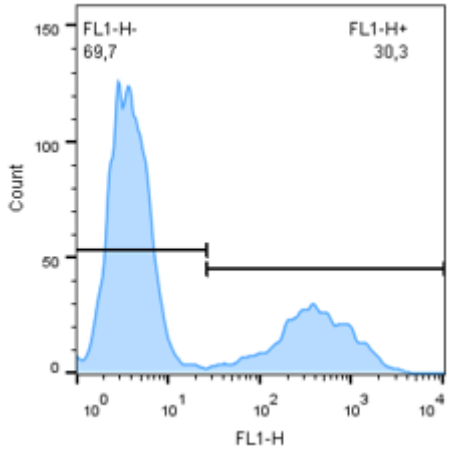


Fig. 4.9 Cardiomyocyte differentiation efficiency for DF6 cell line assessed by flow cytometry.

For neuronal differentiation, we used previously harvested cDNA at sequential stages of differentiation and we directly performed qRT-PCR using those samples, since an efficient neuronal differentiation has been previously demonstrated by our group[97].

Therefore, we assessed the expression of final candidate genes: *TERF2IP*, *RFX4* and *ZHX1*, regarding neuronal differentiation; and *HOXB2-6* and *RARB*, for cardiac differentiation. In addition, we also analysed the expression of genes specific for neuronal and cardiac differentiation, *PAX6* and *TNNT2*, respectively. Primers used for qRT-PCR are presented in Supplementary Table 7.6. Globally, the results obtained by qRT-PCR validated the RNA-Seq data.

From day 5 of cardiac differentiation, there was expression of transcripts representing CM (*TNNT2*), achieving a peak of relative expression at day 15 (Fig. 4.10A). Moreover, we decided to test *HOXB2-6* (homeobox genes) and *RARB* (Retinoic Acid Receptor Beta) expression at the transcript level, since evidence suggests that *HOXB* genes have close associations with RARB[132]. Indeed, transgenic analyses in mice identified novel enhancers, that respond to exogenously added RA and modulate the expression of genes from the *HOXB* complex during embryonic cardiac development[133]. However, in humans, these relationships have not been proven yet. Additionally, *HOXB4* and *HOXB6* were also final candidate putative regulators identified in our analysis and *HOXB5* was predicted by STRING database as an interactive protein. Therefore, to avoid redundant functions between these *HOXB* genes, we decided to assess *HOXB2-6* expression.

HOXB6 expression levels were lower than from the remaining *HOXB* genes (*HOXB2-5),* which is consistent with the RNA-Seq data derived from both datasets (data not shown). *HOXB2*-5 and *RARB* relative expression started on day 1, reached a peak of expression at day 5, reaching a plateau until day 15. Notably, *HOXB4* relative expression levels were higher than the remaining genes, and *HOXB4* and *HOXB5* expression levels slightly increased particularly from day 9 to day 15, in agreement with the transcriptomic data. Moreover, evidence shows that *HOXB2* has been previously identified as a potential cardiac regulator[134,135], which our qRT-PCR data also corroborate. In addition, *HOXB3* has been shown to be involved in the development of blood vessel, vasculogenesis and angiogenesis[136].

Regarding neuronal differentiation, since it was not possible yet to assess the relative expression of genes during more time points throughout neuronal differentiation, we only present the results from day 12 and 16 in comparison with day 0, which is not enough to define an expression kinetic. However, it was noticeable that the relative expression of *TERF2IP* and of neural TF *PAX6* was lower on day 16 than on day 12 (Fig. 4.10B). Conversely, the expression level of *RFX4* was higher on day 16 than on day 12. Indeed, during RNA-Seq, *RFX4* was significatively expressed from day 9 until day 16, increasing progressively along those two time points (data not shown). Hence, regarding *RFX4*, our transcriptomic data revealed a log_2(fold-change) of 5.8 on day 12 and of 7.9 on day 16 in hiPSC-Neurons (data not shown), consistent with the qRT-PCR results. In addition, *RFX4* has been reported to be required for neural tube and brain development and morphogenesis, as a modulator of Shh signaling during development of the CNS, and as a regulator of ciliogenesis, playing also ciliogenesis-independent roles during neural development[137-139].

ZHX1 did not reveal a significative expression in qRT-PCR. Moreover, *TERF2IP* did not match the transcriptomic data, since its expression was not detected on day 12, only on day 16 during RNA-Seq. *TERF2IP* has been reported as being involved in telomere length regulation[140]. Despite this exception, overall our qRT-PCR results validated the RNA-Seq data. However, it is important to consider that the flow cytometry and qRT-PCR results were obtained from only one experiment (n=1) and should be furtherly validated with more experiments in order to become statistically significative.

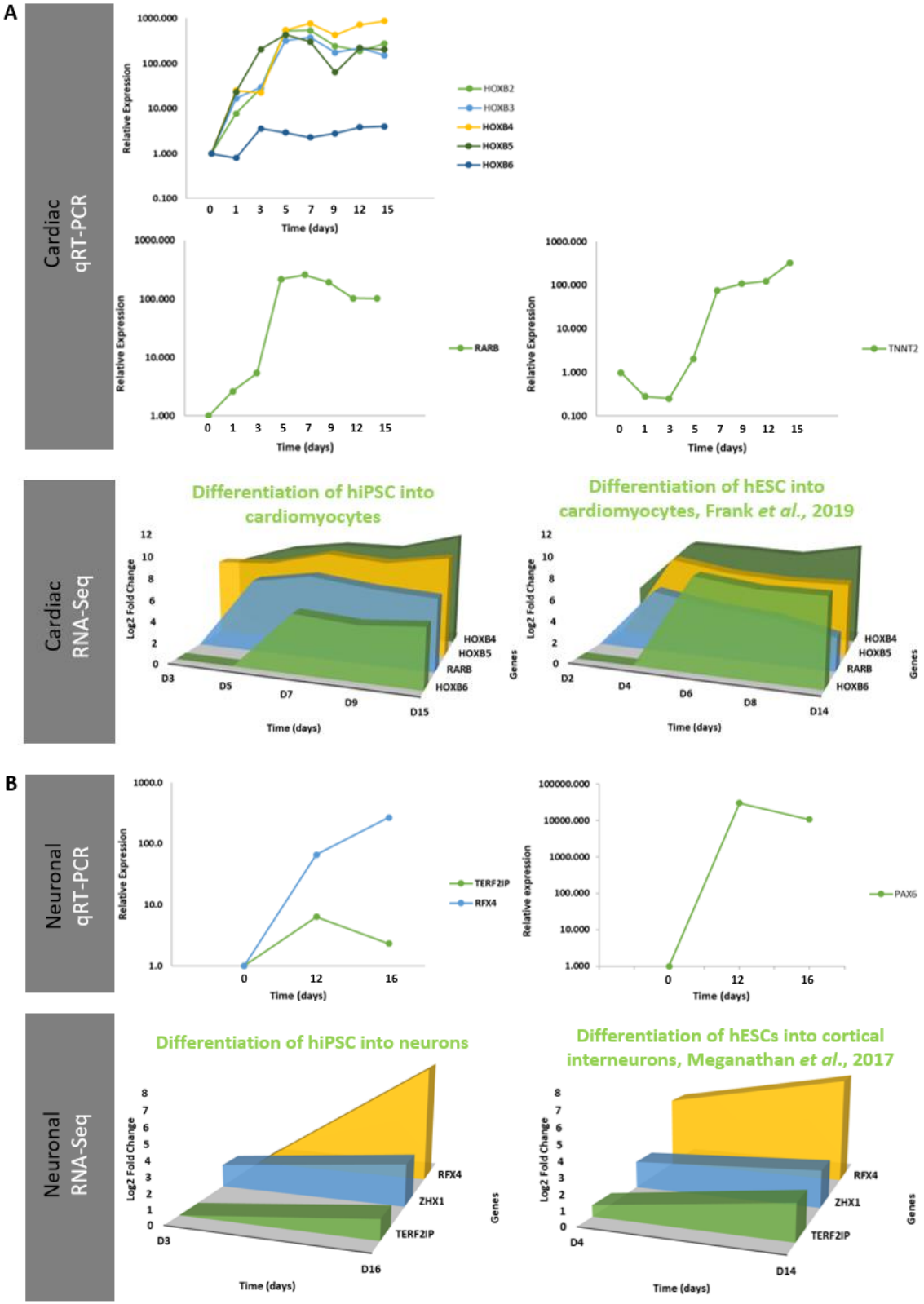


Fig. 4.10 Expression profiles during hiPSC cardiomyocyte and neuronal differentiation.
(A) Gene expression was assessed by qRT-PCR and RNA-Seq at sequential stages for putative novel regulators of cardiomyocyte differentiation identified through the transcriptomic analysis (*HOXB4, HOXB6* and *RARB*). It is also displayed the expression of *HOXB5,* a gene that has been predicted by STRING database as an interactive protein, representing another putative novel regulator. The expression of *HOXB2* and *HOXB3,* as well as of key genes that unveil the progression throughout cardiac differentiation (*TNNT2)* was also assessed by qRT-PCR. (B) Gene expression was assessed by qRT-PCR and RNA-Seq at sequential stages for putative novel regulators of neuronal differentiation identified through the transcriptomic analysis (*TERF2IP, RFX4* and *ZHX1)*. The expression of key genes that unveil

the progression throughout neuronal differentiation was also assessed by qRT-PCR, with *PAX6*. The analysis was performed using the $^{\Delta\Delta}$Ct, therefore expression is always represented as relative to day 0 and normalized against the expression of the housekeeping gene GAPDH. n=1 qRT-PCR experiment for each time point. For RNA-Seq graphs, X axis shows the differentiation time points, y axis shows *log*$_2$ (fold-change), and z axis shows genes encoding transcription factors. Novel regulators predicted in our transcriptomic analysis are highlighted in bold.

5 Conclusion and Future Perspectives

Here, we developed a comprehensive transcriptomic and network modeling approach to investigate the critical changes that may occur during human early neuronal, cardiac and hepatic differentiation. Additional published datasets were analysed to cover multiple cellular lineages and construct their developmental trajectories.

We identified temporal gene and pathway activation patterns during controlled differentiation of hPSC. These approaches paved the way for a deeper understanding of the regulatory mechanisms governing distinct cellular subpopulations present at each differentiation stage. Our analysis led us also to identify novel putative regulators of cardiogenesis and neurogenesis. We expect that this knowledge might one day be exploited for regenerative medicine, disease modeling and drug discovery applications.

Furthermore, we anticipate our assay to be a starting point to better understand the mechanisms that control human pluripotent stem cell fate transitions, and we hope that this will remain a highly dynamic field in the future. Indeed, we propose below some main strategies with important contributions, that should be furtherly developed.

Therefore, putative novel key regulators of cardiogenesis and neurogenesis predicted in our transcriptomic analysis should be subsequently validated by small interfering RNA (siRNA). This strategy would be required to undoubtedly confirm that they play a crucial role in cardiac and neuronal differentiation, whenever their targeted knockdown prevents differentiation.

Moreover, hiPSC-derived hepatocytes should also be analysed by RNA-Seq in order to perform a broader analysis of human early development, integrating cell lineages derived from all three human germ layers. This would therefore provide essential additional insights about differentially regulated mechanisms that govern each lineage decision-making.

In addition, the same type of study that we developed in a 2D culture system could be employed to a 3D system, in order to unveil the similarities and different features of both culture systems. Cells cultured in 3D mimic the *in vivo* microenvironment more closely. However, they can present higher cellular heterogeneity/plasticity and different morphology, in comparison with conventional 2D systems, due to differences among gradients of oxygen, growth factors, nutrients and complex cell-matrix or cell-cell interactions. Thus, it would be beneficial to perform a single-cell RNA-Seq in these 3D cell culture systems, taking into consideration single-cell measurements of the heterogeneous transcriptional responses. Although this technique was not required in our differentiated cells in a 2D culture system, since our protocols derived highly homogenous cell populations, it would be more adequate for a 3D culture comparative analysis.

Hopefully, the present study opens a wide range of exciting possibilities that would aid in enhancing the efficiency of hPSC differentiation as well as their clinical applications. Overall, our analysis reveals the cellular-state landscape of hPSC early differentiation, offering new insights in terms of gene expression profiles and signaling pathways crosstalk, that will shed light on novel molecular mechanisms governing cell fate decisions.

6 References

1 Dulak, J., Szade, K., Szade, A., Nowak, W. & Jozkowicz, A. Adult stem cells: hopes and hypes of regenerative medicine. *Acta biochimica Polonica* **62**, 329-337, doi:10.18388/abp.2015_1023 (2015).

2 Daley, G. Q. Stem cells and the evolving notion of cellular identity. *Philosophical transactions of the Royal Society of London. Series B, Biological sciences* **370**, 20140376, doi:10.1098/rstb.2014.0376 (2015).

3 Sobhani, A. *et al.* Multipotent Stem Cell and Current Application. *Acta medica Iranica* **55**, 6-23 (2017).

4 Yang, Y. *et al.* Derivation of Pluripotent Stem Cells with In Vivo Embryonic and Extraembryonic Potency. *Cell* **169**, 243-257 e225, doi:10.1016/j.cell.2017.02.005 (2017).

5 Theunissen, T. W. & Jaenisch, R. Mechanisms of gene regulation in human embryos and pluripotent stem cells. *Development* **144**, 4496-4509, doi:10.1242/dev.157404 (2017).

6 Takahashi, K. & Yamanaka, S. Induction of pluripotent stem cells from mouse embryonic and adult fibroblast cultures by defined factors. *Cell* **126**, 663-676, doi:10.1016/j.cell.2006.07.024 (2006).

7 Robinton, D. A. & Daley, G. Q. The promise of induced pluripotent stem cells in research and therapy. *Nature* **481**, 295-305, doi:10.1038/nature10761 (2012).

8 De Los Angeles, A. *et al.* Hallmarks of pluripotency. *Nature* **525**, 469-478, doi:10.1038/nature15515 (2015).

9 Crane, G. M., Jeffery, E. & Morrison, S. J. Adult haematopoietic stem cell niches. *Nature reviews. Immunology* **17**, 573-590, doi:10.1038/nri.2017.53 (2017).

10 Angelos, M. G. & Kaufman, D. S. Pluripotent stem cell applications for regenerative medicine. *Current opinion in organ transplantation* **20**, 663-670, doi:10.1097/MOT.0000000000000244 (2015).

11 Trounson, A. & DeWitt, N. D. Pluripotent stem cells progressing to the clinic. *Nature reviews. Molecular cell biology* **17**, 194-200, doi:10.1038/nrm.2016.10 (2016).

12 Shi, Y., Inoue, H., Wu, J. C. & Yamanaka, S. Induced pluripotent stem cell technology: a decade of progress. *Nature reviews. Drug discovery* **16**, 115-130, doi:10.1038/nrd.2016.245 (2017).

13 Passier, R., Orlova, V. & Mummery, C. Complex Tissue and Disease Modeling using hiPSCs. *Cell stem cell* **18**, 309-321, doi:10.1016/j.stem.2016.02.011 (2016).

14 Dutta, D., Heo, I. & Clevers, H. Disease Modeling in Stem Cell-Derived 3D Organoid Systems. *Trends in molecular medicine* **23**, 393-410, doi:10.1016/j.molmed.2017.02.007 (2017).

15 Rami, F. *et al.* Recent Advances in Therapeutic Applications of Induced Pluripotent Stem Cells. *Cellular reprogramming* **19**, 65-74, doi:10.1089/cell.2016.0034 (2017).

16 Udi Sarig & Marcelle Machluf. Engineering cell platforms for myocardial regeneration, Expert Opinion on Biological Therapy, 11:8, 1055-1077, DOI: 10.1517/14712598.2011.578574. (2011).

17 Stoltz, J. F. *et al.* Stem cells and applications: a survey. *Bio-medical materials and engineering* **25**, 3-26, doi:10.3233/BME-141225 (2015).

18 Dabrowska, A. M. & Skopinski, P. Stem cells in regenerative medicine - from laboratory to clinical application - the eye. *Central-European journal of immunology* **42**, 173-180, doi:10.5114/ceji.2017.69360 (2017).

19 Odorico, J. S., Kaufman, D. S. & Thomson, J. A. Multilineage differentiation from human embryonic stem cell lines. *Stem cells* **19**, 193-204, doi:10.1634/stemcells.19-3-193 (2001).

20 Niakan, K. K., Han, J., Pedersen, R. A., Simon, C. & Pera, R. A. Human pre-implantation embryo development. *Development* **139**, 829-841, doi:10.1242/dev.060426 (2012).

21 Freour, T. & Vassena, R. Transcriptomics analysis and human preimplantation development. *Journal of proteomics* **162**, 135-140, doi:10.1016/j.jprot.2016.10.004 (2017).

22 Gardner, D. K. & Balaban, B. Assessment of human embryo development using morphological criteria in an era of time-lapse, algorithms and 'OMICS': is looking good still important? *Molecular human reproduction* **22**, 704-718, doi:10.1093/molehr/gaw057 (2016).

23 Blakeley, P. *et al.* Defining the three cell lineages of the human blastocyst by single-cell RNA-seq. *Development* **142**, 3613, doi:10.1242/dev.131235 (2015).

24 Nishioka, N. *et al.* The Hippo signaling pathway components Lats and Yap pattern Tead4 activity to distinguish mouse trophectoderm from inner cell mass. *Developmental cell* **16**, 398-410, doi:10.1016/j.devcel.2009.02.003 (2009).

25 Yamanaka, Y., Lanner, F. & Rossant, J. FGF signal-dependent segregation of primitive endoderm and epiblast in the mouse blastocyst. *Development* **137**, 715-724, doi:10.1242/dev.043471 (2010).
26 Strumpf, D. *et al.* Cdx2 is required for correct cell fate specification and differentiation of trophectoderm in the mouse blastocyst. *Development* **132**, 2093-2102, doi:10.1242/dev.01801 (2005).
27 Piliszek, A., Grabarek, J. B., Frankenberg, S. R. & Plusa, B. Cell fate in animal and human blastocysts and the determination of viability. *Molecular human reproduction* **22**, 681-690, doi:10.1093/molehr/gaw002 (2016).
28 Simunovic, M. & Brivanlou, A. H. Embryoids, organoids and gastruloids: new approaches to understanding embryogenesis. *Development* **144**, 976-985, doi:10.1242/dev.143529 (2017).
29 Sleeman, J. P. & Thiery, J. P. SnapShot: The epithelial-mesenchymal transition. *Cell* **145**, 162 e161, doi:10.1016/j.cell.2011.03.029 (2011).
30 Denker, H. W. Early human development: new data raise important embryological and ethical questions relevant for stem cell research. *Die Naturwissenschaften* **91**, 1-21, doi:10.1007/s00114-003-0490-8 (2004).
31 Sun, X., Meyers, E. N., Lewandoski, M. & Martin, G. R. Targeted disruption of Fgf8 causes failure of cell migration in the gastrulating mouse embryo. *Genes & development* **13**, 1834-1846 (1999).
32 Beddington, R. S. & Robertson, E. J. Axis development and early asymmetry in mammals. *Cell* **96**, 195-209 (1999).
33 Perea-Gomez, A., Rhinn, M. & Ang, S. L. Role of the anterior visceral endoderm in restricting posterior signals in the mouse embryo. *The International journal of developmental biology* **45**, 311-320 (2001).
34 Ozair, M. Z., Kintner, C. & Brivanlou, A. H. Neural induction and early patterning in vertebrates. *Wiley interdisciplinary reviews. Developmental biology* **2**, 479-498, doi:10.1002/wdev.90 (2013).
35 Vieira, C. *et al.* Molecular mechanisms controlling brain development: an overview of neuroepithelial secondary organizers. *The International journal of developmental biology* **54**, 7-20, doi:10.1387/ijdb.092853cv (2010).
36 Simoes-Costa, M. & Bronner, M. E. Establishing neural crest identity: a gene regulatory recipe. *Development* **142**, 242-257, doi:10.1242/dev.105445 (2015).
37 Rex, M. *et al.* Dynamic expression of chicken Sox2 and Sox3 genes in ectoderm induced to form neural tissue. *Developmental dynamics : an official publication of the American Association of Anatomists* **209**, 323-332, doi:10.1002/(SICI)1097-0177(199707)209:3<323::AID-AJA7>3.0.CO;2-K (1997).
38 Colas, J. F. & Schoenwolf, G. C. Towards a cellular and molecular understanding of neurulation. *Developmental dynamics : an official publication of the American Association of Anatomists* **221**, 117-145, doi:10.1002/dvdy.1144 (2001).
39 Le Dreau, G. & Marti, E. Dorsal-ventral patterning of the neural tube: a tale of three signals. *Developmental neurobiology* **72**, 1471-1481, doi:10.1002/dneu.22015 (2012).
40 Catela, C., Shin, M. M. & Dasen, J. S. Assembly and function of spinal circuits for motor control. *Annual review of cell and developmental biology* **31**, 669-698, doi:10.1146/annurev-cellbio-100814-125155 (2015).
41 Verrier, L., Davidson, L., Gierlinski, M., Dady, A. & Storey, K. G. Neural differentiation, selection and transcriptomic profiling of human neuromesodermal progenitor-like cells in vitro. *Development* **145**, doi:10.1242/dev.166215 (2018).
42 Kriegstein, A. & Alvarez-Buylla, A. The glial nature of embryonic and adult neural stem cells. *Annual review of neuroscience* **32**, 149-184, doi:10.1146/annurev.neuro.051508.135600 (2009).
43 Dulken, B. W., Leeman, D. S., Boutet, S. C., Hebestreit, K. & Brunet, A. Single-Cell Transcriptomic Analysis Defines Heterogeneity and Transcriptional Dynamics in the Adult Neural Stem Cell Lineage. *Cell reports* **18**, 777-790, doi:10.1016/j.celrep.2016.12.060 (2017).
44 Puelles, L., Harrison, M., Paxinos, G. & Watson, C. A developmental ontology for the mammalian brain based on the prosomeric model. *Trends in neurosciences* **36**, 570-578, doi:10.1016/j.tins.2013.06.004 (2013).
45 Tam, P. P. & Beddington, R. S. The formation of mesodermal tissues in the mouse embryo during gastrulation and early organogenesis. *Development* **99**, 109-126 (1987).
46 Loh, K. M. *et al.* Mapping the Pairwise Choices Leading from Pluripotency to Human Bone, Heart, and Other Mesoderm Cell Types. *Cell* **166**, 451-467, doi:10.1016/j.cell.2016.06.011 (2016).

47 de Bree, K., de Bakker, B. S. & Oostra, R. J. The development of the human notochord. *PloS one* **13**, e0205752, doi:10.1371/journal.pone.0205752 (2018).
48 Ward, L., Pang, A. S. W., Evans, S. E. & Stern, C. D. The role of the notochord in amniote vertebral column segmentation. *Developmental biology* **439**, 3-18, doi:10.1016/j.ydbio.2018.04.005 (2018).
49 Tzahor, E. Head muscle development. *Results and problems in cell differentiation* **56**, 123-142, doi:10.1007/978-3-662-44608-9_6 (2015).
50 Ranghini, E. J. & Dressler, G. R. Evidence for intermediate mesoderm and kidney progenitor cell specification by Pax2 and PTIP dependent mechanisms. *Developmental biology* **399**, 296-305, doi:10.1016/j.ydbio.2015.01.005 (2015).
51 Tanaka, M. Molecular and evolutionary basis of limb field specification and limb initiation. *Development, growth & differentiation* **55**, 149-163, doi:10.1111/dgd.12017 (2013).
52 Lee, J. H., Protze, S. I., Laksman, Z., Backx, P. H. & Keller, G. M. Human Pluripotent Stem Cell-Derived Atrial and Ventricular Cardiomyocytes Develop from Distinct Mesoderm Populations. *Cell stem cell* **21**, 179-194 e174, doi:10.1016/j.stem.2017.07.003 (2017).
53 Bulatovic, I., Mansson-Broberg, A., Sylven, C. & Grinnemo, K. H. Human fetal cardiac progenitors: The role of stem cells and progenitors in the fetal and adult heart. *Best practice & research. Clinical obstetrics & gynaecology* **31**, 58-68, doi:10.1016/j.bpobgyn.2015.08.008 (2016).
54 Ruiz-Villalba, A., Hoppler, S. & van den Hoff, M. J. Wnt signaling in the heart fields: Variations on a common theme. *Developmental dynamics : an official publication of the American Association of Anatomists* **245**, 294-306, doi:10.1002/dvdy.24372 (2016).
55 Ema, M., Takahashi, S. & Rossant, J. Deletion of the selection cassette, but not cis-acting elements, in targeted Flk1-lacZ allele reveals Flk1 expression in multipotent mesodermal progenitors. *Blood* **107**, 111-117, doi:10.1182/blood-2005-05-1970 (2006).
56 Moretti, A. *et al.* Multipotent embryonic isl1+ progenitor cells lead to cardiac, smooth muscle, and endothelial cell diversification. *Cell* **127**, 1151-1165, doi:10.1016/j.cell.2006.10.029 (2006).
57 Wu, S. M. *et al.* Developmental origin of a bipotential myocardial and smooth muscle cell precursor in the mammalian heart. *Cell* **127**, 1137-1150, doi:10.1016/j.cell.2006.10.028 (2006).
58 Spater, D., Hansson, E. M., Zangi, L. & Chien, K. R. How to make a cardiomyocyte. *Development* **141**, 4418-4431, doi:10.1242/dev.091538 (2014).
59 Paige, S. L., Plonowska, K., Xu, A. & Wu, S. M. Molecular regulation of cardiomyocyte differentiation. *Circulation research* **116**, 341-353, doi:10.1161/CIRCRESAHA.116.302752 (2015).
60 Brown, D. D. *et al.* Tbx5 and Tbx20 act synergistically to control vertebrate heart morphogenesis. *Development* **132**, 553-563, doi:10.1242/dev.01596 (2005).
61 Calderon, D., Bardot, E. & Dubois, N. Probing early heart development to instruct stem cell differentiation strategies. *Developmental dynamics : an official publication of the American Association of Anatomists* **245**, 1130-1144, doi:10.1002/dvdy.24441 (2016).
62 Pinto, A. R. *et al.* Revisiting Cardiac Cellular Composition. *Circulation research* **118**, 400-409, doi:10.1161/CIRCRESAHA.115.307778 (2016).
63 Lewis, S. L. & Tam, P. P. Definitive endoderm of the mouse embryo: formation, cell fates, and morphogenetic function. *Developmental dynamics : an official publication of the American Association of Anatomists* **235**, 2315-2329, doi:10.1002/dvdy.20846 (2006).
64 Rodriguez, A. M. & Downs, K. M. Visceral endoderm and the primitive streak interact to build the fetal-placental interface of the mouse gastrula. *Developmental biology* **432**, 98-124, doi:10.1016/j.ydbio.2017.08.026 (2017).
65 Gordillo, M., Evans, T. & Gouon-Evans, V. Orchestrating liver development. *Development* **142**, 2094-2108, doi:10.1242/dev.114215 (2015).
66 Zorn, A. M. & Wells, J. M. Vertebrate endoderm development and organ formation. *Annual review of cell and developmental biology* **25**, 221-251, doi:10.1146/annurev.cellbio.042308.113344 (2009).
67 Bort, R., Signore, M., Tremblay, K., Martinez Barbera, J. P. & Zaret, K. S. Hex homeobox gene controls the transition of the endoderm to a pseudostratified, cell emergent epithelium for liver bud development. *Developmental biology* **290**, 44-56, doi:10.1016/j.ydbio.2005.11.006 (2006).
68 Goldman, O. *et al.* Endoderm generates endothelial cells during liver development. *Stem cell reports* **3**, 556-565, doi:10.1016/j.stemcr.2014.08.009 (2014).

69 Semrau, S. & van Oudenaarden, A. Studying lineage decision-making in vitro: emerging concepts and novel tools. *Annual review of cell and developmental biology* **31**, 317-345, doi:10.1146/annurev-cellbio-100814-125300 (2015).
70 Murry, C. E. & Keller, G. Differentiation of embryonic stem cells to clinically relevant populations: lessons from embryonic development. *Cell* **132**, 661-680, doi:10.1016/j.cell.2008.02.008 (2008).
71 Rao, J. & Greber, B. Concise Review: Signaling Control of Early Fate Decisions Around the Human Pluripotent Stem Cell State. *Stem cells* **35**, 277-283, doi:10.1002/stem.2527 (2017).
72 Watabe, T. & Miyazono, K. Roles of TGF-beta family signaling in stem cell renewal and differentiation. *Cell research* **19**, 103-115, doi:10.1038/cr.2008.323 (2009).
73 Park, C. *et al.* A hierarchical order of factors in the generation of FLK1- and SCL-expressing hematopoietic and endothelial progenitors from embryonic stem cells. *Development* **131**, 2749-2762, doi:10.1242/dev.01130 (2004).
74 Lindsley, R. C., Gill, J. G., Kyba, M., Murphy, T. L. & Murphy, K. M. Canonical Wnt signaling is required for development of embryonic stem cell-derived mesoderm. *Development* **133**, 3787-3796, doi:10.1242/dev.02551 (2006).
75 Xu, Z. *et al.* Wnt/beta-catenin signaling promotes self-renewal and inhibits the primed state transition in naive human embryonic stem cells. *Proceedings of the National Academy of Sciences of the United States of America* **113**, E6382-E6390, doi:10.1073/pnas.1613849113 (2016).
76 Kubo, A. *et al.* Development of definitive endoderm from embryonic stem cells in culture. *Development* **131**, 1651-1662, doi:10.1242/dev.01044 (2004).
77 Chambers, S. M. *et al.* Highly efficient neural conversion of human ES and iPS cells by dual inhibition of SMAD signaling. *Nature biotechnology* **27**, 275-280, doi:10.1038/nbt.1529 (2009).
78 Pera, M. F. & Tam, P. P. Extrinsic regulation of pluripotent stem cells. *Nature* **465**, 713-720, doi:10.1038/nature09228 (2010).
79 Wnt signaling and a Smad pathway blockade direct the differentiation of human pluripotent stem cells to multipotent neural crest cells. *Proceedings of the National Academy of Sciences* **109**, 9220-9220, doi:10.1073/pnas.1207810109 (2012).
80 Lian, X., Zhang, J., Zhu, K., Kamp, T. J. & Palecek, S. P. Insulin inhibits cardiac mesoderm, not mesendoderm, formation during cardiac differentiation of human pluripotent stem cells and modulation of canonical Wnt signaling can rescue this inhibition. *Stem cells* **31**, 447-457, doi:10.1002/stem.1289 (2013).
81 Yan, Y. *et al.* Tbx1 modulates endodermal and mesodermal differentiation from mouse induced pluripotent stem cells. *Stem cells and development* **23**, 1491-1500, doi:10.1089/scd.2013.0488 (2014).
82 Teo, A. K. *et al.* Activin and BMP4 synergistically promote formation of definitive endoderm in human embryonic stem cells. *Stem cells* **30**, 631-642, doi:10.1002/stem.1022 (2012).
83 McLean, A. B. *et al.* Activin a efficiently specifies definitive endoderm from human embryonic stem cells only when phosphatidylinositol 3-kinase signaling is suppressed. *Stem cells* **25**, 29-38, doi:10.1634/stemcells.2006-0219 (2007).
84 Cheung, C., Bernardo, A. S., Trotter, M. W., Pedersen, R. A. & Sinha, S. Generation of human vascular smooth muscle subtypes provides insight into embryological origin-dependent disease susceptibility. *Nature biotechnology* **30**, 165-173, doi:10.1038/nbt.2107 (2012).
85 Lian, X. *et al.* Directed cardiomyocyte differentiation from human pluripotent stem cells by modulating Wnt/beta-catenin signaling under fully defined conditions. *Nature protocols* **8**, 162-175, doi:10.1038/nprot.2012.150 (2013).
86 Gessert, S. & Kuhl, M. The multiple phases and faces of wnt signaling during cardiac differentiation and development. *Circulation research* **107**, 186-199, doi:10.1161/CIRCRESAHA.110.221531 (2010).
87 Wang, Z., Gerstein, M. & Snyder, M. RNA-Seq: a revolutionary tool for transcriptomics. *Nature reviews. Genetics* **10**, 57-63, doi:10.1038/nrg2484 (2009).
88 Costa-Silva, J., Domingues, D. & Lopes, F. M. RNA-Seq differential expression analysis: An extended review and a software tool. *PloS one* **12**, e0190152, doi:10.1371/journal.pone.0190152 (2017).
89 Marioni, J. C., Mason, C. E., Mane, S. M., Stephens, M. & Gilad, Y. RNA-seq: an assessment of technical reproducibility and comparison with gene expression arrays. *Genome research* **18**, 1509-1517, doi:10.1101/gr.079558.108 (2008).

90 Cloonan, N. *et al.* Stem cell transcriptome profiling via massive-scale mRNA sequencing. *Nature methods* **5**, 613-619, doi:10.1038/nmeth.1223 (2008).
91 Barbazuk, W. B., Emrich, S. J., Chen, H. D., Li, L. & Schnable, P. S. SNP discovery via 454 transcriptome sequencing. *The Plant journal : for cell and molecular biology* **51**, 910-918, doi:10.1111/j.1365-313X.2007.03193.x (2007).
92 Petropoulos, S., Panula, S. P., Schell, J. P. & Lanner, F. Single-cell RNA sequencing: revealing human pre-implantation development, pluripotency and germline development. *Journal of internal medicine* **280**, 252-264, doi:10.1111/joim.12493 (2016).
93 Meganathan, K. *et al.* Regulatory networks specifying cortical interneurons from human embryonic stem cells reveal roles for CHD2 in interneuron development. *Proceedings of the National Academy of Sciences of the United States of America* **114**, E11180-E11189, doi:10.1073/pnas.1712365115 (2017).
94 Frank, S. *et al.* yylncT Defines a Class of Divergently Transcribed lncRNAs and Safeguards the T-mediated Mesodermal Commitment of Human PSCs. *Cell stem cell* **24**, 318-327 e318, doi:10.1016/j.stem.2018.11.005 (2019).
95 Li, Q. *et al.* A sequential EMT-MET mechanism drives the differentiation of human embryonic stem cells towards hepatocytes. *Nature communications* **8**, 15166, doi:10.1038/ncomms15166 (2017).
96 Chen, G. *et al.* Chemically defined conditions for human iPSC derivation and culture. *Nature methods* **8**, 424-429, doi:10.1038/nmeth.1593 (2011).
97 Gomes, A. R. Engineering region-specific brain organoids from human induced pluripotent stem (hiPS) cells for disease modeling of Rett syndrome. *PhD in Bioengineering - Cell Therapies and Regenerative Medicine* (2017-2020).
98 Nicholas, C. R. *et al.* Functional maturation of hPSC-derived forebrain interneurons requires an extended timeline and mimics human neural development. *Cell stem cell* **12**, 573-586, doi:10.1016/j.stem.2013.04.005 (2013).
99 Maroof, A. M. *et al.* Directed differentiation and functional maturation of cortical interneurons from human embryonic stem cells. *Cell stem cell* **12**, 559-572, doi:10.1016/j.stem.2013.04.008 (2013).
100 Branco, M. Development of Human Pluripotent Stem Cell - Derived Cardiovascular Organoid Microtissues for Cardiotoxicity Screening. *PhD dissertation in Biological Engineering, MIT Portugal* (2017-2020).
101 Rao, J. *et al.* Stepwise Clearance of Repressive Roadblocks Drives Cardiac Induction in Human ESCs. *Cell stem cell* **18**, 554-556, doi:10.1016/j.stem.2016.03.008 (2016).
102 Waldmann, T. *et al.* Design principles of concentration-dependent transcriptome deviations in drug-exposed differentiating stem cells. *Chemical research in toxicology* **27**, 408-420, doi:10.1021/tx400402j (2014).
103 Metsalu, T. & Vilo, J. ClustVis: a web tool for visualizing clustering of multivariate data using Principal Component Analysis and heatmap. *Nucleic acids research* **43**, W566-570, doi:10.1093/nar/gkv468 (2015).
104 Mi, H., Muruganujan, A., Casagrande, J. T. & Thomas, P. D. Large-scale gene function analysis with the PANTHER classification system. *Nature protocols* **8**, 1551-1566, doi:10.1038/nprot.2013.092 (2013).
105 Semrau, S. *et al.* Dynamics of lineage commitment revealed by single-cell transcriptomics of differentiating embryonic stem cells. *Nature communications* **8**, 1096, doi:10.1038/s41467-017-01076-4 (2017).
106 Gupta, M. K. *et al.* Global transcriptional profiles of beating clusters derived from human induced pluripotent stem cells and embryonic stem cells are highly similar. *BMC developmental biology* **10**, 98, doi:10.1186/1471-213X-10-98 (2010).
107 Shinozawa, T. *et al.* Gene expression profiling of functional murine embryonic stem cell-derived cardiomyocytes and comparison with adult heart: profiling of murine ESC-derived cardiomyocytes. *Journal of biomolecular screening* **14**, 239-245, doi:10.1177/1087057108330112 (2009).
108 Handel, A. E. *et al.* Assessing similarity to primary tissue and cortical layer identity in induced pluripotent stem cell-derived cortical neurons through single-cell transcriptomics. *Human molecular genetics* **25**, 989-1000, doi:10.1093/hmg/ddv637 (2016).

109 Almalki, S. G. & Agrawal, D. K. Key transcription factors in the differentiation of mesenchymal stem cells. *Differentiation; research in biological diversity* **92**, 41-51, doi:10.1016/j.diff.2016.02.005 (2016).

110 Liang, Q. *et al.* The roles of Mesp family proteins: functional diversity and redundancy in differentiation of pluripotent stem cells and mammalian mesodermal development. *Protein & cell* **6**, 553-561, doi:10.1007/s13238-015-0176-y (2015).

111 Morrow, Z. T. *et al.* tbx6l and tbx16 are redundantly required for posterior paraxial mesoderm formation during zebrafish embryogenesis. *Developmental dynamics : an official publication of the American Association of Anatomists* **246**, 759-769, doi:10.1002/dvdy.24547 (2017).

112 Faial, T. *et al.* Brachyury and SMAD signalling collaboratively orchestrate distinct mesoderm and endoderm gene regulatory networks in differentiating human embryonic stem cells. *Development* **142**, 2121-2135, doi:10.1242/dev.117838 (2015).

113 Bruneau, B. G. Signaling and transcriptional networks in heart development and regeneration. *Cold Spring Harbor perspectives in biology* **5**, a008292, doi:10.1101/cshperspect.a008292 (2013).

114 Ma, Q., Zhou, B. & Pu, W. T. Reassessment of Isl1 and Nkx2-5 cardiac fate maps using a Gata4-based reporter of Cre activity. *Developmental biology* **323**, 98-104, doi:10.1016/j.ydbio.2008.08.013 (2008).

115 Pandur, P., Sirbu, I. O., Kuhl, S. J., Philipp, M. & Kuhl, M. Islet1-expressing cardiac progenitor cells: a comparison across species. *Development genes and evolution* **223**, 117-129, doi:10.1007/s00427-012-0400-1 (2013).

116 Suter, D. M., Tirefort, D., Julien, S. & Krause, K. H. A Sox1 to Pax6 switch drives neuroectoderm to radial glia progression during differentiation of mouse embryonic stem cells. *Stem cells* **27**, 49-58, doi:10.1634/stemcells.2008-0319 (2009).

117 Watson, L. M., Wong, M. M. K., Vowles, J., Cowley, S. A. & Becker, E. B. E. A Simplified Method for Generating Purkinje Cells from Human-Induced Pluripotent Stem Cells. *Cerebellum* **17**, 419-427, doi:10.1007/s12311-017-0913-2 (2018).

118 Wackerage, H. R., D. P .D.; Judson, R. N.; Sudol, M.; Sadoshima, J. The Hippo signal transduction network in skeletal and cardiac muscle. *Science signaling* **7**, doi:DOI: 10.1126/scisignal.2005096 (2014).

119 Zhou, Q., Li, L., Zhao, B. & Guan, K. L. The hippo pathway in heart development, regeneration, and diseases. *Circulation research* **116**, 1431-1447, doi:10.1161/CIRCRESAHA.116.303311 (2015).

120 Sumi, T., Tsuneyoshi, N., Nakatsuji, N. & Suemori, H. Defining early lineage specification of human embryonic stem cells by the orchestrated balance of canonical Wnt/beta-catenin, Activin/Nodal and BMP signaling. *Development* **135**, 2969-2979, doi:10.1242/dev.021121 (2008).

121 Lian, X. *et al.* Robust cardiomyocyte differentiation from human pluripotent stem cells via temporal modulation of canonical Wnt signaling. *Proceedings of the National Academy of Sciences of the United States of America* **109**, E1848-1857, doi:10.1073/pnas.1200250109 (2012).

122 Sheng, Z. *et al.* Cardiotrophin 1 (CT-1) inhibition of cardiac myocyte apoptosis via a mitogen-activated protein kinase-dependent pathway. Divergence from downstream CT-1 signals for myocardial cell hypertrophy. *The Journal of biological chemistry* **272**, 5783-5791, doi:10.1074/jbc.272.9.5783 (1997).

123 Chandrasekaran, A. *et al.* Comparison of 2D and 3D neural induction methods for the generation of neural progenitor cells from human induced pluripotent stem cells. *Stem cell research* **25**, 139-151, doi:10.1016/j.scr.2017.10.010 (2017).

124 Hebert, J. M. & Fishell, G. The genetics of early telencephalon patterning: some assembly required. *Nature reviews. Neuroscience* **9**, 678-685, doi:10.1038/nrn2463 (2008).

125 Corbin, J. G., Rutlin, M., Gaiano, N. & Fishell, G. Combinatorial function of the homeodomain proteins Nkx2.1 and Gsh2 in ventral telencephalic patterning. *Development* **130**, 4895-4906, doi:10.1242/dev.00717 (2003).

126 Yan, Y., Song, L., Madinya, J., Ma, T. & Li, Y. Derivation of Cortical Spheroids from Human Induced Pluripotent Stem Cells in a Suspension Bioreactor. *Tissue engineering. Part A* **24**, 418-431, doi:10.1089/ten.TEA.2016.0400 (2018).

127 Li, P. *et al.* Transcriptional Reactivation of OTX2, RX1 and SIX3 during Reprogramming Contributes to the Generation of RPE Cells from Human iPSCs. *International journal of biological sciences* **12**, 505-517, doi:10.7150/ijbs.14212 (2016).

128 DeRosa, B. A. *et al.* hVGAT-mCherry: A novel molecular tool for analysis of GABAergic neurons derived from human pluripotent stem cells. *Molecular and cellular neurosciences* **68**, 244-257, doi:10.1016/j.mcn.2015.08.007 (2015).
129 Rajala, K., Pekkanen-Mattila, M. & Aalto-Setala, K. Cardiac differentiation of pluripotent stem cells. *Stem cells international* **2011**, 383709, doi:10.4061/2011/383709 (2011).
130 Costello, I. *et al.* The T-box transcription factor Eomesodermin acts upstream of Mesp1 to specify cardiac mesoderm during mouse gastrulation. *Nature cell biology* **13**, 1084-1091, doi:10.1038/ncb2304 (2011).
131 Rao, J. *et al.* Stepwise Clearance of Repressive Roadblocks Drives Cardiac Induction in Human ESCs. *Cell stem cell* **18**, 341-353, doi:10.1016/j.stem.2015.11.019 (2016).
132 Marshall, H. *et al.* A conserved retinoic acid response element required for early expression of the homeobox gene Hoxb-1. *Nature* **370**, 567-571, doi:10.1038/370567a0 (1994).
133 Nolte, C., Jinks, T., Wang, X., Martinez Pastor, M. T. & Krumlauf, R. Shadow enhancers flanking the HoxB cluster direct dynamic Hox expression in early heart and endoderm development. *Developmental biology* **383**, 158-173, doi:10.1016/j.ydbio.2013.09.016 (2013).
134 Paige, S. L. *et al.* A temporal chromatin signature in human embryonic stem cells identifies regulators of cardiac development. *Cell* **151**, 221-232, doi:10.1016/j.cell.2012.08.027 (2012).
135 den Hartogh, S. C., Wolstencroft, K., Mummery, C. L. & Passier, R. A comprehensive gene expression analysis at sequential stages of in vitro cardiac differentiation from isolated MESP1-expressing-mesoderm progenitors. *Scientific reports* **6**, 19386, doi:10.1038/srep19386 (2016).
136 Loo, Z. X., Kunasekaran, W., Govindasamy, V., Musa, S. & Abu Kasim, N. H. Comparative analysis of cardiovascular development related genes in stem cells isolated from deciduous pulp and adipose tissue. *TheScientificWorldJournal* **2014**, 186508, doi:10.1155/2014/186508 (2014).
137 Blackshear, P. J. *et al.* Graded phenotypic response to partial and complete deficiency of a brain-specific transcript variant of the winged helix transcription factor RFX4. *Development* **130**, 4539-4552, doi:10.1242/dev.00661 (2003).
138 Ashique, A. M. *et al.* The Rfx4 transcription factor modulates Shh signaling by regional control of ciliogenesis. *Science signaling* **2**, ra70, doi:10.1126/scisignal.2000602 (2009).
139 Sedykh, I. *et al.* Zebrafish Rfx4 controls dorsal and ventral midline formation in the neural tube. *Developmental dynamics : an official publication of the American Association of Anatomists* **247**, 650-659, doi:10.1002/dvdy.24613 (2018).
140 Kotla, S. *et al.* Endothelial senescence is induced by phosphorylation and nuclear export of telomeric repeat binding factor 2-interacting protein. *JCI insight* **4**, doi:10.1172/jci.insight.124867 (2019).

7 Supplementary Information

Supplementary Table 7.1 Literature-based curation list of ectoderm, mesoderm and endoderm-associated genes.

Ectoderm associated genes			Mesoderm associated genes							Endoderm associated genes					
EED	EDA2R	KRT6B	HOXA11	ETV2	ITGB4	ITGA2	NF2	POU5F1	TCF21	PAF1	CDC73	MESP1	HNF4A	ONECUT1	ZNF202
ENC1	ELF5	NF2	ITGA3	MEST	KDM6B	CITED2	POFUT2	EPHA2	FOXD3	LAMA3	COL4A2	SMAD3	APELA	NANOG	FGF4
IGF1	RBM46	ADA	MATK	LHX2	TBX3	TBXT	MESP2	MESTP4	HOXB5	COL11A1	ITGA7	CTNNB1	LHX1	SOX17	LHX1
CUL3	FZD7	CTR9	BTK	BMPR1A	GDF11	TBX20	TXNRD1	MESTP2	PHOX2A	WNT8A	LAMC1	ONECUT1	CD24	ARC	HOXC11
TP63	NODAL	SMURF1	FOXC1	FGF8	KLF4	SHH	BMPR2	MESTP3	RNF111	ITGB5	NKX2-1	SOX7	MTRR	MIR410	MIR1263
CNOT3	ETS2	VPS52	EYA2	DKK1	TRIM15	SOX17	MIR1-2	MESTP1	CDX4	MMP2	GATA4	BPTF	HNF1B	BMP7	PELO
ITGA6	GRHL3	LHX1	TIE1	WNT3	YAP1	MESP1	TSPY3	PUS7	PRDM14	KIF16B	RTF1	SMAD2	TGFB1	MED12	FOXL1
FOXF1	ZBTB7B	MTHFR	FGFR2	PRKAR1A	LEF1	SMAD3	TSPY8	CDX2	TBX18	LAMB1	DUSP5	FOXL1	MTHFR	ZEB2	GDF3
BMPR1A	EOMES	EZH2	NUP133	SRF	SECTM1	TSPY2	TSPY4	NPPA	PRRX2	MMP9	ITGAV	SOX2	AMN	HHEX	WNT7B
ZPR1	CITED2	KLHL25	PRKACA	VEGFA	SMAD4	ZFPM2	TSPY10	MTRR	TMEM88	MMP15	DUSP6	SETD2	LINC00261	PDX1	GSC
STX2	SHH	HDAC2	RPS6KA6	HAND1	OSR1	SMAD1	APELA	TET2	RNF130	TGFB1	GATA6	EXT1	CTNNB1	PAX9	CCNB3
CNOT2	CDX2	DNMT3B	TP63	PPP2CA	SFRP2	SIX2	TSPY1	NANOG	HCNR617	BMPR1A	SMAD4	NOG	NOTCH1	EXT1	MIXL1
BYSL	AMER2	CHUK	TEAD2	WNT5A	KDM6A	RPL38	ITGB3	GATA4	HHC2:066683	FGF8	COL6A1	GDF3	MTOR	PDGFA	
SRF	CTNNB1	MTRR	PAX2	ACVR2B	CER1	OVOL1	SCX	EDNRB	HHC2:066628	DKK1	COL8A1	MED12	CXCR4	FOXA2	
HAND1	ITGAM	YY1	FGFR1	EPB41L5	CRB2	SMAD2	WNT3	FGF1	LOC111258509	VTN	NOTCH1	MIXL1	SMAD3	NOG	
EPB41L5	JUNB	HDAC3	ITGA8	ACVR1	TBX6	CTDNEP1	PHOX2B	NKX2-5	LOC111258518	NANOG	HMGA2	ZFP36L1	HDAC1	GATA6	
VAX2	HOPX	CTCF	PALB2	TLX2	HMGA2	FOXL1	PRRX1	RNASE3	LOC111258517	COL12A1	PELO	LAMB3	SMAD4	MIR200A	
ZBTB17	SP3	ITGA4	WNT11	WLS	ITGB1	FOXC2	MIER1	MIR1-1	LOC111258516	COL7A1	HHEX	ARC	EZH2	NODAL	
AEBP2	EZH1	ANXA5	CHRD	MESD	EXT2	IRX3	MTHFR	WIF1	LOC100128248	EPB41L5	NODAL	CTR9	APC	LAMC1	
SUZ12	SCN5A	AGR2	DLL3	NR4A3	MSGN1	HES7	CTNNB1	TAL1		ITGA4	MMP14	PAX9	KDR	FERMT1	
FERMT1	HDAC1		JAK2	ACVR2A	POU4F1	DAND5	CDH1	FOXA2		FN1	DUSP2	COL5A2	FGF2	ITGB5	
WNT7A	DNMT1		BMP7	INHBA	WNT3A	SETD2	MTOR	TBX5		ID2	ITGB2	POU5F1	BMP4	KDM6B	
ITGB7	CHI3L1		HCK	SNAI1	NODAL	EXT1	CXCR4	FERMT1		MMP8	ITGA5	DKK1	SOX7	APELA	
ITGAE	BMP4		BMX	BMP4	ETS2	NOG	MIER3	MIR335		DUSP1	EOMES	SMAD2	KDM6A	CER1	
WNT3	DNMT3A		AXIN1	TCF15	FOXH1	TBX1	MIER2	SENP1		DUSP4	PITX2	NKX2-1	WNT2	EPB41L5	
CHD1L	PAX6		FOXF1	SMO	IKZF3	GDF3	TLNRD1	DLL1		INHBA	TBX20	HSBP1	EOMES	KIF16B	
PPP1R8	LOC266694		EYA1	TWSG1	POGLUT1	MIXL1	CARMN	FERMT2		BMP4	SOX17	AFP	BPTF	LEAP2	
TGS1	AGR3		AMH	EXOC4	EOMES	IKZF1	IGF2	HOPX		COL5A1	LEO1	ITGAV	TBX20	GSC-DT	

Supplementary Table 7.2 Gene clustering analysis.
Each cluster represents a protein-protein interaction network based on protein encoding genes involved in human ectodermal, mesodermal and endodermal specification. These interaction networks were obtained by STRING database, that predicts associations based on *in vivo* and *in vitro* experimental assays. Proteins were clustered using the Markov Cluster Algorithm (MLC) with an inflation parameter = 3. 10 more proteins were predicted for the network. This table represents the biological processes (FDR < 0.05) and KEGG pathways (FDR < 0.05) associated with the analysed proteins. (*) Protein encoding genes from individual cluster colours that did not present statistically significant results were grouped and analysed together.

	Cluster Color	Gene Count	Biological Process	Protein Name	Pathway Description
Ectoderm	Red	13	chromatin organization; regulation of transcription from RNA polymerase II promoter; ectodermal placode development; nervous system development.	DNMT3A; HDAC1; HDAC2; EZH2; EZH1; EED; AEBP2; DNMT3B; DNMT1; SP3; SUZ12; CTCF; YY1.	Notch signaling pathway
	Light green	12	developmental process; regulation of transcription; cell differentiation; ectoderm development; embryonic organ development; stem cell differentiation; axis specification. pattern specification process; forebrain regionalization; telencephalon regionalization .	CDX2; LHX1; SHH; TP63; STX2; EOMES; BMP4; NODAL; BMPR1A; PAX6; ELF5; FOXF1.	Hedgehog signaling pathway TGF-beta signaling pathway
	Cyan	7	integrin-mediated signaling pathway; ectoderm development; extracellular matrix organization; ectodermal cell differentiation; cell-matrix adhesion.	ITGAM; ITGA6; ZBTB17; ITGA4; ITGAE; ITGB7; FERMT1.	PI3K-Akt signaling pathway
	Orchid	5	trophectodermal cell differentiation; blastocyst formation; blastocyst development; primitive streak formation; gastrulation with mouth forming second; embryo development; anatomical structure formation involved in morphogenesis.	JUNB; HOPX; ETS2; FOS; SRF.	TNF signaling pathway
	Medium Purple	5	positive regulation of transcription elongation from RNA polymerase II promoter; histone H2B ubiquitination; gastrulation; stem cell differentiation.	CDC73; RTF1; LEO1; CTR9; PAF1.	-
	Pale Violet Red	3	positive regulation of cytoplasmic mRNA processing body assembly; nuclear-transcribed mRNA poly(A) tail shortening. negative regulation of intracellular estrogen receptor signaling pathway; regulation of stem cell population maintenance.	CNOT1; CNOT2; CNOT3.	-
	Olive	2	-	AGR3; AGR2.	-
	Medium Sea Green Purple Pink Green Yellow Brown Dark Golden Rod Yellow Green	9 4 4 2 2 2 2 2	regulation of translation; biosynthetic process. (*)	WNT7A; SCN10A; VPS53; CTNNB1; WNT3; HAND1; VPS52; PPP1R8; CHUK; CALM2; CUL3; NF2. AXIN1; SCN5A; ENC1; IGF1; FZD7; CHD1L; SMURF1; TGS1; ANXA5; HDAC3; KLHL25; MTHFR; ADA; ZBTB7B; MTRR.	Wnt signaling pathway TNF signaling pathway PI3K-Akt signaling pathway Hippo signaling pathway
Mesoderm	Red	33	mesoderm development; regionalization; gastrulation; pattern commitment process; embryonic morphogenesis; anterior/posterior pattern commitment; mesoderm morphogenesis; tissue morphogenesis; mesoderm formation formation of primary germ layer; negative regulation of developmental process; epithelium development; mesenchyme development; embryonic organ development; regulation of organ morphogenesis.	CDX2; WLS; TBX3; BMPR2; SHH; WNT5A; WNT11; WIF1; MESP1; CER1; EOMES; TWSG1; AMH; FGFR1; NOG; BMP4; SOX17; WNT3; GDF3; BMPR1A; FOXF1; BMP7; KLF4; WNT3A; DKK1; PAX2; HOXB5; MIXL1; SMO; FGF8; SFRP2; CHRD; FOXA2.	Wnt signaling pathway Hedgehog signaling pathway Hippo signaling pathway TGF-beta signaling pathway
	Aquamarine	17	mesoderm development; mesoderm formation; mesoderm morphogenesis; pattern commitment process; activin receptor signaling pathway; embryonic morphogenesis; formation of primary germ layer; gastrulation; transmembrane receptor protein serine/threonine kinase signaling pathway; tissue morphogenesis; skeletal system development; mesenchyme development; regionalization; cardiovascular system development.	SMAD1; INHBA; SMAD3; GDF11; TLX2; FOXH1; ACVR1; HMGA2; ACVR2A; NODAL; TAL1; SMAD2; POGLUT1; SMAD4; ACVR2B; TMEM88; SNAI1.	Hippo signaling pathway TGF-beta signaling pathway
	Medium Purple	12	viral gene expression; viral transcription; single-organism membrane organization; cellular component disassembly; nuclear-transcribed mRNA catabolic process, nonsense-mediated decay; mesoderm development; translational initiation, elongation and termination; protein transport.	RPLP2; PPP2CA; RPS20; NUP133; RPS27; NUP107; RPL30; RPL36; CTDNEP1; RPL38.	-
	Violet	11	mesenchyme development; cardiac septum morphogenesis and development; muscle structure development; heart morphogenesis; morphogenesis of an epithelium; cardiac chamber morphogenesis; heart development; cardiovascular system development; circulatory system development; striated muscle tissue development.	GATA4; HAND1; NPPA; TCF21; ZFPM2; TBX20; NKX2-5; KDM6A; TBX1; TBX5; TCF15.	-
	Purple	9	mesoderm development; blood vessel morphogenesis; positive regulation of cell division; cardiovascular system development; circulatory system development; positive regulation of MAPK cascade; cell surface receptor signaling pathway; anatomical structure morphogenesis; angiogenesis.	BMX; ETV2; CXCR4; APLNR; VEGFA; FGF1; FGFR2; IGF2; TIE1;	Ras signaling pathway Rap1 signaling pathway PI3K-Akt signaling pathway
	Hot Pink	8	somitogenesis; somite development; chordate embryonic development; paraxial mesoderm development.	TBX18; DLL3; HES7; MESP2; TBX6; FOXC2; FOXC1; DLL1.	Notch signaling pathway
	Pink	8	mesodermal cell differentiation; integrin-mediated signaling pathway; cell-matrix adhesion.	ITGA8; ITGA2; ITGB4; ITGA3; ITGB1; FERMT2; FERMT1; ITGB3.	PI3K-Akt signaling pathway
	Light Coral	5	mesoderm development; nephron development; kidney morphogenesis; tissue morphogenesis.	SIX2; IRX3; OSR1; EYA1; EYA2.	-
	Salmon	4	cellular response to heat; cell cycle arrest; regulation of transcription from RNA polymerase III promoter. cellular response to stress; regulation of intracellular signal transduction; cellular response to insulin stimulus.	MTOR; RPTOR; MLST8; RPS6KA6.	mTOR signaling pathway PI3K-Akt signaling pathway
	Olive	3	autonomic nervous system development; regulation of respiratory gaseous exchange; neural crest cell migration.	PHOX2A; PHOX2B; EDNRB.	-
	Yellow	3	exocytosis.	EXOC3; EXOC5; EXOC4.	-
	Brown	3	heparan sulfate proteoglycan biosynthetic process, polysaccharide chain biosynthetic process.	EXT2; EXT1; NR4A3.	-
	Sky Blue	6	mesoderm development; positive regulation of transcription, DNA-templated; tissue development; positive regulation of transcription from RNA polymerase II promoter; gene expression; embryo development.	RNF111; RNF130; MSGN1; CDX4; AXIN1; CTNNB1.	Hippo signaling pathway Wnt signaling pathway
	Medium Orchid Cornflower Blue Medium Slate Blue Pale Violet Red Fire Brick Sandy Brown Dark Golden Rod Blue Cyan Orchid Green	2 4 2 6 4 3 3 2 2 2 2	mesoderm formation, morphogenesis and development; embryo development; gastrulation. (*)	JAK2; TET2; LEF1; IKZF3; IKZF1; BTK; CITED2; LHX2; TEAD2; PRRX2; HOPX; PRRX1; ETS2; SRF; MATK; FOXD3; POU5F1; NANOG; PRDM14; MTHFR; TXNRD1; MTRR; PRKAR1A; NF2; PRKACA; TP63; MEST; DAND5; CDH1; FOXL1; MESDC2; SENP1; YAP1; SETD2; KDM6B; HCK; EPHA2.	Hippo signaling pathway mTOR signaling pathway Wnt signaling pathway
Endoderm	Red	19	cell differentiation; developmental process; system development; pattern specification process; transcription from RNA polymerase II promoter.	PITX2; NOTCH1; GATA4; DKK1; TBX20; AXIN1; ZEB2; SOX2; SOX7; RBPJ; POU5F1; ID2; CTNNB1; APC; WNT2; MED12; WNT8A; GDF3; NANOG.	Hedgehog signaling pathway Notch signaling pathway.
	Light Green	18	signal transduction; response to stimulus; cell communication; cell proliferation; MAPK cascade; developmental process; regulation of phosphate metabolic process; biological regulation; transmembrane receptor protein tyrosine kinase signaling pathway; angiogenesis; locomotion; behavior.	ZFP36L1; MTOR; VEGFA; TGFB1; RPS6KB1; MMP2; RICTOR; CXCR4; APLNR; AKT1; KDR; PDGFA; RPTOR; DUSP5; MLST8; MMP9; VEGFC; FGF2.	mTOR signaling pathway; MAPK signaling pathway; TGF-beta signaling pathway
	Green	13	mesoderm development; anatomical structure morphogenesis; developmental process.	HHEX; LHX1; SOX17; MESP1; ZNF202; HNF1B; AFP; EOMES; GATA6; PDX1; HNF4A; ONECUT1; FOXA2.	-
	Cyan	13	developmental process; transmembrane receptor protein serine/threonine kinase signaling pathway; cell surface receptor signaling pathway; cell differentiation; signal transduction; mesoderm development; protein phosphorylation; regulation of transcription from RNA polymerase II promoter; MAPK cascade.	BMP7; NOG; MIXL1; FGF4; NKX2-1; FGF8; CER1; GSC; PAX9; EXT1; BMP4; BMPR1A; WNT7B.	TGF-beta signaling pathway Hippo signaling pathway Hedgehog signaling pathway
	Light Sea Green	13	biological adhesion; cell adhesion; signal transduction; cell communication.	LAMB1; VTN; ITGB2; ITGB5; FN1; ITGA7; ITGA4; LAMB3; LAMC1; FERMT1; LAMA3; ITGA5; ITGAV.	PI3K-Akt signaling pathway
	Blue	8	mesoderm development.	COL12A1; COL11A1; COL8A1; COL4A2; COL6A1; COL7A1; COL5A1; COL5A2.	PI3K-Akt signaling pathway
	Violet	6	transmembrane receptor protein serine/threonine kinase signaling pathway; intracellular signal transduction; cell surface receptor signaling pathway; regulation of transcription from RNA polymerase II promoter; RNA metabolic process; primary metabolic process; cell communication; transcription, DNA-dependent.	INHBA; HMGA2; NODAL; SMAD3; SMAD2; SMAD4.	TGF-beta signaling pathway FoxO signaling pathway Hippo signaling pathway
	Purple	6	biosynthetic process; transcription elongation from RNA polymerase II promoter; nitrogen compound metabolic process.	CDC73; RTF1; SETD2; LEO1; CTR9; PAF1.	-
	Pink	5	chromatin organization.	KDM6A; HDAC1; BPTF; EZH2; KDM6B.	-
	Brown	4	regulation of catalytic activity; regulation of molecular function; MAPK cascade; response to stimulus; regulation of phosphate metabolic process; signal transduction; cell communication; biological regulation.	DUSP6; DUSP2; DUSP4; DUSP1.	MAPK signaling pathway
	Dark Golden Rod Yellow Olive Drab	3 2 1	Without statistically significant results. (*)	TIMP2; MMP8; MMP15; MTHFR; MTRR; MMP14.	-

Supplementary Table 7.3 Transcription factors guiding human pluripotent stem cells towards cardiac mesoderm and neuroectoderm specification.

According to RNA-Seq expression results, statistically significant upregulated genes encoding transcription factors were selected based on the following cutoff: adjusted *p*-value < 0.05 and fold-change > 2.

*** Genes previously predicted *in silico* (on Supplementary Table 7.1).**

Neuroectoderm differentiation									Cardiac mesoderm differentiation													
Day 1	Day 3	Day 6					Day 9			Day 1	Day 3			Day 5				Day 9				
PRDM8	PRDM8	NR2E1	NR4A3*	ZNF439	ZBED3	NR2F1	BHLHE22	SOX1	FOXK1	EOMES*	SOX17*	MEIS1	HHEX*	SOX17*	TWIST1	ZFPM2*	PREX1	FOXA1	PDX1*	SOX18	SNAI2	RHOXF1
SOX1	BHLHE22	BHLHE22	SNAI2	ZNF280C	GTF2H2	ZMAT3	NR2E1	FOXG1	RFX3	TBXT*	EOMES*	FOXD2	IRX5	GATA3	UNCX	SIX3	ID2*	HAND2	ZBTB16	ZNF467	FOXP4	FOXK1
HES5	NR2E1	PRDM8	MECOM	GRHL1	ZNF624	ZNF839	PRDM8	SIX3	ARID5B	CDX1	GATA6*	GATA2	TBX6*	EOMES*	DMBX1	GLIS1	DLX3	HOXB3	TWIST1	KLF2	ONECUT2	ZBTB20
SIX3	LMX1A	RAX	ARX	ZNF441	ZFP37	ZC3H3	FEZF2	MEIS1	ZNF333	MIXL1*	HAND1*	PRDM1	DLX3	HAND1*	TAL1*	DPF3	IRX1	GATA6*	DPF3	MITF	GLIS3	NKX3-1
FEZF1	SIX3	NEUROD4	FEZF2	SP8	ZSCAN29	ETV3	NEUROD4	PRDM16	SETBP1	CDX2*	FOXA2*	NR0B1	ZEB1	HOXB3	PLXNA2	MYRF	ELF4	HAND1*	TSHZ2	BMP2	TIPARP	ID3
	PAX6*	PRRX1*	OTX1	VEZF1	KLF13	MBNL2	ARX	EBF3	SOX21	SP5	FOXC2*	FOXF2	SMAD9	GATA6*	HOXA3	PLAGL1	ZNF124	TBX2	PRDM8	PLEK2	MBNL2	ZNF441
	RAX	LMX1A	POU3F2	SOX8	ZNF343	ZSCAN5A	PAX6*	LHX5	ZNF781	HOXA1	HOXB3	ZNF503	PRDM6	HOXB2	SMAD6	GLIS3	DDIT3	NKX2-5*	SOX6	ZFHX3	ID2*	ZNF83
	GLIS3	WNT8B	GLIS3	ZNF835	ZNF438	FOXO3	DMRT3	SNAI2		GSC*	HOXB2	RXRG	DPF3	FOXA1	LEF1*	SOX18	SALL1	HOXB2	PRDM6	MSX1	ELF4	ZNF611
	ZEB2*	SIX3	MEF2C	TIPARP	ZNF138	TOX2	EMX2	RAX		LEF1*	GATA3	RUNX1	ZNF396	FOXA2*	PITX2*	PRDM6	ZNF467	HOXB4	HHEX*	MYRF	ZHX1	ZNF608
	SP5	PAX6*	ZNF20	SOX6	HES4	ZBTB41	BARHL1	ZNF521		GATA6*	FOXA1	GLIS1	ZNF703	GATA4*	ZNF503	ZNF703	NFATC1	MECOM	ONECUT3	FOXP2	SP5	ID1
	NR2F2	NKX6-1	ZNF471	KLF8	PRR3		NR2F1	MECOM		GBX2	FOXC1*	SIX3	IFI16	ISL1	SMAD9	IFI16	SP6	FOXA2*	NKX2-1*	ZEB1	ESRRG	FOSL2
	FOXB1	EBF1	ZNF823	PRDM2	NFATC2		NR2F2	LEF1*		FOXA2*	MIXL1*	MECOM	MYRF	FOXC2*	MESP1*	ZEB1	ID1	HOXB5*	ZNF503	CDX2*	ZNF704	
	PAX2*	LHX9	FOXS1	ZHX1	BRPF1		POU3F2	DMRTA1		PITX2*	GSC*	FOXF1*	ZNF436	HOXB4	NKX2-3	KLF5	ZFY	TBX3*	FLI1	NR4A2	SOX5	
	NEUROD4	ZBTB16	PROX1	ZNF410	ZNF781		HES5	MEIS2		ZEB2*	SP5	HAND2	ZNF611	GATA2	RBM20	IRX5	ZNF441	GATA4*	NR2F2	ZNF703	PRDM1	
	LHX5	NR2F2	ZNF19	ZNF502	ZNF800		NEUROG2	FEZF1		FOXB1	HOXB5*	SOX6	LYL1	TBX2	ZEB2*	ZNF611	ESRRG	GATA2	LEF1*	STAT4	SOX7*	
	LHX2*	BARHL1	TIGD2	PLXNA3	ZNF467		LHX9	HES4		PLAGL1	PITX1	MESP2*	SOX5	PITX1	BMP2	ZNF396		SOX17*	MSX2	IRX3*	SORBS2	
	ZNF503	LHX2*	ZNF521	BMP2	IKZF5		NEUROD1	SP8		HES7*	BHLHE22	HMX1	SP6	TBX3*	SP5	ETV2*		TBX5*	TAL1*	HEY1	HES7*	
	FEZF1	CDX1	LEF1*	ZNF18	ZNF442		LHX2*	ZFHX3		GATA4*	GATA4*	RARB	PKNOX2	HAND2	SOX6	HES4		GATA3	FOXF1*	HLX	KLF5	
	ZFHX4	POU3F3	SATB2	CREB5	ETV2*		DMRTA2	SOX5		ZNF703	PLXNA2	PLAGL1	NR2F2	FOXC1*	TSHZ2	SMAD7		ISL1	FOXF2	EBF4	HES1	
	OTX1	ZNF575	SOX5	ZNF710	ZNF619		ST18	FBN1		MNX1	PAX2*	BMP2	EBF4	HOXB5*	NR2F1	ID4		MEIS2	HOPX*	PLAGL1	PCSK6	
	EMX2	ZEB2*	ZNF396	NFRKB	NRF1		ZNF503	PROX1		TCF7	MESP1*	PITX2*	KLF1	TBR1	FOXL1*	NR2F2		TFAP2A	FOXC1*	HES4	NFATC2	
	FOXA1	FOXA1	HES7*	ZNF763	ZNF426		LMX1A	SKIDA1		PLXNA2	TBR1	SOX18	PREX1	NKX2-5*	PRDM1	MIXL1*		NKX2-8	PAX1	BNC1	CASZ1	
	SOX5	DMRT3	FEZF1	MEIS2	ZSCAN20		ZEB2*	POU3F1		MBNL3	HOXB4	MNX1	TCF7	MSX2	FLI1	MESP2*		PITX1	RBM20	ONECUT1*	ZNF546	
	POU3F2	IQCN	TSHZ2	ZNF136	ZNF414		ONECUT3	ZHX1		MSX1	TWIST1	SHOX2	HES4	TFAP2A	FOSB	KLF1		GATA5	WT1	LYL1	FOXP1	
	ZNF521	SIX6	MEIS1	FOXK1	ZNF592		ZEB1	GLI3		NPAS3	MSX2	SNAI2	MEIS3	RARB	ZBTB16	TIPARP		PITX2*	TSHZ1	HIVEP3	CREB3L2	
	SOX1	EOMES*	BHLHE40	ZFYVE26	ZFHX3		EBF1	MYT1		ATF5	TBX3*	GATA5	IRX1	MEIS1	TBX5*	SOX5		HNF4A*	ZEB2*	IFI16	PPARA	
	ZEB1	KLF4*	ZNF174	ZNF780B	ZC3H7A		ZBTB16	ZNF441		ZFHX3	LEF1*	DMBX1		NKX2-1*	MSX1	EBF4		HOXA3	SMAD9	ID4	BCL6	
	POU3F1	ZNF503	FOXB1	FBN1	TOX3		PLAGL1	ID4			PRRX1*	TSHZ2		MEIS2	FOXF1*	SOX7*		RARB	PRRX1*	ZNF436	DLX3	
	LEF1*	ATOH8	FOXN4	HMBOX1	ZNF454		NHLH1	FOXB1			MEIS2	IRX3*		FOXF2	HHEX*	ZNF546		RFX6	PLXNA2	EOMES*	ZNF124	
	ZBTB16	ZEB1	ZSCAN1	RPA4	SOX21		SOX6	MBNL2			MSX1	RBM20		PAX2*	FOS	KLF8		MEIS1	RXRG	PLXNC1	SKIDA1	
	HESX1	TSHZ1	HES5	ESRRG	POU3F1		ZFHX4	ZNF516			FOXP2	SMAD6		GATA5	TSHZ1	ZNF83		MEF2C	SMAD6	EPAS1	KLF12	
	TSHZ2	PAX8	EMX2	ZNF461	REL		RFX4	TOX3			ZEB2*	HLX		NKX2-8	HLX	PKNOX2		FOXC2*	ISX	PKNOX2	DDIT3	
	SOX21	NEUROG2	BARX1	ZNF713	ZNF385D		GLIS3	MXD3			TBXT*	NFATC1		HNF4A*	SNAI2	ZNF436		NR2F1	HOXB6	IRX5	ZNF471	
	FBN1	SP100	LHX5	CASZ1	ZBTB25		TSHZ2	PLXNA2			ISL1	ZBTB16		MECOM	HEY1	JUN		ZFPM2*	TBX20*	IRX4	MEF2A	
	SETBP1	ZFHX4	ZNF385B	MBNL3	RFX2		OTX1	ZNF385B			CDX2*	KLF5		NR0B1	IRX3*	MXI1		NKX2-3	PROX1	SMAD7	PRDM16	

Supplementary Table 7.4 Expression profiling analysis of fetal and adult tissues by high throughput sequencing.

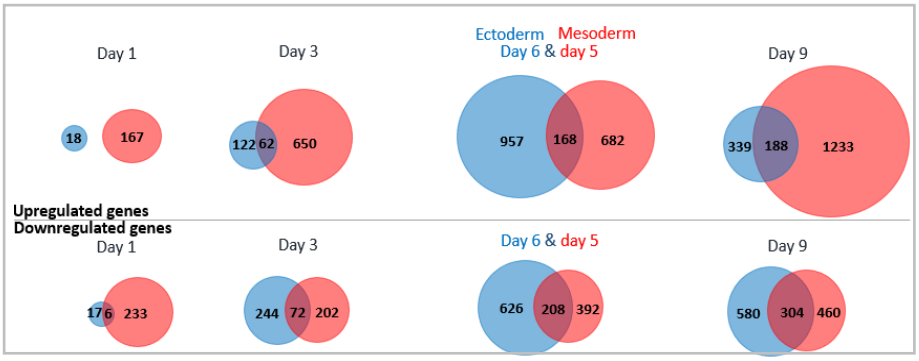
Description of collected datasets deposited with GEO to compare RNA-Seq gene expression profiles from neural, cardiac and hepatic fetal and adult cell types.

GSE nº	Reference	Differentiation	Fetal/adult	Cell type	Gender and age
GSE106640	Long et al., 2018.	neural	fetal	cortical plate of fetal human neocortex, following pregnancy termination	Due to protection of data privacy, gender was not reported. 12, 13 and 14 gestational week
GSE72930	Dincer et al., 2015.	neural	adult, post-mortem	human prefrontal cortex (PFC) neurons	25 controls without known neurological or psychiatric disease (male=18, female=7, mean±s.d. age =30.8±29 years, postmortem interval = 14.75± 9 h, pH=6.5±0.3). From these, seven adult control PFC specimens were used
GSE62913	Kuppusamy et al., 2015.	cardiac	fetal	HFV and HFA cells	Gender was not reported. 3-months-old of human fetal ventricular (HFV) and atrial (HFA) in vivo samples
GSE93841	Mills et al., 2017.	cardiac	adult	The adult human heart sample was obtained from Clontech and pooled from three hearts	Males. 30- to 39-y-old Caucasian males who died from trauma
GSE58557	Wang et al., 2016.	hepatic	fetal	fetal hepatocytes	Male. Fetal liver tissues at 19 gestational weeks were obtained from abortion patient with consent
GSE75606	MacRae et al., 2015.	hepatic	adult	Three adult human non-tumor liver samples	Gender and age were not reported

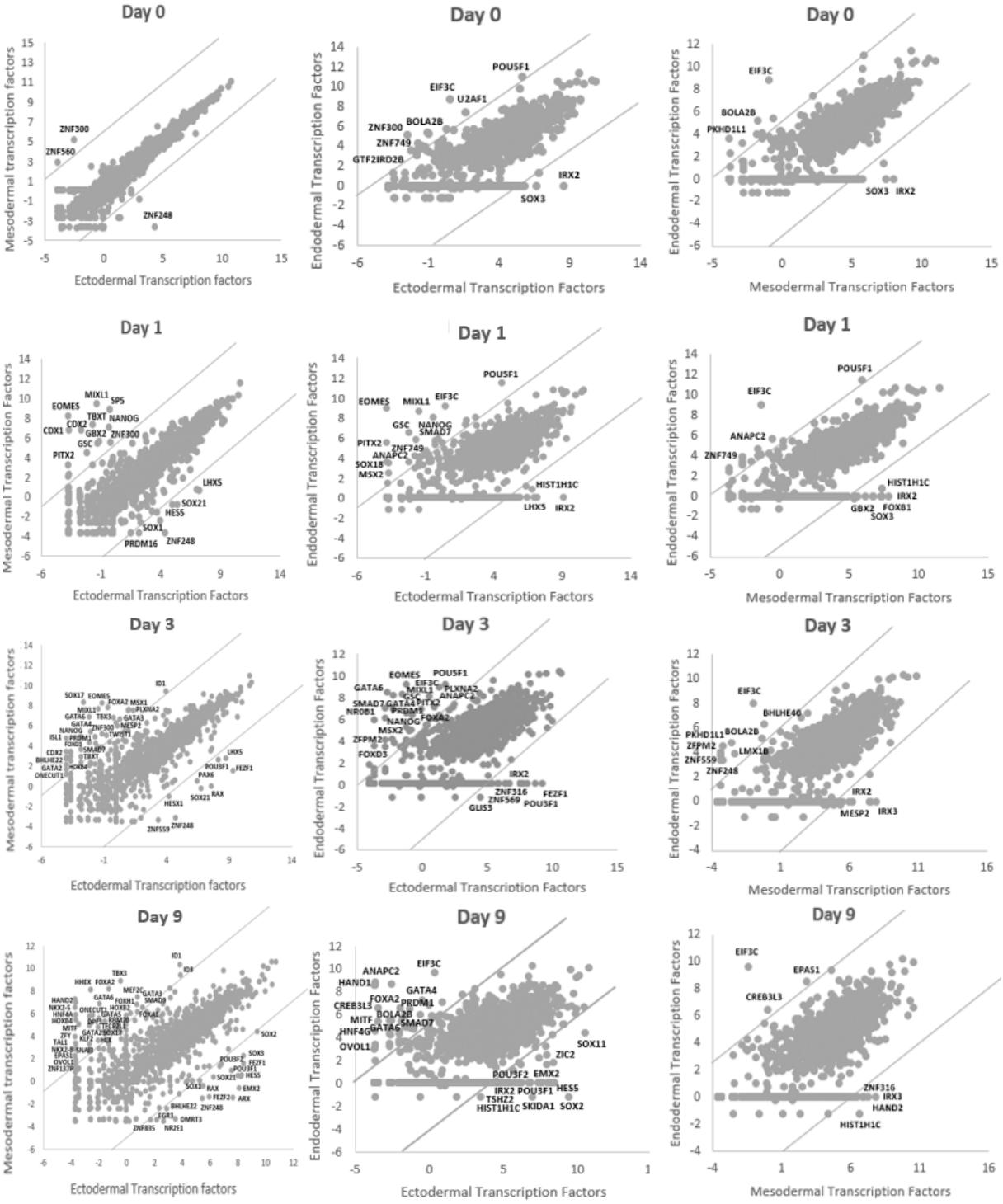
Supplementary Table 7.5 Expression profiling analysis by high throughput sequencing.

Description of collected datasets deposited with GEO to compare RNA-Seq gene expression profiles from neural, cardiac and hepatic differentiation.

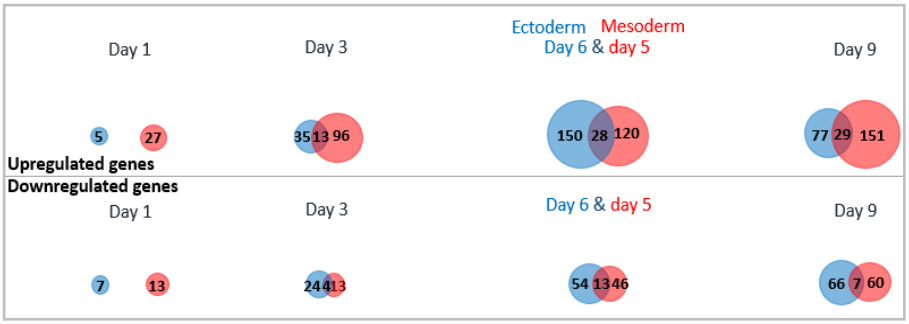
GSE nº	Reference	Time points	Differentiation	Cell type	Differentiation medium
GSE99937	Meganathan et al., 2017	0, 4, 14, 25, 35	Neural	H9 and HES3 hESC lines	ventral NPC (VNPC) medium, 1xb27 supplement without vitamin A, 10 µM SB431542, 1.5 µM purmorphamine, 100 nM LDN-193189, 2 µM XAV-939, and 10 µM Y-27632. After 15 days of EB culture, small molecules from the VNPC medium were removed, following the addition of the γ-secretase inhibitor DAPT (10 µM), cAMP (200 µM), ascorbic acid (200 µM), and BDNF (20 ng). At day 35, Neural Rosette Selection reagent was used to purify.
GSE115575	Frank et al., 2019	0, 1, 2, 4, 6, 8, 14	Cardiac	Human: WA01 (H1 human ESCs), genotype:male; Human: HuES6 (human ESC), genotype:female; Human: HEK293T cells, genotype: N/A	ITS medium with 10 mM Y-27632, 25 ng/ml FGF2, 1-2 ng/ml BMP4 and 1-2 mM CHIR99021. After 24 h, medium was changed to TS medium. After 48 h, medium was changed to TS medium supplemented with 10 mM IWP-2 for 48 h, followed by TS media with 250 mM 2-Phospho-L-Ascorbic Acid. After 48 h, medium was changed to fresh TS until beating cells were observed at days 8.
GSE70741	Li et al., 2017	0, 1, 2, 3, 5, 7, 9, 11, 13, 21	Hepatic	Human H1 ES cells (WiCell)	mTeSR1 medium to maintain hESC. RPMI/B27 medium (Insulin minus), supplemented with Activin A for 3 days, followed by 4 days with 20 ng ml1 BMP2 and 30 ngml1 FGF-4 in RPMI/B27 (complete with Insulin), then 6 days with 20 ng ml1 HGF and KGF in RPMI/B27 (complete with Insulin), then 8 days with 20 ng ml1 Oncostatin-M in hepatocyte culture media, supplemented with SingleQuots (without EGF).



Supplementary Fig. 7.1 Differentially expressed genes (DEGs) overview in ectoderm and mesoderm differentiation. Venn diagrams highlight the number of DEGs in ectoderm (blue) and mesoderm (red) differentiation. Significantly DEGs were selected based on the following cutoff: adjusted *p*-value < 0.05 and fold-change > 2. Day 6 of ectoderm differentiation was compared to day 5 of mesoderm differentiation in order to obtain more early time points to analyse.



Supplementary Fig. 7.2 Transcription factors regulating human pluripotent stem cell fate transitions.
Scatter plots comparing the indicated samples. Genes encoding TFs were screened, according to their *log*$_2$(CPM), at sequential stages of differentiation. An additional published endodermal dataset (Li *et al.*, 2017) was used to broaden our analysis to all three-germ layer-derived cell lineages.



Supplementary Fig. 7.3 Differential expression of key regulators in ectoderm and mesoderm differentiation.
Venn diagrams highlight the number of DEGs in ectoderm (blue) and mesoderm (red) differentiation. Significantly DEGs encoding transcription factors (TFs) were selected based on the following cutoff: adjusted *p*-value < 0.05 and fold-change > 2. Day 6 of ectoderm differentiation was compared to day 5 of mesoderm differentiation in order to obtain more early time points to analyse.

Supplementary Table 7.6 Primer sequences used for quantitative real-time polymerase chain reaction.
Melting temperatures (Tm), amplicon size and primer specificity were estimated using Primer Blast.

Gene	Primer	Tm	GC %	Product length
HOXB2	FW TCCCCTAGCCTACAGGGTTC	60.03	60	180
	RW GGTGGGAGAGGCTCGATTTT	59.75	55	
HOXB3	FW CACTTCTAGCCCCCGGTTTT	59.96	55	185
	RW AGTCCTAACAACCTGCCTGC	59.96	55	
HOXB4	FW CTCCCGATACCCAGCGAAAG	60.25	60	97
	RW ATGGGCACGAAAGATGAGGG	60.11	55	
HOXB5	FW GGCCGCATACATAGCAAAACG	60.6	52.38	157
	RW CGCCCCGAGAAGGAGTTTAC	60.46	60	
HOXB6	FW TGGAGGGTAAATGGACAATCTGC	60.62	47.83	118
	RW TGACTTGGGAGAGAGGGTGTT	60.41	52.38	
RARB	FW CTTCCTCCCCCTCGAGTGTA	60.03	60	111
	RW ACTTCTGCGGAAAAAGCCCT	60.18	50	
TERF2IP	FW GGGCCAGGAGCATAAGTACC	59.89	60	114
	RW TTCTGTGGTTCCCCGCTATC	59.46	55	
RFX4	FW GTGAGTGAGACGGGCAAGAA	59.97	55	79
	RW GCAGCAGTGTTCCGAGTTTG	60.04	55	
ZHX1	FW GGCAAGCAGGCGAAAATCAA	60.04	50	108
	RW GGTGTAAGCACAGGAGGACC	60.04	60	