TIAGO MANUEL SANTOS JUSTO

Role of *Ccbe1* during cardiac differentiation of mouse ESCs

Doutoramento em Ciências Biomédicas

Trabalho efetuado sob a orientação de:

Professor Doutor José António Henriques de Conde Belo

UNIVERSIDADE DO ALGARVE

Departamento de Ciências Biomédicas e Medicina

2016

Role of *Ccbe1* during cardiac differentiation of mouse ESCs

Declaração de autoria de trabalho

Declaro ser o autor deste trabalho, que é original e inédito. Autores e trabalhos consultados estão devidamente citados no texto e constam da listagem de referências incluída.

_____

Researchers have recreated in the laboratory the environmental conditions that they believe existed early in the earth's history. In these experiments, a few scientists have manufactured some of the molecules found in living things. If the chemicals in the experiment represent the earth's early environment and the molecules produced represent the building blocks of life, whom or what does the scientist who performed the experiment represent? Does he or she represent blind chance or an intelligent entity?

If Scientists ever did construct a cell, they would accomplish something truly amazing – but would they prove that the cell could be made by accident? If anything, they would prove the very opposite, would they not?

“You are worthy, Jehovah our God, to receive the glory and the honor and the power, because you created all things, and because of your will they came into existence and were created.” – Revelation 4:11

Acknowledgements

I hereby acknowledge all the support given to me throughout these pays years from all the people in my life. It has been a pleasure to meet and work closely with a wonderful lab crew, whether in normal days or heavy load ones. I reckon all the patience and kindness shown to me by everyone around me, in and out of *campus.* There are a few names that I will always bear closely in mind and thought from those who worked together with me, helping me to achieve this important mark on my academic life and perhaps professional life. Starting with my co-supervisor Paulo Pereira, who many times had to deal with a quite a lot of naive questions and had them explained to me through several perspectives. That was quite a job, and another PhD itself! Anyways I got to quit poetic writing in scientific manuscripts! I thank Professor José Belo for his support, for giving me the chance of developing this work in his team, and for being a great person in real life. I thank Fernando for his friendship, and sharing a house in a Lisbon adventure during our last PhD year was a remarkable experience, and so it says the oven!

There are quite a few other good friends worth mentioning by name for their incredible support. I thank Ana Tátá for having collaborated with me, but specially for being an amazing person, knowing always how to cheer someone. I also thank Professor José Bragança for all instructive discussions throughout my work. I thank all the Belinhos for being cheerful and amazing people: to Zé Inácio, to Rubi, to Carol, to the 2 Martinhas, to Margaret (mostly because of the wine), to Oriol, to the 2 Joões and to Sara. I thank all the CEDOC crew and the afternoon toast buddies (Inês was the main bread source). Also the weekly cake society!

My pre-PhD reallife mates also deserve a good shout of thanks. I thank João Nuno for being a great friend and adviser, Renato, Nagila, Marta, Sara, Angelina, Joaninha Cristo, my fav cous Abe, my brother Luke and to all in my loong next of kin list: Marta, Rute Tiago, Matias, Adriana, Rui, São, Eduardo, Paulo, Sara, Eduardo II, Samuel, Isaac, Débora, Joana, Jonas, Mariline, Daniel, Leonardo, Alexandre, Pedro, Marco, lil Alice, Isa, Elsa, Zé Manel, Henrique (the policeman), avó Lila, Rosa, Justo, mom and dad (the trumpet guy). Nothing would mean a thing without any of you! I hope we can still live many adventures together and I hope I can learn a lot from all of you.

I have learnt endurance and tenderness with avó Lila, which will always have an important special place in my heart, no matter how long it will take to see her beautiful smile again. Thank you for that, and I miss all of our adventures and late night talks!

I thank again my parents Ana and Manuel, and to my four grandparents for providing me all the education I needed to come this far in life, because above everyone it was thanks to them, and I will be for them!

Resumo

Das suas quatro cavidades à sua síncrona rede elétrica, o coração foi perfeitamente projetado para servir de interface entre cada órgão presente no corpo humano. Devido à sua complexidade, as doenças cardiovasculares englobam também um grande conjunto de manifestações clínicas incluindo miocardites, hipertensão arterial, defeitos congénitos cardíacos e doenças isquémicas. Muitas destas patologias traduzem-se geralmente na perda de tecido cardíaco funcional e por outro lado pela formação de tecido fibrótico não funcional. Similarmente ao que ocorre nos países desenvolvidos, em Portugal também as doenças cardiovasculares continuam a ser uma das maiores causas de morbidade e mortalidade.

Devido à limitada capacidade regenerativa do coração e ao facto das terapias existentes para tratar doenças cardiovasculares serem ineficientes ou implicarem enormes riscos para o paciente, é urgente desenvolver novas terapias mais eficazes. Nesse sentido, o uso de células multi e pluripotentes tem contribuído na última década para um franco avanço nesta área. Muitos ensaios clínicos têm sido feitos, ou decorrem ainda, onde se avalia a capacidade regenerativa de células estaminais de diferentes origens na reposição dos tecidos cardíacos danificados. Além disto pensa-se que certos nichos de células progenitoras de cardiomiócitos residentes no coração adulto possam representar um mecanismo endógeno de regeneração. De modo a explorar este mecanismo tem-se recorrido a técnicas de isolamento destas células para transplante em doentes cardíacos. No entanto, até agora as melhorias evidenciadas por essas terapias celulares parecem estar associadas a efeitos parácrinos que as células transplantadas exercem sobre os tecidos envolventes, em detrimento da sua implantação no tecido danificado e consequente diferenciação em novo tecido cardíaco. Em paralelo às terapias celulares tem-se feito um esforço para desenvolver *patches* e *scaffolds* que possam complementar estas terapias por facilitar o *homing* de células transplantadas ao constituírem uma matriz onde estas células possam ser envolvidas e desempenhar a sua função.

Outra alternativa ao uso de células estaminais para uso em terapias de regeneração cardíaca é o uso de células já diferenciadas com identidade semelhante à do tecido a ser substituído. No caso do miocárdio, será potencialmente interessante o uso de cardiomiócitos como fonte em transplantes para a regeneração do tecido danificado. Tal abordagem é especialmente interessante visto terem sido identificadas no coração populações de novos cardiomiócitos derivados de cardiomiócitos já existentes, que contribuem para o *turnover* normal do miocárdio. No entanto, para explorar este mecanismo é necessário criar e otimizar protocolos eticamente aceitáveis para experimentação humana de derivação em grande escala de cardiomiócitos a partir de células pluripotentes. Tal objetivo pode ser alcançado através do uso de fatores segregados que possam ser utilizados para estimular o potencial cardiogénico das células pluripotentes.

A procura de genes envolvidos na cardiogénese têm-se tornado cada vez mais importante com o objetivo de identificar potenciais fatores que possam modular este processo biológico quer *in vitro* como *in vivo*. De facto, é possível modelar *in vitro* com grande rigor os estadios iniciais da cardiogénese através da diferenciação de células estaminais. Tal como ocorre *in vivo*, a especificação das linhagens cardiovasculares *in vitro* implica uma transição para populações de células progenitoras cardíacas com potencial de diferenciação cada vez mais restrito e específico. Começando num estado de pluripotência, durante a sua diferenciação estas especificam-se em mesoderme cardíaca e posteriormente em células de todas as outras linhagens cardíacas. Para monitorizar o seguimento deste processo biológico e para assegurar o correto comprometimento nas várias linhagens cardíacas recorre-se à expressão génica de marcadores genéticos específicos para cada linhagem esperada em cada ponto específico de tempo. Através desta monitorização é possível identificar células de mesoderme cardíaca pela expressão dos genes *Mesp-1* e *Isl-1* a dia 4 de diferenciação das células estaminais, e também diferentes populações de células progenitoras cardíacas pela expressão concomitante de genes como *Isl-1* e *Nkx2.5* em dias posteriores. Assim é possível estabelecer em laboratório um modelo fidedigno e manipulável para se estudar a cardiogénese.

Num rastreio génico efetuado pelo nosso laboratório em células progenitoras cardíacas de galinha com expressão do marcador *Nkx2.5*, foram identificados genes não caracterizados, mas com um potencial envolvimento na cardiogénese. Um destes novos genes identificados foi o *collagen and calcium binding EGF domains 1* ou *Ccbe1.* Na literatura, é possível hoje ver que em modelos animais *knockout* para este gene, um outro processo biológico é afetado i.e. a linfangiogénese. Estes animais apresentam uma total ausência de vasos linfáticos. Este fenótipo deve-se em parte ao papel já identificado que o CCBE1 tem na maturação do fator pro-linfangiogénico VEGF-C. Em humanos a síndrome de Hennekam (associado também a mutações em CCBE1), é caracterizada pela existência de uma rede linfática disfuncional fazendo com que estes apresentem um edema generalizado. Não obstante estes estudos, recentemente verificou-se em ratinho e galinha a expressão deste gene nas regiões embrionárias que dão origem ao coração, sugerindo assim também um potencial papel neste processo. De facto, trabalho efectuado no nosso laboratório veio a demonstrar que o silenciamento deste gene em galinha leva ao desenvolvimento de defeitos cardíacos incompatíveis com a vida, associados a uma redução da proliferação das células cardiacas. Também, em ratinhos *knockout* para este gene é possível identificar um miocárdio subdesenvolvido pelo estreitamento da camada compacta do miocárdio também associado a problemas na proliferação. Assim, no presente trabalho propusemo-nos a estudar mais detalhadamente o envolvimento deste gene nos estadios iniciais da cardiogénese. Como este gene codifica para uma proteína secretada, a verificar-se um importante papel na cardiogénese, a sua manipulação como um fator de crescimento torna-se de grande interesse visando a otimização de protocolos para derivação de cardiomiócitos.

Para estudar os estadios iniciais da cardiogénese recorremos ao uso de uma linha de células estaminais duplamente transgénica que nos permite acompanhar o processo de diferenciação para linhagens cardíacas pois expressam a proteína fluorescente GFP sob o controlo do promotor de *Nkx2.5* e a proteína fluorescente dsRed sob um promotor específico de cardiogénese de *Mef2c.* Assim pode-se confirmar que é possível obter células progenitoras cardíacas *in vitro* correspondentes aos estadios iniciais do desenvolvimento do coração de ratinho. De seguida analisámos o padrão de expressão de *Ccbe1* e

verificou-se que coincide com o aparecimento da expressão dos marcadores genéticos cardíacos, mostrando que *in vitro* a sua expressão ocorre aquando da especificação das células para as linhagens cardíacas. Posteriormente gerámos duas linhas estáveis de células estaminais com silenciamento de *Ccbe1* para avaliar o seu impacto na cardiogénese. Os resultados demonstram que ao diferenciar estas células em agregados 3D conhecidos como corpos embrióides (nome dado devido à sua semelhança física e funcional com um embrião nos estadios iniciais do desenvolvimento), estas células são incapazes de se especificar em mesoderme cardíaca pois apresentam a expressão de *Mesp-1* e *Isl-1* reduzida. Em paralelo com estes resultados, foi possível verificar que os corpos embrióides gerados a partir de células estaminais com silenciamento de *Ccbe1* apresentam um tamanho muito reduzido. Este defeito é devido não a um aumento da morte celular mas sim a um défice na proliferação das células estaminais silenciadas. Estes defeitos na proliferação estão de acordo com outros estudos efetuados pela nossa equipa, em que fibroblastos embrionários derivados de ratinhos *knockout* apresentam grandes problemas na proliferação. Adicionalmente, em embriões de galinha foi verificado necessidade de *Ccbe1* para a correta proliferação de células precursoras cardíacas para formar o tubo cardíaco. Em conjunto, estes resultados demonstram que CCBE1 tem um papel importante em proliferação. Tais resultados são corroborados por experiências onde foi feita a adição de CCBE1 recombinante ao meio de cultura e se observou a recuperação parcial dos corpos embrióides silenciados. Apesar das dificuldades em produzir quantidades elevadas desta proteína recombinante, os resultados indicam que CCBE1 foi capaz de aumentar a proliferação dos corpos embrióides silenciados. No entanto, as células demonstram-se incapazes de se especificar em mesoderme cardíaca, sugerindo que para além deste papel que *Ccbe1* tem em proliferação, o seu papel na cardiogénese é independente deste mecanismo.

Conclui-se assim que *Ccbe1* é indispensável para a especificação das células em diferenciação em mesoderme cardíaca. Para vir a ser utilizado no futuro como fator de crescimento em células estaminais em diferenciação, para derivar grandes quantidades de células cardíacas, é necessário desenvolver

ainda mais estudos que permitam ultrapassar as limitações associadas à sua produção e à sua bioatividade.

Paralelamente a estes estudos, uma outra parte do meu trabalho incidiu numa colaboração com uma equipa de bioinformática, na qual nos propusemos a analisar o transcriptoma de diferentes tipos de células progenitoras cardíacas. O objetivo desta análise seria primariamente identificar através de sequenciação RNA novas isoformas de genes envolvidos na cardiogénese, e adicionalmente identificar novos genes não caracterizados com potencial impacto na cardiogénese. Para tal utilizámos a linha de células estaminais duplamente transgénica já referida, da qual isolámos diferentes populações de células progenitoras cardíacas em dias de diferenciação diferentes. Conseguimos analisar o *dataset* resultante utilizando algumas ferramentas bioinformáticas, que nos permitiu construir uma lista de genes potencialmente envolvidos em cardiogénese ainda não caracterizados. Deste trabalho resultam alguns genes que merecerão um estudo funcional mais detalhado visto estarem claramente expressos nas regiões embrionárias cardiogénicas.

Palavras-chave: cardiogénese; cardiomiócitos; diferenciação de células estaminais; terapia regenerativa; doenças cardiovasculares; *Ccbe1*; sequenciação de RNA.

Abstract

The identification and use of new growth factors to stimulate the cardiogenic potential of pluripotent cells is a safe and alternative approach to develop cell therapies to address the limited regenerative capacity of the heart.

Collagen and calcium binding EGF domains 1 (*Ccbe1*) was firstly identified in our laboratory, which encodes for a secreted protein with potential involvement in cardiogenesis. Knockout animal models for this gene and humans with mutations in *CCBE1,* have lymphangiogenic defects, resulting in the absence of lymphatic vessels. This is in part due to the known described role that CCBE1 has in the processing of the pro lymphangiogenic factor VEGF-C. However, *Ccbe1* is also expressed in the embryonic cardiogenic regions of both mouse and chick and in fact, silencing this gene in chick embryos leads to the development of heart defects incompatible with life. Noteworthy, knockout mice show an underdeveloped myocardium. The objective of the present work is to perform a detailed study of the involvement of this gene in the early stages of cardiogenesis.

The results demonstrate that silencing the expression of *Ccbe1* or blocking CCBE1 in differentiating stem cells, impairs their specification towards cardiac mesodermal lineages. Additionally, we found that differentiating *Ccbe1* KD ESCs have a reduced proliferation rate that leads to smaller EBs. In agreement with this result, when supplementing the differentiating *Ccbe1* KD ESCs lines with recombinant CCBE1, we were able to partially rescue the size of the EBs, but the expression of the cardiac mesoderm markers remained downregulated. These data suggest that those defects are independent from each other, but are intimately related to the disruption of *Ccbe1*, placing CCBE1 as a direct regulator of cell proliferation and cardiac mesoderm specification during ESC differentiation.

Keywords: Cardiogenesis; cardiomyocytes; ESCs diferentiation; cardiovascular disease; *Ccbe1*; RNA sequencing.

List of contents

List of Figures

Chapter I

Chapter III

Chapter IV

List of abbreviatures, acronyms and symbols

#

3D – Three dimensional

A

Afp – Alpha fetoprotein
αMHC – Alpha myosin heavy chain
ao – Aorta
ASC - Adipocyte-derived stem cell

B

BMP – Bone morphogenic protein
BMMSC – Bone marrow-derived mesenchymal stem cell

C

Ccbe1 – Collagen and calcium-binding EGF domain-containing protein
CF – Cardiac fibroblast
Cm/s – centimetre per second
CNTN2 – Contactin-2
Col - Collagen
CPC – Cardiac progenitor cell
CX – Connexin

D

Da - Dalton
dNTP – Deoxynucleotide

E

E – Embryonic day
EB – Embryoid body
EC – Endothelial cell
ECM – Extracellular matrix
EGF – Endothelial growth factor
eGFP – Enhanced green fluorescent protein
EMT – Epithelial-to-mesenchymal transition
EP – Electrophysiological
ESC – Embryonic stem cell

F

FACS – Fluorescence-activated cell sorting
FBS – Fetal bovine serum
FGF – Fibroblast growth factor
FHF – First heart field
Flk1 – Kinase insert domain receptor
FN – Fibronectin
FOXA2 – Forkhead box protein A2

G

Gata4 - GATA binding protein 4
GO-BP – Gene ontology biological process database
GO-MF – Gene ontology molecular function database

H

HCN4 – Potassium/sodium hyperpolarization-activated cyclic nucleotide-gated channel 4
HFR – Heart forming regions

I

Igf2 – Insulin-like growth factor 2
iPSC – Induced pluripotent stem cell
Isl1 - ISL1 transcription factor, LIM/homeodomain

K

KD – Knockdown
KP – Kegg pathways database

L

LA – Left atrium
LEC – Lymphatic endothelial cells
LIF - Leukemia inhibitory factor
LN – Laminin
LSCV – Left superior caval vein
LV – Left ventricle
Lyve1 – Lymphatic vessel endothelial hyaluronan receptor 1

M

MEF – Mouse embryonic fibroblast
Mef2c – Myocyte enhancer factor 2C
Mesp-1 – Mesoderm posterior protein1
MLC2a/v – Myosin light chain 2a and/or 2v
MSC – Mesenchymal stem cell
MYH – Myosin heavy chain

N

NEAA – Non-essential amino acids
NF – Neural fold
Nkx2.5 – NK2 homeobox 5
nN/mm^2 – nanoNewton per squared millimeter
NPPA – Natriuretic peptide precursor A
NRG1 – Neuregulin 1

O

Oct4 - POU class 5 homeobox 1
OFT – Outflow tract

P

PBS – Phosphate buffered saline
PCR – Polymerase chain reaction
PDGF – Platelet-derived growth factor
PDGFR – PDGF receptor
PEA – Poly ethyl acrylate
PN – Primitive node
pt – Pulmonary trunk
PSCs – Pluripotent stem cells
Prox1 – Prospero homeobox 1
PV – Pulmonary vein

Q

Q – Quadrant
qPCR – Quantitative PCR

R

RA – Right atrium
RG – Mouse *Nkx2.5*-eGFP/SHF-dsRed ESCs
RNA-seq – RNA sequencing
rRNA – Ribosomal RNA
RSCV – Right superior caval vein
RV – Right ventricle

S

SAP – Self-assembly peptide
Sca1 – Stem cell antigen 1
SCN5A – Sodium channel protein type 5 subunit α
sFrps – Frizzle-like proteins
SHF – Second heart field
Shh – Sonic hedgehog
shRNA – Short hairpin RNA
siRNA – Short interfering RNA
SMC – Smooth muscle cell
SNP – Single nucleotide polymorphism
SOX – SRY-related high-mobility-group box

T

T – Brachyury
TnC – Troponin C
TnT – Troponin T

V

VEGF – Vascular endothelial growth factor
VEGFR – VEGF receptor

W

Wnt – Wingless-type MMTV integration site family, member 1
Wt1 - Wilm's tumor protein-1

Chapter I
General Introduction

1. Definition and prevalence of cardiovascular disease

From its four different chambers to its synchronous electric network, the heart is perfectly engineered to act as an interface to every single different system present in human organism. Due to heart's complexity, cardiovascular disease can enclose a vast set of cardiac manifestations including inflammatory heart disease, hypertensive heart disease, congenital heart disease and ischemic heart disease. Despite all of these different etiologies cardiovascular disease can have, the ultimate outcome is with no exception very similar – ectopic cardiac function that ultimately leads to scarred and/or dead heart tissue. For example, in ischemic heart disease, coronary insufficiency results in myocardial infarction, and ultimately cardiomyocyte loss.

In Portugal cardiovascular disease is the leading cause of morbidity and mortality in the adult population, being it a proper reflection of that what occurs in developed countries (INE 2013; Jessup and Brozena, 2003).

1.1. Limited cardiac regeneration capacity

As the heart has a very limited regeneration capacity, all injuries caused in heart tissue represent a major medical challenge when it comes to the replacement of the lost tissue. Looking at the major component of the heart the myocardium after an ischemic infarction, contractile myocardial tissue is replaced by non-contractile scar tissue (Cao et al., 2008). Cardiac transplantation has been the standard therapy to overcome a conditioned poorly functioning heart, however is limited by the number of available donors (Jing et al., 2008) and to a series of associated risks such as immunoreactivity, organ rejection and the side effects of immunosuppressive therapies (NHLBI, 2012).

2. Current strategies to regenerate the heart

With the advancement of tissue regeneration technologies on the past two decades, a different light started to be shed on cardiac regeneration, setting in motion the investigation on what could be the real potential of such therapies in restoring lost tissues in damaged hearts. From where we stand now, a lot of progresses have been made in such therapies, as I am going to explain in more detail on the next sections.

2.1. Pluripotent and multipotent cell-based therapies in cardiac repair

The ultimate goal for any cell-based therapy is to regenerate diseased or damaged tissues or cells by the use of autologous, allogenic or xenogenic cells. In the case of the latter two, it would be optimal that these cells lacked immunogenicity in order for the cells to engraft the injured area without triggering an immunological response that could lead to cell rejection and local inflamation. However, life-long lasting immunosupressive therapies often will have to be combined with the use of those cells. On the other hand, autologous cell-based therapies is the ultimate optimal option as this major limitation would be overcome, withdrawing the need to use immunosupressive therapies. Cell-based therapies can comprise diverse delivery strategies, in order to deliver cells into the injured sites or areas, such as systemic intravenous administration or, more specifically, *in situ* administration, eg. intracoronary administration in myocardial repair strategies (Hastings et al, 2014).

In cardiac repair approaches, it was thought that pluripotent or multipotent stem cells could drive regeneration by differentiating and repopulating the damaged tissue in the heart. Hence, types of cells that preserved to some extent a pluri/multipotent capacity have been so far tested aiming this goal. In fact, there are already excellent reviews about the most various cell types explored in order to develop the most efficient therapy, that are currently on phase I and II clinical trials (Boyle et al., 2006; Sanganalmath and Bolli, 2013; Aguirre et al., 2013; Hastings et al., 2014). Accordingly, a meta-analysis from 50 different clinical studies confirmed that overall local benefit was significant, as ejection fraction increased by 3.96 % for a period of at least 2 years in patients with or without myocardial infarction; while present infarct size was reduced by more than 4% (Jeevanantham et al., 2012).

Most of these cell-based therapies rely on the cardiogenic potential that some cell niches have been identified to preserve in adult mammalians (Kim et al., 2015). Apart from the pluripotent potential of embryonic stem cells (ESCs; discussed in more detail on section 4), other cell niches that have been manipulated aiming towards the same goal include mesenchymal stem cells (MSCs), adipocyte-derived stem cells (ASCs), bone marrow-derived

mesenchymal stem cells (BMMSCs) and induced pluripotent stem cells (iPSCs). Although iPSCs have similar differentiation potential as ESCs, one of the major advantages its cell-based therapies offer is that they are patient-specific, meaning that there is a reduced chance of transplant rejection, and are also easy to generate (e.g. with a patient fibroblast sample). In addition, the other cell lines used can only differentiate into more restricted fates as they are multipotent instead of pluripotent like ESCs and iPSCs (Takahashi and Yamanaka, 2006; Gnecchi and Melo, 2009; Nardi and Meireles, 2006; Zuk et al., 2001). However, on the other hand due to their pluripotent state, ESCs and iPSCs have higher tumorigenicity than multipotent or even differentiated cells.

Interstingly, the main evidences so far in large mammals and on the ongoing clinical trials have related the benefits of such therapies more likely to a paracrine effect that transplanted cells exert on the surrounding cells rather than to *in situ* differentiation into new tissue, as initially envisaged (Boyle et al., 2006; Sanganalmath and Bolli, 2013; Aguirre et al., 2013).

2.2. Adult cells-based therapies

Another promising cell population that has been described to have regeneration potential is adult cardiac progenitor cells (CPCs). Despite the heterogeneity, and-nonconsensual origin of CPCs, it has been described that the Sca-1$^+$ compartment of CPCs can contribute, even though at a low rate, to myocardial turnover (extensively reviewed in Valente et al., 2014). Indeed, a clinical trial using cardiospheres-derived Sca1$^+$ cells has shown that these cells contribute to cardiac improvements after myocardial infarction. In this trial it was shown reductions in scar mass (p=0·001), increments in viable heart mass (p=0·01) and regional contractility (p=0·02), and regional systolic wall thickening (p=0.015). However, there were not identified improvements in the left ventricular ejection fraction, the most-expected functional outcome when regenerating the myocardium (Makkar et al., 2012). Therefore there is a need to try to understand at the single cell level, the differences that may exist between the overall Sca1$^+$ CPCs and other Sca1$^+$ stromal cells, such as cardiac fibroblasts (CD90$^+$). It is not clear which cell population, nor to which extent, are

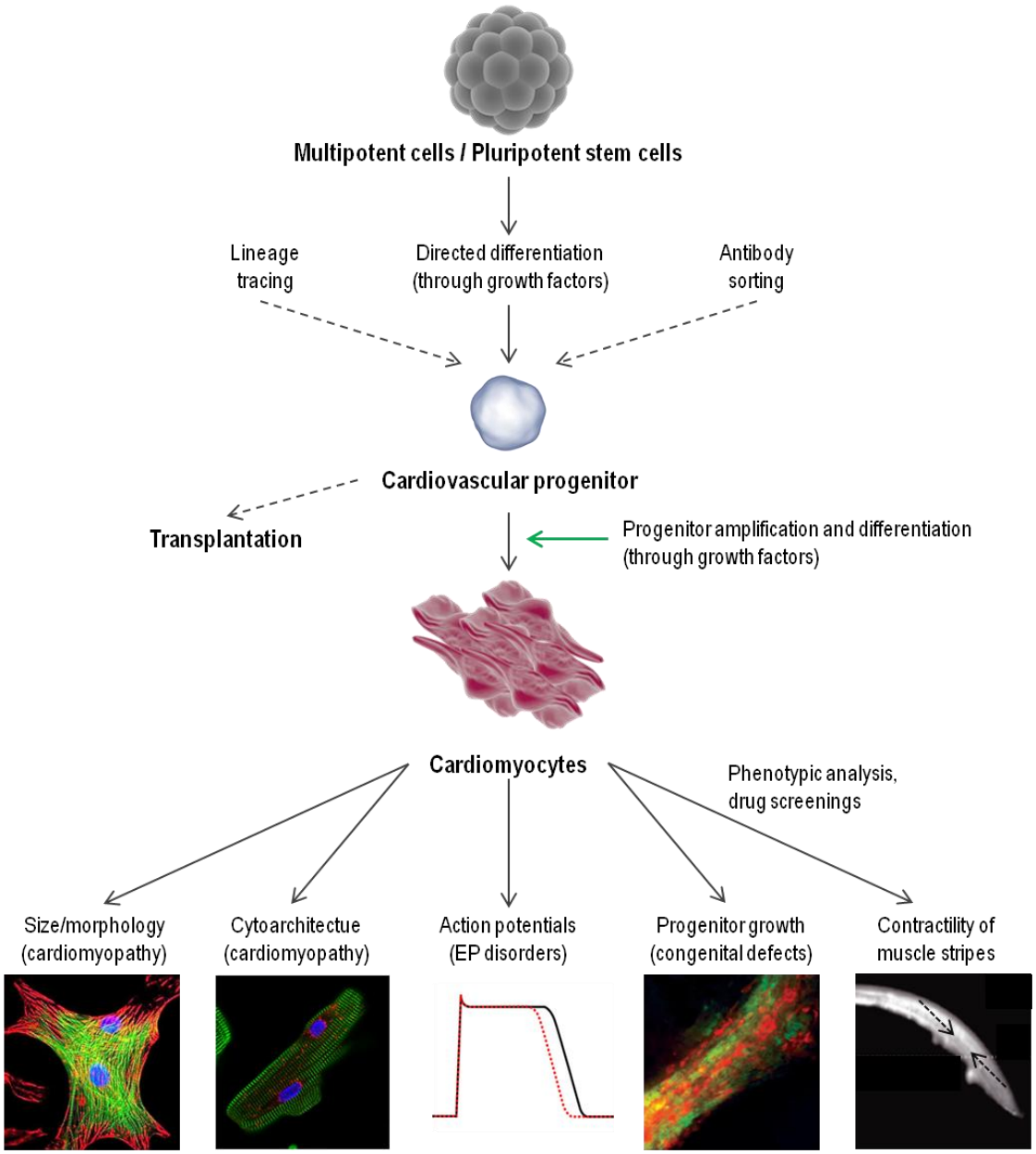


Figure 1.1 – Schematic representation for the potential uses of cardiovascular progenitors and cardiomyocytes in cardiac regenerative therapies. Deriving cardiovascular progenitors from multipotent cells or pluripotent stem cells using growth factors, is a major step for ultimately derive large amounts of cardiomyocytes. On the other hand isolating and purifying adult CPCs populations or cardiovascular progenitors can be transplanted directly into an injured myocardium, or alternatively, be expanded into larger numbers, followed by further differentiation into functional cardiomyocytes. Derived patient-specific cardiomyocytes can be used in various cellular assays, several examples of which are shown, to study and develop therapies for a variety of cardiovascular disorders, including cardiomyopathy, electrophysiological (EP) disorders, and congenital defects. One major goal the production of cardiomyocytes aims is to be used in developing efficient cardiomyocyte transplantation techniques for myocardial regeneration.

these different $Sca1^+$ cell populations contributing for cardiac regeneration (Valente et al., 2014). Being confirmed the existence of the multipotent compartment of cells amongst $Sca1^+$ CPCs, it could be speculated that this therapy could also offer the patient-specific benefits, since after patient's CPCs isolation it would be possible to expand and transplant them into the injured myocardium. Indeed, in Figure 1.1 there is a schematic representation of how isolated cardiovascular progenitors (which can comprehend adult CPC's) could be used (1) for direct transplantation or (2) cultured and differentiated as a source of cardiomyocytes for further transplantation or patient-specific disease modeling (Davis and Stewart, 2011; Garbern et al., 2013).

In an early 2013 study published in Nature by Senyo and colleagues, it was shown how pre-existing cardiomyocytes could be the major source of new cardiomyocytes found in adult mammals' hearts, contributing to myocardial turnover. Interestingly, these new cardiomyocytes derived from already existing cardiomyocytes showed to be more abundant in areas adjacent to myocardium injuries, correlating them to a strong contribution for myocardium regeneration (Senyo et al., 2013). As so, exploring this mechanism – transplanting already differentiated cells into the injured areas (e.g. cardiomyocytes) – can be identified as another approach for cardiac cell-based therapies. Such strategy seems a safer alternative than all of the ones considered so far, as with engraftment cells would substitute the exact same cell types lost during an ischemic event, and their capacity to cause tumors in the host organism is rather lower then pluripotent cells. In fact, earlier studies on this approach in rats has proven the technique to be feasible, for the transplanted cardiomyocytes engrafted the host tissue, proliferated and formed cardiac tissue. In addition, transplanted cells were connected to each other by intercalated disks and the newly formed tissue was also more vascularized then the remaining scarred tissue, however their overall arrangement was disorganized when compared to the host cardiac tissue (Li et al., 1996; Sakakibara et al., 2002). Even though this proves it is possible to transplant cardiomyocytes into ischemic injuries, the authors address some concerns with this technique, such as the used cardiomyocytes being from newborn mice and have being rejected after several weeks post-transplantation. Interestingly, in a

different study it was indicated that cardiomyocyte transplantation only inhibited the progress of cardiac remodeling in chronic myocardial infarction and did not improve cardiac function significantly (Sakakibara et al., 2002). Nevertheless, to test the viability and efficiency of such approach in humans there would be starting limitations needing to be addressed such as developing protocols that allow a scalable, yet ethical, production of cardiac cells for further transplantation.

One attractive and safe way to achieve this goal could be the use of secreted factors that promote the cardiogenic potential of pluripotent stem cell (both ESCs and iPSCs) or CPCs (Hansson and Lendahl, 2013), but examples in the literature of such factors are still limited (Czyz and Wobus, 2001; Hashimoto and Yuasa, 2013; Khezri et al., 2007; Sato et al., 2006; Takahashi et al., 2003; Zeng et al., 2013). However, combining factors such as hypoxia and bioreactor's hydrodynamics has shown to be an interesting approach to efficiently maximize the production of cardiomyocytes from iPSCs (Correia et al., 2014). Nonetheless, there is still room for significant improvements in such protocols – on how to alternatively modulate pluripotent or multipotent cell lineages to achive their full cardiogenic potential – as another challenging limitation is related to the non-maturation state these engineered cardiomyocytes present.

2.3. Cardiac patches and scaffolds in cardiac repair

Myocardial infarction leads to ventricular weakening by replacement of the cardiac muscle fibers by non-functional fibrotic scar tissue, leading to ventricular dilation and wall thinning. To avoid this, actual research is also being developed to design cardiac patches that help to improve these defects upon injury. Cardiac patches are three-dimensional scaffolds engineered from natural or synthetic polymers which aim to be engrafted on the site of the injury to help avoiding the progressive impairment of surrounding healthy tissue, and on the other hand also by improving the restoration of the lost functions. By mimicking the extracellular matrix (ECM), cardiac patches can stimulate to a limited extent biological processes such as cell adhesion, proliferation and migration, which intend to drive tissue regeneration. In fact, Holubec et al. reviews how the use

of different porcine small intestinal submucosa ECM-derived products in clinical cardiac surgery offers the potential for cellular repopulation and growth in different damaged cardiac tissues (Holubec et al., 2014). However, some approaches of engineering such patches can include the encapsulation of living cells into the polymer mesh, resembling a sheet of living tissue with closer similarity to native myocardial tissue.

To have a clinically relevant effect on damaged hearts, such cardiac patches should have around 1 cm of thickness, be able to generate between 20-50nN/mm^2 and also be able to propagate electrical impulses around 25 cm/s (as clearly reviewed and described in Radisic and Christman, 2013). Even though so far the small size of viable cardiac patches for transplantation has been a limitation, Martínez-Ramos and colleagues have recently developed a scalable way to produce injury-size patches which can be grafted into the site of the injured myocardium. While producing scaffolds with similar myocardial physical properties such as elasticity, flexibility and stiffness had been a challenge, this latter group was able to combine a poly (ethyl acrylate) (PEA) scaffold, with self-assembly peptide (SAP) hydrogel RAD16-I and ASCs biohybrid patch that overcomes such limitations (Martínez-Ramos et al., 2014). This clearly shows that synergisms can be created by combining different regenerative strategies. Indeed, in the 6-months follow-up study to their animal models, enhanced systolic and diastolic parameters and also the reduction of the infarct area were identified. Along with the regeneration of the lost myocardium tissue, proper local vascularization of the injured tissue, or even of the engrafted patch, is a requirement for a successful strategy. To meet that need, Ichiara's team have recently developed a biodegradable surgical patch for high pressure systems that enables the incorporation of endothelial cells (ECs) and smooth muscle cells (SMCs) in the scaffold, therefore opening the way to create vascular grafts (Ichiara et al., 2015).

2.4. Current challenges and future directions on cardiac tissue repair

As some of the cell-based therapies have shown a significant but yet modest improvement of the cardiac function, primarily related to paracrine effects of the transplanted cells, cell-homing can play an important role in raising the

efficiency of such approaches. Taghavi and George review many cell-adhesion markers, growth-factors, chemokines, endothelial nitric oxide synthase and hormones responsible to enhance the homing of transplanted cells to the injured myocardium (Taghavi and George, 2013). Despite such promising alternatives to enhance the efficacy of current cell-based therapies, there seems to be strong suggestions that combining cardiac patches with cell-based therapies can already increase the time window of transplanted cells' homing and residency on the local of interest. So whether is by increasing the time of exposure of the damaged tissue to the paracrine effects of metabolites produced by the engrafted cells, or whether is by actually easing the homing of more differentiated cell populations (eg. cardiomyocytes and endothelial cells) into the injury site, the way these strategies are evolving offer a very promising fashion in helping the regeneration of the heart.

3. Mammalian Heart development: from defined progenitor populations to a 4 chambered organ

A better and more detailed understanding of the mechanisms involved in mammalian embryonic cardiogenesis would be beneficial for the development of novel cardiac regenerative therapeutic approaches.

With the growth and development of the embryos, which limits the access to oxygen and nutrients to all the cells, novel embryonic cardiovascular structures are formed to ensure a sufficient supply of nutrients and oxygen to all the cells, and to remove efficiently the cellular waste products. Consequently, around the 3rd week in human embryos and embryonic day (E) 8.5 in mouse embryos, populations of cardiogenic cells start forming the heart to ensure these functions, being the heart the 1st organ to be formed during development (Brade et al., 2013; Carlson, 2014). Due to the conserved similarity in mammalian heart development, mouse cardiogenesis is considered to be a good model for unraveling mechanisms of human heart development. Interestingly, studies in mouse models have determined that specific regions of the embryo are pre-assigned to give rise to specific cardiac structures.

During gastrulation, both in mouse and human embryos, precardiac mesodermal cells expressing *Brachyury* (*T*) and *Mesp1* leave the primitive streak associated with endodermal cells composing the splanchnic mesoderm (Figure 1.2A). Later, these cells migrate anteriorly and adopt a U-shaped disposition – the cardiogenic mesoderm, also called the cardiac crescent (Brade et al., 2013). The endodermal cells that migrate most anteriorly from the

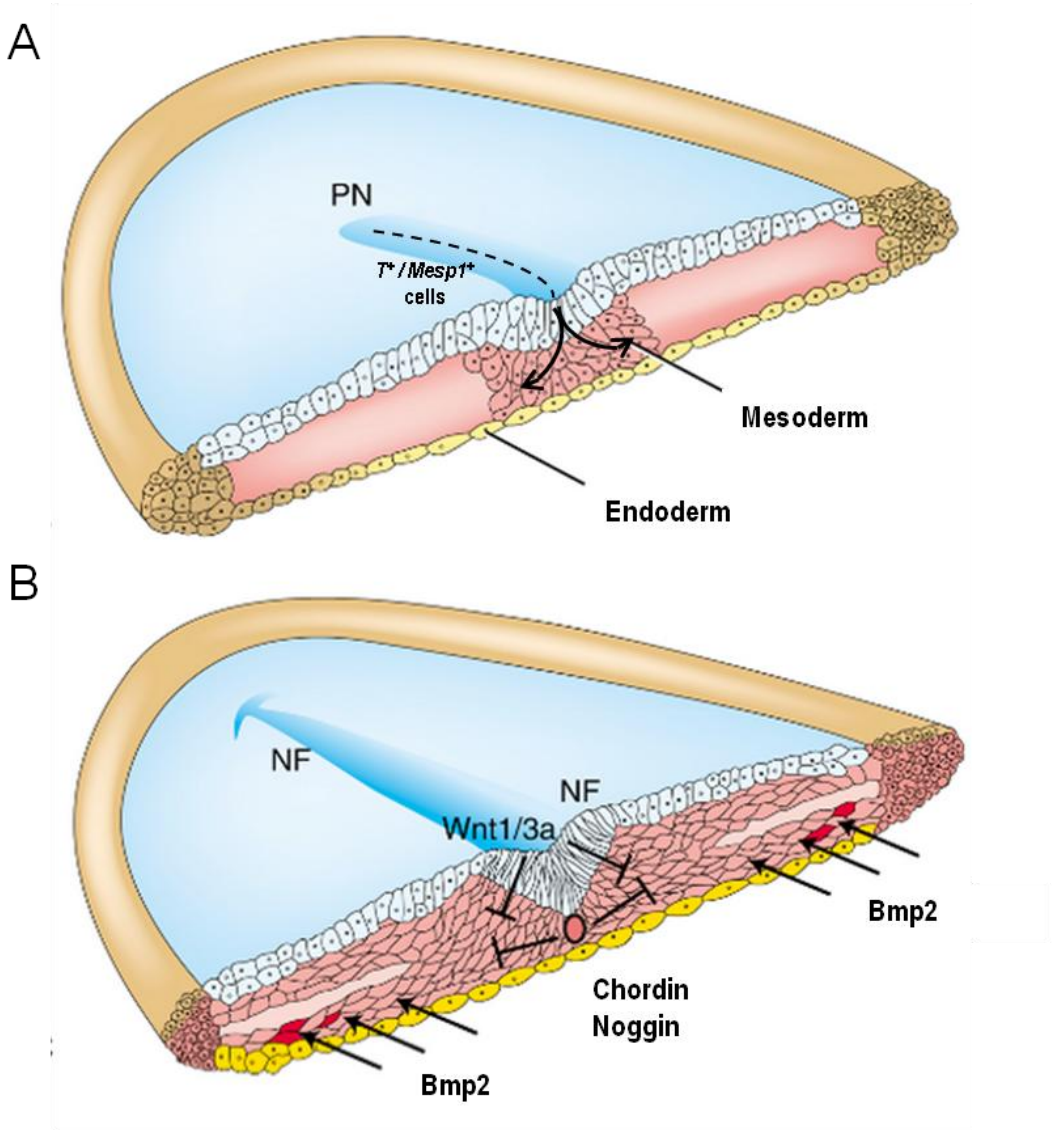


Figure 1.2 – Cardiac mesoderm formation during gastrulation is conserved in human and mice embryos. A) In human, during gastrulation T^+ mesodermal cells, *Mesp1*$^+$ precardiac mesodermal cells and endodermal cells leave the primitive streak migrating anteriorly and formatting the mesoderm and the endoderm; **B)** Bmps, released from the newly formed endoderm, signal the formation of a cardiogenic lineage from the mesoderm (red cells), but their influence is limited to the lateral mesoderm because of the release of chordin and noggin from the notochord and Wnt1/3a from the forming neuroectoderm. NF, Neural fold; PN, primitive node. Adapted from Schoenwolf, 2015.

primitive streak will form the definitive endoderm which is crucial for cardiac mesodermal cells specification. This process is conserved in mice and human embryos (Lewis and Tam, 2006). During its formation, this structure secretes cardiogenic inductive signals such as bone morphogenic proteins (Bmps), fibroblast growth factor (Fgf), activin, insulin-like growth factor 2 (Igf2) and sonic hedgehog (Shh). These factors contribute to the commitment of mesodermal cells to cardiac fates, and also promote their proliferation and survival (Lewis and Tam, 2006; Schoenwolf, 2015).

3.1. Endoderm genetic networks defines cardiogenic mesoderm

However all the mesoderm in exposed to these signals, only the cranial part of the lateral mesoderm commits to cardiac fates. In one hand, the lateral specification is related to the inhibitory effects that secreted factors by both the notochord and the neural tube exert on the Bmp signaling (Figure 1.2 B). Chorddin and Noggin are secreted by the notochord, and act by sequestering Bmps, keeping them from binding to their receptors. Wnt1 and Wnt3a are secreted by the neural tube and are antagonizers of the Bmp signaling (Figure 1.2 B; Schoenwolf, 2015). On the other hand, the cranial specification results from the secretion of dickkopf proteins by the cranial definitive endoderm, and from the presence of frizzle-like proteins (sFrps) on those same cells. While sFrps will sequester the Wnt molecules secreted in the cranial mesoderm, dickkopfs molecules will act by binding both to the Wnts and its co-receptors, abrogating their cardiogenic inhibitory signal (Schoenwolf, 2015). Bmp2 signaling will hence act restrictively as an early stimulus to the expression of early cardiogenic transcription factors within the lateral mesoderm, such as *Nkx2.5* and *Gata4*, and its role is also conserved in other vertebrates (Carlson, 2014, Andrée et al., 1998; Schultheiss et al., 1997). For this reason these early mesodermal cardiogenic fields are located bilaterally and later merge to form the cardiac crescent. From this structure two distinct pools of cells can be identified through the expression of unique markers which give rise to specific cardiac structures: the first heart field (FHF) and second heart field (SHF) populations (Kelly et al., 2001; Abu-Issa et al., 2004).

3.2. Two defined cardiac progenitor populations

The FHF cardiac progenitors are known to express *Nkx2.5* (red cell population in Figure 1.3 and Figure 1.4) soon after the onset of gastrulation under the influence of Bmps secreted by the adjacent endoderm, and are derived from splanchnic mesoderm, which gives rise to the heart tube and subsequently will contribute to the left ventricle and atria (Dehaan 1963, Zaffran et al., 2004; Meilhac et al., 2014; Carlson, 2014). The SHF progenitors are characterized by the expression of *Isl1* (green cell population in Figure 1.3 and Figure 1.4),, which together with the *Gata* transcription factors will drive the expression of a specific SHF enhancer of *Mef2c* in pharyngeal mesoderm, the embryonic region

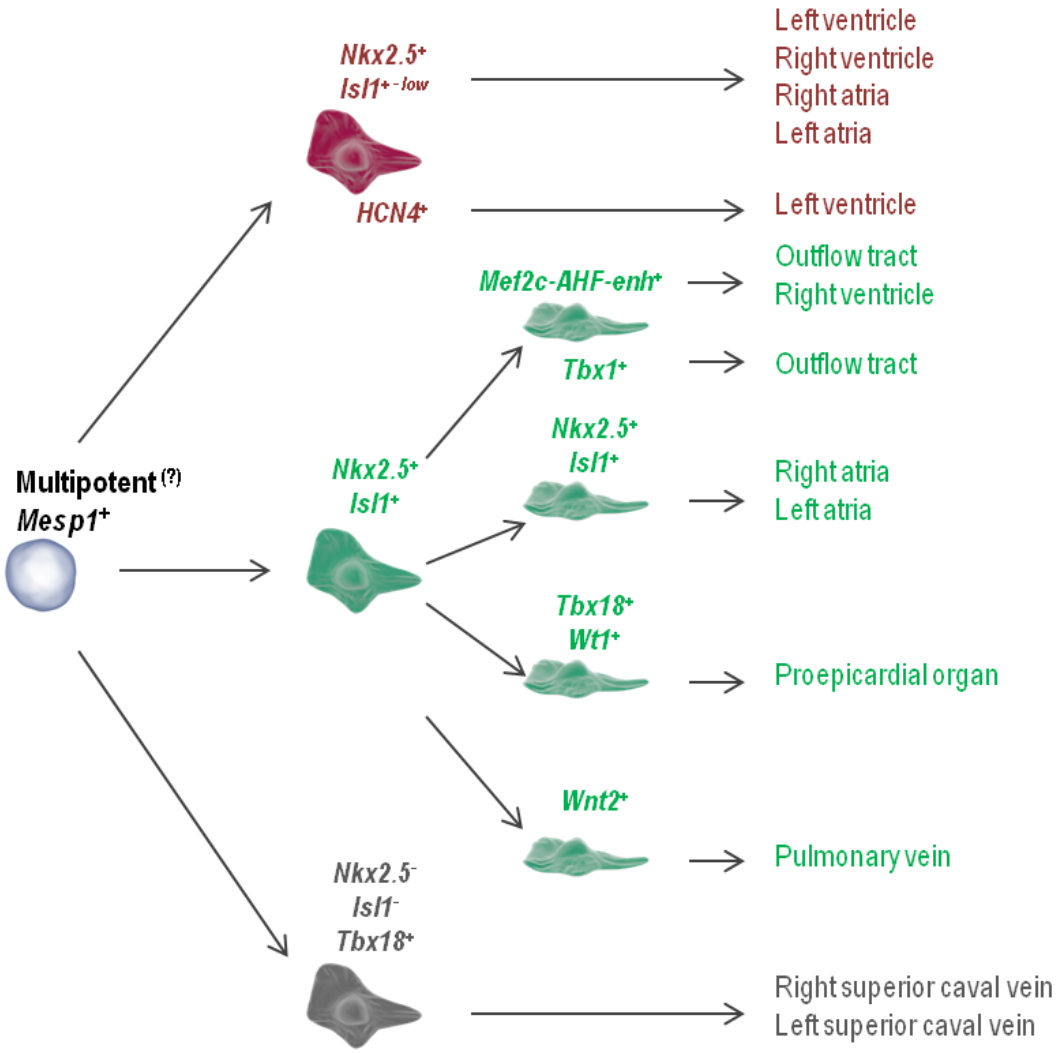


Figure 1.4 – Genetic origin of cardiac components. Genetic tracing with a *Mesp1-Cre* and *Rosa26* conditional reporter shows that almost all cardiac cells in the heart are labeled, so that *Mesp1* marks all cardiac progenitors.

where this lineage is derived from. *Nkx2.5* enhancer in the SHF is then activated by these two transcription factors, leading to its expression in this second pool of embryonic cardiac progenitor cells (Meilhac et al., 2014; Kelly and Evans, 2010). SHF progenitors hence lie medial and slightly caudal to the FHF within the lateral plate mesoderm (Figure 1.3; Schoenwolf, 2015). At E8.0 in mouse and 3rd week in humans, the primordial heart tube is composed mainly by FHF progenitors when the cardiac crescent fuses at midline, after which it starts beating and undergoes rightward looping (Zaffran et al., 2004). Proliferating cells from the SHF start to migrate to the newly-formed heart tube contributing to its elongation and growth at both arterial and venous poles (Figure 1.3). SHF progenitors will give rise to the outflow tract, right ventricle and atria of the developing heart (Buckingham et al. 2005; Kelly et al. 2001). At day 32 in human gestation and E10.5 in mice, the heart presents a well-defined 4 chamber structure, which resembles the form it will have as a mature heart (Brade et al., 2013).

While most of the FHF- and SHF- derived cells are going to mature into cardiomyocytes and compose the myocardium, other cell types found in the heart, i.e. smooth muscle cells (SMCs) and cardiac fibroblasts (CF), will arise

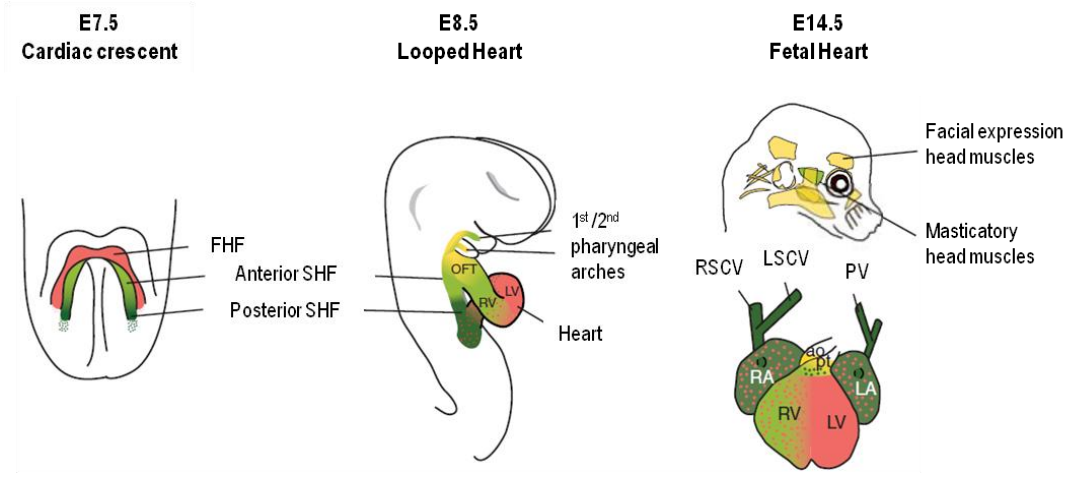


Figure 1.3 – Contribution of the heart fields to the mature tissues of the mature heart and head. First heart field (red; FHF) and second heart field (SHF; green) and anterior (pale green/yellow) or posterior (dark green) subdomains of the SHF are shown at different stages of heart and head development. Regions of the heart with a dual origin are shown with colored dots. ao, aorta; LA, left atrium; LSCV, left superior caval vein; LV, left ventricle; OFT, outflow tract; pt, pulmonary trunk; PV, pulmonary vein; RA, right atrium; RSCV, right superior caval vein; RV, right ventricle. Adapted from Meilhac et al., 2014

from the epicardium. The proepicardial organ is marked by the expression of *Tbx18* and *Wilm's tumor protein-1* (*Wt1*), and is the embryonic structure that gives rise to the epicardium (Figure 1.4). This structure derives from a specialized group of cells within the splanchnic mesoderm during E9.5 in the caudal dorsal mesocardium/septum transversum junction (Meilhac et al., 2014). As the heart looping starts, these cells will start migrating in order to cover all the surface of the myocardium and form the epicardium, and then will undergo epithelial-to-mesenchymal transition (EMT) to enter the myocardium and give rise to CFs and SMCs of the coronary vasculature (Meilhac et al., 2014; Carlson, 2014; Schoenwolf, 2015).

3.3. ECM in cardiogenesis: collagens and fibronectin

The ECM provides structural support for the formation and maintenance of 3 dimensional organs and tissues within an organism. However, the ECM is also a communication net of molecules that allow cells to sense and interpret the environment around them. As a response to those stimuli cells can undergo many cellular processes such as adhesion, migration, proliferation, apoptosis, transformation, and even secrete to that same net additional factors, giving back a response to the surrounding environment. How cells interact with each other and with the surrounding ECM is hence important for the continued understanding of cardiogenesis and cardiac defects (Bowers and Baudino, 2010).

One of the earliest contributions of the ECM to the developing embryonic heart happens before the migration of the mesenchymal cardiac cells to an acellular compartment called the cardiac jelly, located between the myocardium and endocardium of the primitive heart tube. In the mammalian heart the ECM is mostly composed by collagens (Col) of types I, III, IV, VI, fibronectin (FN), laminin (LN) and elastin (Schenke-Layland et al., 2011; Burggren and Keller, 1997). Even though it is not possible yet to understand the role of each one of these single components of cardiac ECM, relevant information is already known on the roles of the FN and Col I, Col IV and LN. In the case of the FN it is known that its loss-of-function in mice leads to severe cardiac malformations (George et al., 1993). More recently, is was described that proliferating niches

of *Isl1+/Flk1+* cardiac progenitor cells within the right ventricular free wall, the atria and outflow tract of both mouse and human developing hearts, were characterized by a rich Col IV and LN ECM. Indeed, data from the same report strongly suggested that such ECM composition was important to maintain cardiac progenitors in an undifferentiated state, prior to their migration to populate other parts of the heart. Interestingly, while cardiac progenitors migrated from the niche, the surrounding ECM rich in Col I and FN promoted their differentiation towards cardiomyocytes and vascular cells as *Isl1* expression was downregulated and cells started to express Troponin C (TnC; Schenke-Layland et al., 2011).

It is getting clear now the determining role of the ECM composition in the developing heart. Now the aim is to try to understand what particular cellular processes do these components regulate and promote in the heart, in order to also exploit such mechanisms in therapeutics (Bowers and Baudino, 2010). In the next section, with the explanation of why are pluripotent stem cells (PSCs) a good model for embryonic cardiogenesis, there will also be addressed some findings that have been made to better understand the role of Col I, Col IV and FN in controlling cellular processes using *in vitro* models.

4. Mouse and human pluripotent stem cell differentiation in vitro recapitulate cardiac differentiation

As mentioned earlier, ESCs offer a very promising potential as a source of cells for heart regeneration. In fact, ESCs and iPSCs by being PSCs, when manipulated and differentiated *in vitro* allow the possibility to recapitulate some of the crucial steps for cardiac specification. Indeed, it is possible to derive cells expressing specific genetic markers of both early- and late- cardiogenesis by differentiating PSCs. However, whether is *in vivo* or *in vitro*, the specification of the cardiovascular lineages involves a transition through a sequence of increasingly restricted progenitor cells, proceeding from a pluripotent state to mesoderm and then to cells committed to cardiovascular fates (Figure 1.5; Laflamme and Murry 2011). Culturing and differentiating mouse PSCs as cell aggregates, called embryoid bodies (EBs), has become a routine in many laboratories, since it was proved back in 1985 that spontaneous *in vitro*

differentiation of ESCs could give rise to cells from the 3 embryonic germ-layers (Doetschman et al. 1985). At the same time the differentiation protocol is triggered on PSCs after removal of *Leukemia inhibitory factor* (LIF) from the culture medium the cells are cultured in hanging drops (Figure 1.6Figure 1.). This technique, which requires the preparation of a cell suspension with a precise cell density, allows the aggregation of the cells with the stimulus of the gravity since these cell suspension drops are cultured in inverted bacterial dishes. After they aggregate, they become EBs, resembling the inner cell mass of early embryos. Simultaneously with this physical change, cells also lose the expression of the pluripotency genes *Nanog, Oct4* and *Sox2*, and start expressing germ-layer specific genes. As the EBs grow in size, they form an outer shell-like layer composed of cells and an enriched collagen IV and laminin

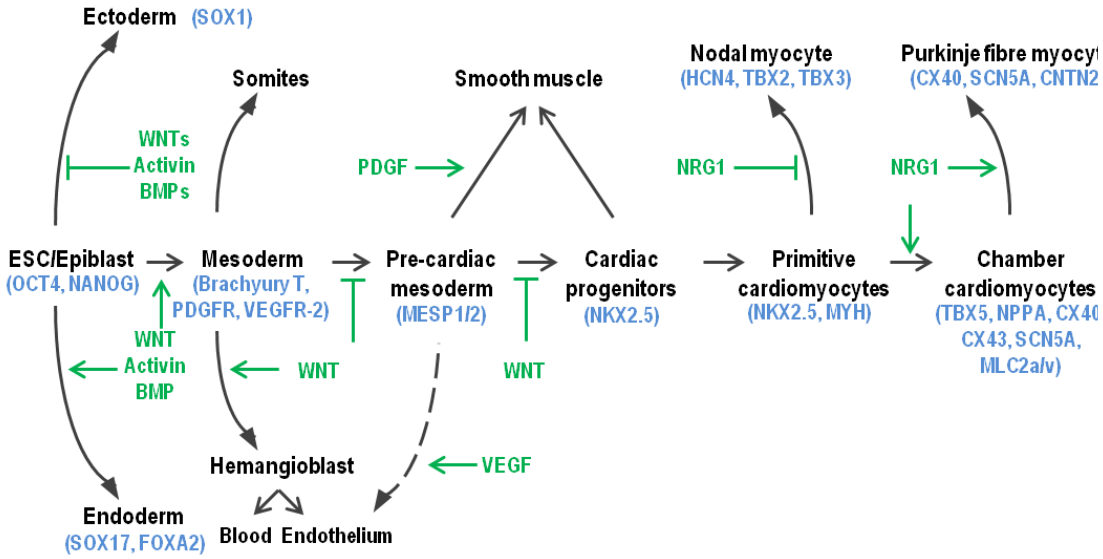


Figure 1.5 – Growth factors and key transcription factors that regulate fate choices during early embryonic cardiogenesis and ESCs differentiation. Growth factors that regulate fate choices are listed at branch points (green), and key transcription factors and surface markers for each cell state are listed under the cell types (blue). The growth factors are useful for directing the differentiation of ESCs, whereas the markers are useful for purifying cells at defined developmental states. BMPs, bone morphogenetic proteins; CNTN2, contactin-2; CX, connexin; FOXA2, forkhead box protein A2; HCN4, potassium/sodium hyperpolarization-activated cyclic nucleotide-gated channel 4; MESP, mesoderm posterior protein; MLC2a/v, myosin light chain 2a and/or 2v; MYH, myosin heavy chain; NPPA, natriuretic peptide precursor A; NRG1, neuregulin 1; PDGF, platelet-derived growth factor; PDGFR, PDGF receptor; SCN5A, sodium channel protein type 5 subunit α; SOX, SRY-related high-mobility-group box; TBX, T-box transcription factor; VEGF, vascular endothelial growth factor; VEGFR-2, VEGF receptor-2.

extracellular matrix (ECM), that can act similar to the endoderm in the embryos by secreting mesoderm inducing morphogens (Li et al., 2001). In fact around day 3, such external layer of cells starts expressing the endoderm-specific genes α-fetoprotein (α*fp*) and galactosamine epitopes until all the outer shell-like layer is formed by day 4 (Weitzer, 2006). Also during day 3, *T* which is responsible for inducing mesoderm formation in the embryo starts to be expressed, and is therefore an early mesodermal indicator in differentiating PSCs. After mesoderm specification it is possible to identify transient *Mesp-1* and *Mesp-2* expression around day 4 in EBs, which are markers of early cardiac commitment. Similarly, in the embryo its expression can be found transiently on the primitive streak prior to migration to the cranial region of the embryo where they become cardiac progenitors (Weitzer 2006; Schoenwolf, 2015). When occurs cardiac mesoderm specification, different early cardiac markers start to be expressed, such as *Nkx2.5, Isl1, Gata4* and *Tbx18* (reviewed in Meilhac et al., 2014) and such markers can be found to be expressed in differentiating PSCs. After plating the EBs on gelatin-coated culture dishes at day 6, they form cellular structures that within a couple of days start beating and are hence called beating *foci*. These structures are indicative of cardiac differentiation, and so it is that mature-specific cardiac markers like *cardiac Troponin T (TnT)* and myosin chains encoding genes (light and heavy chains) start to be expressed. Nonetheless, contrasting with the organized way gastrulation occurs in developing embryos, the way cells commit to cardiac fates in differentiating EBs is rather stochastic, and hence referred to spontaneous differentiation (Weitzer 2006).

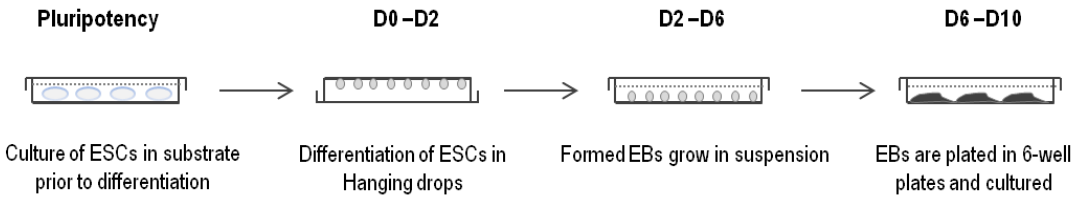


Figure 1.6 – Schematic representation of the hanging droplet method used to differentiate ESCs. Undifferentiated ESCs are maintained in culture for 2 passages prior to differentiation with medium supplemented with LIF. A cell suspension is prepared and cells are put to grow in droplets on an inverted bacterial dish for 2 days, and EBs grow in suspension four additional days. At day 6 of differentiation the formed EBs are plated in to gelatin-coated 6-well plates.

Similar to what was described *in vivo* ColI, ColIV and FN have important roles in promoting the differentiation of PSCs towards cardiac lineages, meaning that the ECM found in differentiating EBs plays also an important role. In fact at day 4 of differentiation, as the EBs commit to early cardiac lineages expressing *Mesp1, Isl1* and *Flk1*, it is shown that the ECM surrounding these cells is rich in ColIV (Schenke-Layland et al., 2011). In addition, culturing undifferentiated ESCs in ColIV substrates was shown to effectively enhance the fold number of *Flk1*+ cells, while posterior culturing *Flk1*+ cells in a FN substrate led to expression of α*Mhc*, an early indicator of mature cardiomyocytes. In the one hand, this strongly supports that ColIV is indeed required for the earlier stages of cardiac commitment, proliferation and maintenance of cardiac progenitors in an undifferentiated state. On the other hand, these data supports that once the cells are committed to a cardiac fate, they get sensitive to FN and respond to that stimulus by differentiating into cardiomyocytes (Schenke-Layland et al., 2011). Additionally, inhibiting the synthesis of ColI or blocking its interaction with β1 integrin receptors in differentiating PSCs leads to failure on cardiac lineages commitment and specification, showing it also has an important role on cardiac differentiation and hence is an important cardiac ECM component (Zeng et al., 2013). Of interest, our molecule of study is also a secreted protein and a prospective cardiac ECM component, and was already implied in cellular processes such as cell migration and proliferation in the heart, as it will be explained in detail in section 6. However, its role and function in cardiac differentiation is not well known.

As described before, there are mainly two pools of progenitor cells in the embryo that give rise to the heart - FHF and SHF. These progenitors have only recently been described in mouse ESCs. To do so, these authors generated a transgenic mouse with the red fluorescent protein dsRed under the control of an *Isl1*-dependent enhancer of the *Mef2c* gene whose expression is restricted to the SHF. This mouse line was then bred with another transgenic mouse line containing the enhanced green fluorescent protein (eGFP) controlled by the cardiac-specific *Nkx2.5* enhancer (Domian et al., 2009). With the expression of these fluorescent markers it was possible to isolate from developing hearts, cell populations expressing one, both or none of these reporters, corresponding to

populations of the FHF, SHF and non-cardiac cells as a control. After, blastocysts from these transgenic mice were isolated in order to establish a double transgenic ESC line model, which aimed to become a powerful tool to study the divergent origin of the different cardiac progenitor populations (Domian et al., 2009). By being a good model for embryonic cardiogenesis we chose to work with this double transgenic ESC line as they differentiate into some of the different cardiac populations existing in the developing heart.

5. Identification of genes (splice variants) involved in cardiogenesis: DNA microarrays Vs RNA sequencing

Growth factors or genes that regulate fate choices of differentiating PSCs can in fact offer a powerful tool to boost the production of a desired cell type (known examples in Figure 1.5). Since there are a lot of complex genetic networks interacting during cardiogenesis, the need to identify the factors that could ultimately be manipulated to increase in vivo or in vitro the number of cardiomyocytes or cardiac progenitors for heart regeneration applications, is a field worth exploring. Performing DNA microarrays has been the predominant technique used in the past decade to measure gene expression levels, to identify transcription factors' binding sites and to genotype single-nucleotide-polymorphisms (SNP), allowing biologists to explore vast amounts of complex digital data. According to this, in our lab was carried out a differential screening using Affymetrix GeneChip system technologies to enable us to identify and study genes expressed and involved in the correct development and differentiation of the vertebrate cardiac progenitor cell lineages. Indeed, this screening led to the identification of more than 700 transcripts differentially expressed in the heart forming regions (HFR), which after bioinformatical analysis and in vivo validation, this number was cut down to a few more than 150. Collagen and calcium-binding EGF domain-containing protein 1 (Ccbe1), our gene of study and interest, was identified among the new genes potentially expressed in the heart precursor cells (Bento et al., 2011). However DNA microarrays have been incredibly useful in a wide variety of applications, they can be particularly problematic for gene families and for genes with multiple splice variants (Bumgarner, 2013).

During pre-mRNA maturation to mRNA, alternative splice sites in one gene transcript may give rise to different protein isoforms that differ in their peptide sequence, having consequently different biological and chemical activities. And in fact many genes are known to have several splicing patterns or even thousands (Black, 2003). Now, as the costs of sequencing became cheaper, sequencing is a feasible unbiased approach to measure which nucleic acids are present in a given solution. In addition, it is also independent of our prior knowledge of which nucleic acids may be present and so it can detect closely related gene sequences, novel splice forms or RNA editing that may be missed due to cross hybridization on DNA microarrays (Bumgarner, 2013). RNA sequencing (or simply RNA-seq) is a whole genome transcriptome analysis that allows us to quantify the gene expression in a genome-wide fashion and in a given moment in time. Taking advantage of such technique presents an additional way to identify some of the yet unknown genes or growth factors that regulate cardiogenesis. Nonetheless, at the same time will allow us to identify if some of the already well known cardiac genetic players have different isoforms depending on the nature of the cardiac progenitor population, or if within the same population these change with time.

6. Ccbe1

6.1. Protein Structure and Identity

Ccbe1 encodes a 408 amino acid secreted protein with a calcium-binding EGF-like domain, which has 89% identity with the 406 amino acid human ortholog CCBE1. Even though it is still not possible to determine the correct structure of this protein, it possible to highlight some of the features it has by analyzing its primary amino acid sequence. By aligning mouse and human Ccbe1 sequences (Figure 1.7) and blasting against protein databases, Ccbe1 is shown to have 1 peptide signaling for secretion, 2 collagen domains, 1 calcium binding EGF-like domain, 1 predicted EGF-like domain and 2 glycosylation sites that could be responsible for further protein modifications.

6.2. Lymphangiogenesis and Hennekan Syndrome

During E9.5 in mouse development *Ccbe1* is shown to be expressed in tissues surrounding the anterior cardinal vein where *Prox1*+ lymphatic endothelial cells

Figure 1.7 – Mouse and Human Ccbe1 protein alignment. Protein domain sequences, predicted domain sequences, ion binding and glycosylation sites and disulfide bonding cysteines are highlighted in different color code. Signalling peptide for secretion (pink); Predicted EGF-like domain (blue); Calcium-binding EGF-like domain (yellow); Calcium (Ca^{2+}) binding sites (red arrows); Cysteine residues forming dissulfide bonds (green; cysteines bonding position: 138↔150; 146↔159; 161↔174); Glycosylation sites (orange); Triple helix collagen domains (brown). The alignment was performed using CLUSTAL O (1.2.1) multiple sequence alignment.

(LECs) are also present (Figure 1.9; Facucho-Oliveira et al., 2011; Bos et al., 2011). At this stage of development, *Ccbe1* appears to be mostly involved in the development of the lymphatic system. Indeed, *Ccbe1* loss-of-function in mice leads to prenatal death due to the loss of definitive lymphatic structures resulting in generalized lymphedema, which is a consequence of dramatic reduction of the number of *Prox1*+ and *Lyve1*+ LECs (Bos et al., 2011). Moreover, from E10.5 onwards *Ccbe1* is indeed required for the formation of LECs themselves, but additionally also for their consequent budding and migration from the anterior cardinal veins to give rise to the lymphatic vasculature, where in its absence LECs form dilated sprouts or bag-like sacs, which always remain connected with the cardinal vein (Bos et al., 2011; Hagerling et al., 2013). Interestingly, after being showed that administering CCBE1 together with VEGF-C could increase the yield in lymphangiogenesis in

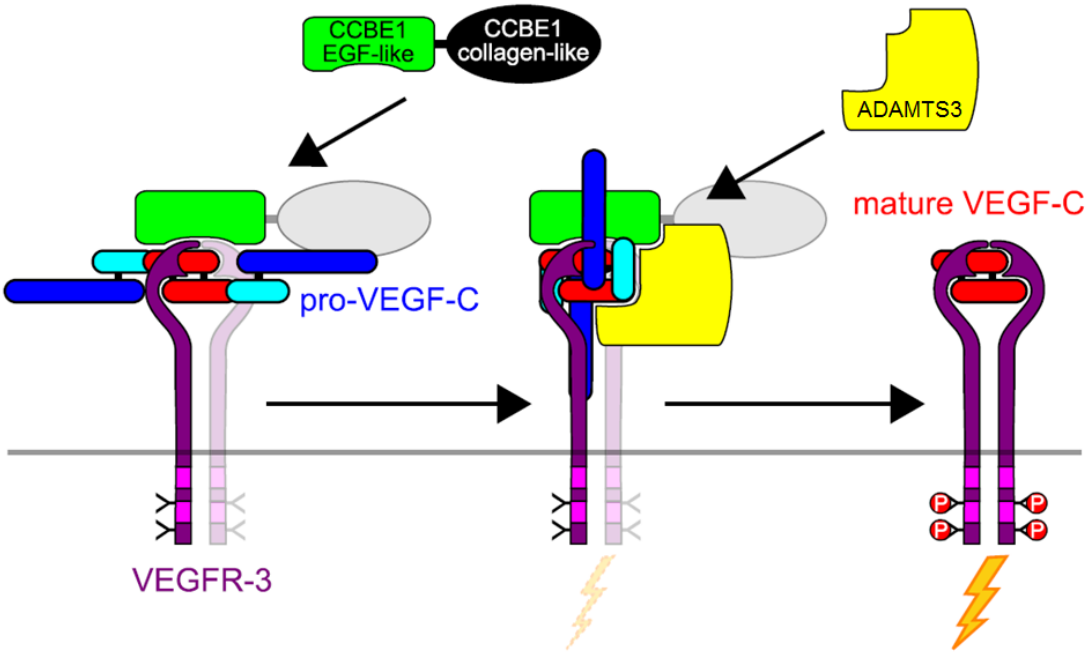


Figure 1.8 - Schematic view of the function of CCBE1 in lymphangiogenesis. Pro–VEGF-C binding to VEGFR-3 is assisted by the N-terminal domain of collagen- and calcium-binding epidermal growth factor domains 1 (CCBE1). Pro–VEGF-C is then proteolytically processed in situ by the metalloprotease ADAMTS3, and the mature VEGF-C activates VEGFR-3. Note that the transparently illustrated elements are hypothetical: VEGFR-3 could be either monomeric or dimeric during the initial binding of VEGF-C, and it is not known whether the removal of the C-terminal domain of CCBE1 is required for the CCBE1 function. Adapted from Jeltsch et al., 2014.

corneas, suggesting a role in VEGF-C modulation, it was shown that CCBE1 functions during lymphangiogenesis by helping in the maturation of the pro-VEGF-C through the interaction with the metalloprotease ADAMTS3 (Figure 1.8; Bos et al., 2011; Jeltsch et al., 2014). Likewise, *ccbe1* in zebrafish has also been associated with the development of the lymphatic vasculature and venous sprouting using the same mechanism, even though due to the lack of early lymphatic markers it is not clear whether *ccbe1* is also needed for lymphatic precursor cells fate commitment (Hogan et al., 2009; Le Guen et al., 2013; Astin et al., 2014). On the other hand overexpression of mature vegf-c in zebrafish is capable of rescuing the deficits found in lymphangiogenesis in the absence of ccbe1, shedding light on the conserved role ccbe1 has in modulating the vegf-c signaling (Le Guen et al., 2013). While it was firstly suggested that the CCBE1 EGF-like domain is essential for the maturation of VEGF-C (Bos et al., 2011; Jeltsch et al., 2014), more recently in an attempt to perform additional functional domain analysis of CCBE1, it was possible to show that its collagen domains

are also essential. Indeed, the collagen domains of CCBE1 were found to function in activating the VEGF-C both *in vivo* and *in vitro* models, rather than binding to the ECM, like many collagen proteins (Roukens et al., 2015). However, as an ECM protein itself there are also proposed interactions of CCBE1 with other ECM components such as vitronectin, in order to localize it *in vivo* in tissues were it mediates activation of growth factors such as VEGF-C, as a cue for cellular processes like migration in the case of LECs in lymphangiogensis (Jeltsch et al., 2014). Indeed, a truncated fraction comprising the EGF-like domain of CCBE1 was found to bind to ECM proteins vitronectin, ColI and ColV (Bos et al., 2011).

In humans, mutations in *CCBE1* have been associated with Hennekam's syndrome (HS), a rare autossomal recessive syndrome characterized by defective lymphatic development. Of the several mutations identified so far, the more commons are found to be in the N-terminal portion of the protein, which corresponds to the EGF-like domain. However some others also have the mutation on the C-terminal portion, corresponding to the collagen domains (Alders et al., 2009). The patients present limb lymphendema, lymphangiectasias, mental retardation, and also unusual facial characteristics (Alders et al., 2009; Connel et al., 2010; Frosk et al., 2015). It is likely that all these patients are hypomorphs, as even though mutation occurs, the protein might still preserve some minimal function, since the total absence of *Ccbe1* in mice is rather lethal (Bos et al., 2011). So altogether, it is clear the indisputable key role of *CCBE1* in lymphagiogenesis, and the way CCBE1 acts is embryonic process is likely by providing positional information for VEGF-C signaling, which orchestrates the migration of LECs (Roukens et al., 2015).

6.3. Carcinogenesis, proliferation and migration

In a thorough expression analysis of the 18q21-qter chromosomal region, where gene losses are frequent in breast cancer, primary breast cancer cell lines were found to have strong downregulation on the expression of *CCBE1* (Yamamoto and Yamamoto, 2007). Down-regulation of *CCBE1* was additionally identified in ovarian cancer cell lines and primary carcinomas (Barton et al., 2010). Because the domains of CCBE1 are found in some of the extracellular matrix proteins, the loss of CCBE1 protein expression could possibly result in changes in

cellular characteristics, such as adhesion and motility. Indeed, siRNA-mediated knockdown of *CCBE1* in ovarian cancer cell lines enhanced cell migration; on the other hand, re-expressing *CCBE1* in the same cell lines was sufficient to reduce cell migration and survival. Hence, in carcinogenesis the way *CCBE1* loss of expression may act is by enhancing migration and cell survival (Yamamoto and Yamamoto, 2007; Barton et al., 2010). Such findings are interesting, as they suggest CCBE1 to be involved in metastization by enhancing cell migration. However, the mechanism through which CCBE1 might be involved in migration of carcinogenic cells is likely different from the mechanism involved in the impaired migration of LECs in lymphangiogenesis. While in lymphangiogenesis the lack of Ccbe1 is rather involved indirectly in migration, as it impedes the maturation of the Vegf-c as the leading migration cue for the LECs, on carcinogenesis its involvement seems related to a more direct mechanism. Cell migration experiments in *Ccbe1* knockout (*Ccbe1*$^{-/-}$) mouse embryonic fibroblasts (MEFs) were carried out by our laboratory to better understand its role in migration. Interestingly, it was shown that comparing to *wild-type* (Wt) MEFs, *Ccbe1*$^{-/-}$ MEFs were found to have enhanced migration capacity, similar to what happens in carcinogenesis (Perestrelo et al., *in preparation*). This recently new data is in agreement with another previous report from our lab where modulating *Ccbe1* expression during chick heart development leads to abnormal cell migration during primitive heart tube fusion (described in section 1.6.4; Furtado et al., 2014). Since these cardiovascular progenitors rely on ECM cues for migration, and it was also shown that CCBE1 interacts with ECM proteins (Bos et al., 2011), it was hypothesized a mechanism where CCBE1 interacts with the ECM providing or inhibiting the migratory stimuli. In fact, *Ccbe1*$^{-/-}$ MEFs have higher expression of proteolytic enzymes, which can be activated through ECM components and be responsible for a higher ECM degradation and hence a higher migration capacity. Therefore our data fits a model where in contexts where the ECM provides the migratory stimuli, such as what happens in carcinogenesis, CCBE1 seems to act as an antagonist (*Unpublished data*). However, further studies are still required to identify what molecules interact with CCBE1 or are involved in such migration process.

While in one hand cell migration is enhanced in *Ccbe1*$^{-/-}$ MEFs when compared to Wt MEFs, survival assays show *Ccbe1*$^{-/-}$ MEFs to have a decreased cell proliferation (*Unpublished data*). In agreement with this data it had been described previously that the loss of *Ccbe1* could also lead to a reduced number of *Lyve1+/Prox1+* lymphatic progenitor cells (Bos et al., 2011). So altogether it is likely that apart from its role in cell migration, *Ccbe1* is also somehow involved in cell proliferation.

6.4. Cardiogenesis

Prior to the lymphatic system development and to its expression on lymphangiogenic regions in the embryo, *Ccbe1* has also expression in cardiogenic embryonic regions. At first, expression analysis carried out by our lab showed that *Ccbe1* is expressed in the early cardiac progenitors of the two bilateral cardiogenic fields at E7.0, and in the cardiogenic mesoderm from E7.5 - E8.0 (Figure 1.9 A-D). At E8.25 (Figure 1.9 E-F) *Ccbe1* showed persistently expression in the pericardium and transiently expression in the myocardium of the primitive heart tube. Later its expression is detected in the proepicardium (Figure 1.9 G-H). Within these regions its expression is detected in all the FHF, SHF and proepicardium cells during early mouse heart organogenesis from 7.0 – 8.75 (Facucho-Oliveira et al., 2011) raising the possibility that Ccbe1 may play a role in cardiac development. Interestingly, prior work has also shown that *Ccbe1* is expressed in the pericardium at E11.0 and E12.5, even though at these stages its expression is more related with lymphangiogenesis as described in section 6.2 (Bos et al., 2011). In developing chick embryos *Ccbe1* was similarly found to be expressed in FHF and SHF populations during early heart development (Furtado et al., 2014).

In HS patients, some of the patients also present congenital heart defects including hypertrophic cardiomyopathy and ventricular septal defects (Alders et al., 2009; Connell et al., 2010; Frosk et al., 2015). Such malformations have not been taken into account a probable involvement of *Ccbe1* in heart development. Intriguingly, diseases that are related to defects in the development of the anterior pole of the heart may be accompanied by abnormalities of the skeletal muscles of the head, and are likely to arise from a defect in a common progenitor of the SHF and other derivatives (Figure 1.3; Schoenwolf, 2015).

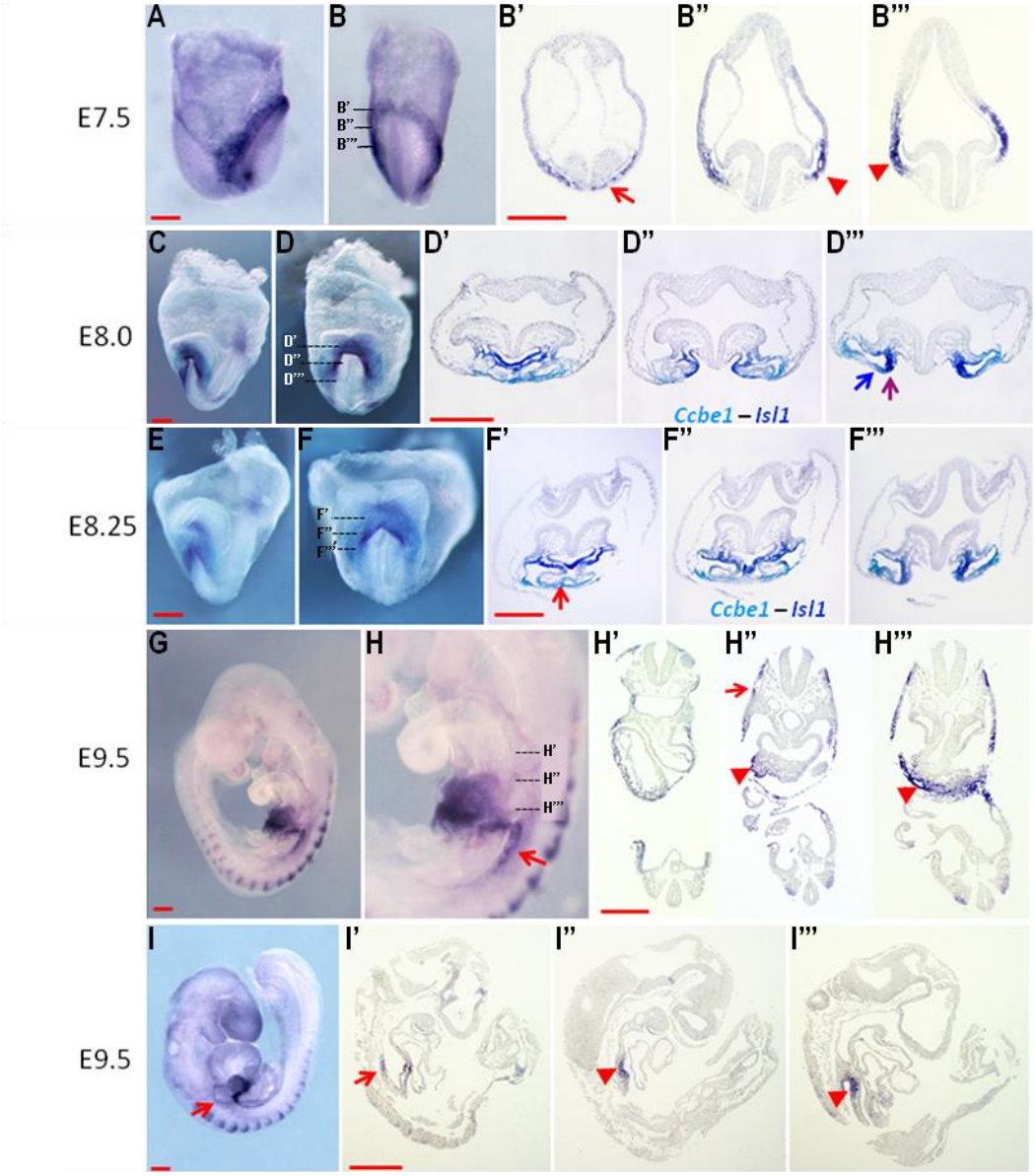


Figure 1.9 – *Ccbe1* expression pattern in cardiogenic regions during mouse embryogenesis. (A, B) A series of transverse sections extending from the head fold level towards to a more caudal positions showed that *Ccbe1* is expressed in the cardiogenic plate (arrow in B') and in the mesothelial precursors of the intra-embryonic coelomic cavity (arrow in B'', B'''); **(C, D)** Double WISH performed to detect *Ccbe1* mRNA (light blue) and Isl1 mRNA (dark blue) at E8.0 demonstrated that *Ccbe1* is mainly expressed in FHF progenitors; although restricted overlap of *Ccbe1* and *Isl1* staining confirms that *Ccbe1* is also expressed in the SHF (C and blue arrow in D'''). **(E, F)** At E8.25, *Ccbe1* continues to delineate the pericardium cavity; lateral view, (E); anterior view (F). *Ccbe1* mRNA was detected in the ventral mesothelium of the pericardium (F'), heart tube tissue adjacent to the ventral pericardium (arrow F') and mesoderm lining the intra-embryonic coelomic cavity (F'''). Double WISH reveals that *Ccbe1* mRNA (light blue) continues to be partially colocalized with *Isl1* (dark blue) at the level of the pharyngeal mesoderm (F'); **(G, H)** Lateral view of E9.5 embryo showed that *Ccbe1* is expressed in the proepicardium, in the anterior cardinal veins (arrow) and in the somites (G, H). Transverse sections show that *Ccbe1* expression in the heart is residual or non-existent (H') but rather highly expressed in the proepicardium

(arrowhead; H'', H'''), in the anterior cardinal vein (arrow; H'', H''') and dermomyotome of the cervical somites (H'', H'''). Sagital sections further demonstrate that *Ccbe1* staining is located in the vicinity of the anterior cardinal vein (arrow) which during mouse development gives rise to important veins of the cardiovascular and lymphatic systems (I). Scale bars, 200 mm. Adapted from Facucho-Oliveira et al., 2011.

Seemingly, the unusual facial characteristics of the HS patients can strongly fit this model, since *Ccbe1* expression is found to be enriched in SHF progenitors of the studied animal models (Facucho-Oliveira et al., 2011; Furtado et al.,2014). in animal models, two recent studies, including one where an alternative *Ccbe1 null* mutant line with a mixedgenetic background that survives until birth was used, suggest that the heart of *Ccbe1* mutant mice develops normally (Burger et al., 2015; Jakus et al., 2014). However, in embryonic cardiogenesis animal models, a recent report by our lab has shown the need for *Ccbe1* in early chick heart development where both gain- and loss-of-function approaches led to incorrect fusion of the bilateral heart fields in order to form the primitive cardiac tube, which resulted in *cardia bifida* and in other cardiac malformations (Furtado et al., 2014). Additionally, it was also identified that *Ccbe1* loss-of-function in mice is responsible for thinner/hypoplasical myocardial walls, a phenotype more evident on the compact layer of the developing right ventricle, which is related to the underdevelopment of the coronary vasculature (Pereira et al., *in preparation*). Additionally, co-culture of ESCs in *Ccbe1*$^{-/-}$ MEFs leads to decreased expression of cardiac marker genes, suggesting that secreted Ccbe1 by the fibroblasts is important for proper cardiac differentiation of ESCs (*Unpublished data*). Even though the exact molecular mechanisms by which this happens require further investigation, all of these data gathered so far strongly suggest that *Ccbe1* plays an important role in heart organogenesis and in cardiac differentiation of ESCs.

7. Objectives

Despite the increasing evidence of a potential important role of *Ccbe1* during cardiogenesis, its role in cardiac differentiation and development has not been further investigated. In this work I mainly aimed to perform an integrated study on the role of *Ccbe1* in cardiac differentiation of ESCs in order to gather

knowledge about its involvement in the early cardiogenesis. To achieve this goal we used a double transgenic mouse ESC line which allows us to isolate different ESC-derived early cardiac progenitors and study *Ccbe1* expression and its involvement in this stage of cardiogenesis. Also, by silencing the expression of *Ccbe1* through short hairpin RNAs (shRNA) to generate stable knockdown ESC lines we were help to unveil on what biological pathways *Ccbe1* plays an important role besides cardiogenesis. Additionally, gain-of-function approaches were planned by producing and supplementing the ESCs differentiation medium with human recombinant CCBE1 to detect whether this protein can act as a guiding molecule in cardiac differentiation, with the ultimate goal of generating large amounts of cardiomyocytes to be used in regenerative applications.

Furthermore, for the systematic study of alternative splicing, we aimed in firstly identify splicing events in mESCs and in cardiac progenitors on a genome-wide level using next-generation RNA-seq technology. After, we analyzed how splicing events can change the structure and function of signaling pathways during the differentiation of embryonic stem cells to cardiomyocytes. We looked closer at how different isoforms of well described cardiac genes can characterize different populations of cardiac progenitors: To achieve this task we isolated the whole transcriptome of different populations of cardiac progenitors in different time points of cardiac differentiation. New genes potentially involved in cardiogenesis were especially taken in account for this analysis. The main tools used in this particular systems biology task involved a combined application of various computational and experimental methods required to cope with the complexity of the studied mechanisms.

The overall work accomplished throughout this project will allow us to better understand the genetic mechanisms underlying cardiogenesis for the progress of the science in the cardiovascular field and translational medicine.

Chapter II

Materials and Methods

2.1. Culture of mouse embryonic fibroblasts (MEFs)

Mouse embryonic fibroblasts (MEFs) were cultured in high glucose Dulbecco's modified eagle's medium (DMEM; Sigma, Poole, Dorset, UK) supplemented with 10% of fetal bovine serum (FBS), 2 mM of L-glutamine (Invitrogen Life Technologies, Paisley, UK), 1% non-essential amino acids (NEAA, Invitrogen Life Technologies) and 1% penicillin/streptomycin solution (Invitrogen Life Technologies). MEFs were grown in 75 cm2 flasks (Sarstedt, Leicester, UK) with 10 ml of media at 37°C/5% CO2 and culture media replaced every other day to ensure optimal growth conditions. MEF feeder layers for culture of mESCs were prepared using mitomycin C (Sigma). Confluent 75 cm2 flasks (Sarstedt, Leicester, UK) of MEFs were incubated with media containing 10 µg/ml of mitomycin C for 2 hours at 37ºC/5% CO2 and washed 3 times with 10 ml of PBS. Inactivated MEFs were cultured overnight or frozen before use as feeder cells for undifferentiated mESCs.

2.2. Culture of mouse ESCs

Mouse *Nkx2.5*-eGFP/SHF-dsRed (RG) ESCs were kindly provided by MD Ibrahim Domian. RG ESCs were cultured in knockout DMEM (Sigma) with 15% ESC screened FBS (Hyclone, Utah, US), 1% penicillin/streptomycin solution (Invitrogen Life Technologies), 2 mM L-glutamine (Invitrogen Life Technologies), 1% NEAA (Invitrogen Life Technologies), 0.1 mM-mercaptoethanol (Sigma) and 1000 U/ml leukaemia inhibitory factor (LIF; Chemicon, Temecula, Ca, USA). RG ESCs were cultured in 60mm ø tissue culture specific plates with 6 mL of ESC media at 37ºC/5% CO2 and passaged when the ESC colonies had about 50% of confluence and before semiconfluency, every 2 or 3 days. The culture medium was replaced daily.

2.3. Differentiation of ESCs

RG ESCs and *Ccbe1* knockdown (KD) ESCs lines were differentiated using the hanging droplet method previously described in the literature (Keller, 2005).

Undifferentiated ESCs were washed with a 1xPBS solution, then incubated with 1xTrypsin for 2 minutes, dissociated into a single cell suspension, centrifuged and resuspended in fresh ESC medium without LIF. Cells were counted on either a hemocytometer or on an improved Neubauer chamber (VWR), and then a cell suspension was prepared with a final concentration of $2{,}2x10^4$ cells/mL. Cells were plated as 20 µL drops (approximately 440 cells) onto the base of an anti-adherent Petri dish which was then inverted into the lid. The cells were cultured in hanging droplets for 48 hours (days 1 and 2). The Petri dishes were inverted back to the original position and 10 mL of ESC media was added to the plates. EBs were cultured in suspension for 4 days (days 3 to 6) before being plated onto gelatin (0.1%) coated wells and cultured up to day 10 of differentiation. Culture medium was replaced every 2 days.

2.4. Fluorescent-activated cell sorting (FACS)

EBs were gently dissociated into a single cell suspension using 1x trypsin and resuspended in 1xPBS. RG ESCs ran through a Becton Dickinson FACSAria II (BD Biosciences, Erembodegem, Belgium) using the FACSDiva 6.1.3 software (BD Biosciences, Erembodegem, Belgium). Flow cytometer was carried out upon excitation with the blue laser (488 nm) being the emission signals measured in the FL1 channel (530/30 nm; Green) and in the FL2 channel (585/42 nm; Red). Sorting was performed with a 100 µm nozzle, using a “purity” precision mode and a flow rate with a maximum of 10,000 events per second. In total, 4 different cell populations were isolated: $GFP^+/dsRed^-$ FHF progenitor (here on named G^+R^-), $GFP^+/dsRed^+$ and the $GFP^-/dsRed^+$ SHF progenitors (here on named G^+R^+ and G^-R^+) and the $GFP^-/dsRed^-$ control cells (G^-R^-). Cells were collected to anti-adherent collection tubes with fresh ESCs medium without LIF. To pellet the cells for RNA isolation, the collection tubes were centrifuged at 1000rpm and the supernatant discarded.

2.5. RNA extraction

Total RNA was extracted from undifferentiated and spontaneously differentiated (days 0-10) ESCs. Total RNA was extracted using the TRizol Reagent (Invitrogen) according to the manufacturer's instructions with the following modifications; cell pellets were washed with ice cold PBS before being lysed using 400 µL of Trizol reagent. The cell lysate was resuspended with a pipette several times, incubated at RT for 5 min. Chloroform (80 µL) was added to the lysate, the tube was mixed by inversion for about 15 seconds and centrifuged at 13,000 rpm for 10 min. The RNA containing aqueous phase was then pipette into a new microtube. 10ng of RNAse-free glycogen was added to each sample as a carrier, to enhance the total amount of RNA extracted. Then 250 µL of ice cold isopropanol was mixed into the samples. The solution was then incubated at 30°C for 5 min before being centrifuged at 13,000 rpm for 30 min. The supernatant was then discarded and the pellet washed with 500 µL of icecold 75% ethanol. After centrifugation, the supernatant was discarded and the RNA pellet allowed to air-dry for approximately 10 min until a tiny meniscus of solution was left around the pellet. Protocol was performed on ice during at all times and centrifugations performed at 4°C. RNA was then resuspended in 25 µL of DNAse mix (Ambion, Life Technologies) and incubated at 37ºC for 30 minutes. After, it was added 5 µL of DNAse inactivator, samples were centrifuged at 15ºC, 13.000rpm for 5 minutes, and around 22 µL of RNA was collected to new collection tubes. RNA concentration was determined using a NanoDrop® ND- 2000c Spectrophotometer (NanoDrop Technologies).

2.6. RNA isolation for RNA Sequencing

After population isolation on the FACS, the cells were pelleted by centrifugation at 1000 rpm and total RNA was extracted with QIAGEN's RNeasy Micro kit, according to the manufacturer's protocol. After RNA extraction, to inquiry its integrity, all the samples underwent an automated electrophoresis with Experion^{TM} equipment from BIO-RAD.

2.7. cDNA Synthesis

The RNA was reverse transcribed using Thermo Scientific RevertAid Reverse Transcriptase (Fermentas). Reactions contained 1000 ng/µl of RNA, 0.5 µg of Oligo (dT) primer, 4 µl of 5x Reaction Buffer, 1 U/µl of RiboLock RNase Inhibitor, 1 mM of dNTP mixture, 10 U/µl of RevertAid Reverse Transcriptase and nuclease-free water (Sigma) up to 20 µl. Reactions were performed in two steps according to the manufacturer's instructions. RNA and Oligo Oligo (dT) primers were firstly incubated at 65°C for 5 min. The remaining components were then added to the reaction and incubated at 42°C for 60 min followed by 70°C for 10 minutes to denature the Reverse Transcriptase enzyme.

2.8. Quantitative PCR

The quantitative PCR (qPCR) reaction contained 2 µL of cDNA, 7.5 µL of 2x SsoFast Evagreen Mix (Bio-Rad), 0.33 µM of each of the forward and reverse primers (Invitrogen Life Technologies) and ultrapure H2O up to 15 µL. Reactions were performed in a CFX real-time PCR machine (Bio-Rad). Initial denaturation was performed at 95°C for 1minutes, followed by 40 cycles of: denaturation at 95°C for 15 seconds; annealing for 15 seconds (temperature was specific for each gene primers as specified in Table 1); and extension at 72°C for 15 seconds. Data were acquired in the FAM/SYBR channel during the extension phase. The PCR product was then denaturated by progressively increasing the temperature from 62ºC to 99ºC (0,5ºC every 5 sec). For the standards, a series of 10-fold dilutions (2 x 10-4 ng/µL to 2 x 10-9 ng/µL) of the target-specific PCR product were generated. Reactions with an efficiency comprised between 90-110%, Pearson correlation coefficient close to 1 and melting curve without non-specific amplifications were considered for current application. Analyses were performed using the Bio-Rad CFX Manager software (version 2.1; Bio-Rad). The relative level of expression of each target gene was calculated using ddCq method (Bustin, 2000). Data is expressed as mean ± standard deviation of the mean (SEM).

Table 1. Primers used for Real-time PCR (qPCR)		
Gene	Forward (5' – 3')	Reverse (5' – 3')
αMhc	GATGGCACAGAAGATGCTGA	CTGCCCCTTGGTGACATACT
Ccbe1	GACACACGTGGACCTACCGAG	CCGTGCACTGCTGTTCACAGG
cTnt	GGAAATCCAAGATCACTGCCTCC	GGGCACTGAGGGACAGACCA
Esm1	TCACATACACGCCACAAAACAAC	TTCCGCAAAGACCATGCAT
Gapdh	GGGAAGCCCATCACCATCTTC	AGAGGGGCCATCCACAGTCT
Isl1	CCTGTGTGTTGGTTGCGGCA	GGGCACGCATCACGAAGTCG
Nkx2.5	CCACTCTCTGCTACCCACCT	CCAGGTTCAGGATGTCTTTGA
Mesp-1	TGTACGCAGAAACAGCATCC	TTGTCCCCTCCACTCTTCAG
Vegf-C	AACGTGTCCAAGAAATCAGCC	AGTCCTCTCCCGCAGTAATCC
Vegfr3	CAGACAGACAGCGGGATGGTGC	AGGCTGTAGTGGGGGTGGGACA
Pgk1	ATGGATGAGGTGGTGAAAGC	CAGTGCTCACATGGCTGACT
Prox1	GCCATCTTCAAAAGCTCGTC	CTGGGCCAATTATCACCAGT

2.9. Production of lentiviral vectors

Commercially available shRNAs targeting codifying regions of *Ccbe1* were purchased (Sigma) and used to produce lentiviral vectors to generate *Ccbe1* knockdown (KD) mouse ESCs lines. The shRNAs we used were named based on its respective commercial reference (Table 2.1). Lentiviral particules were produced using HEK 293T cells for the different shRNA plasmids.

shRNA reference	Sequence
sh91779	CCGGGAGACCGTACTGTCTGGATATCTCGAGATA TCCAGACAGTACGGTCTCTTTTTG
sh91780	CCGGCCCTGACCTGTCTCATATTAACTCGAGTTA ATATGAGACAGGTCAGGGTTTTTG
sh91781	CCGGCAAGTATGTCAATGGTGACAACTCGAGTTG TCACCATTGACATACTTGTTTTTG

Table 2.1. Sequences of the shRNAs plasmids targeting codifying regions of *Ccbe1* used to generate the knockdown cell lines

2.10. Generation of Ccbe1 knockdown mESCs lines

RG ESCs lines expressing shRNAs against *Ccbe1* mRNA were generated using lentiviral particles containing *Ccbe1*-shRNA plasmids and shRNA control plasmids, whose function had already been validated. RG ESCs were cultured in pluripotency for 2 passages prior to lentiviral infection on 60mm ø tissue culture plates. On the day of the infection the culture medium was replaced by

fresh ESCs medium LIF, the viral particles and 4 ug/mL of polybrene, composing a final volume of 3 mL. Cells were left to incubate for up to 6 hours, where after it were added additional 3 mL of fresh ESCs medium containing polybrene. After 24 hours the medium containing the viral particles was withdrawn from the plates and changed to fresh ESCs medium containing. Clonal selection started 48 hours after the infection protocol. To start the selection, puromycin was added to the ESCs medium at a final concentration of 10 mg/mL and lasted for an additional 7 days. Newly generated clones were picked individually, and an initial *Ccbe1* expression analysis was carried out through semi-quantitative PCR to assess the knockdown (KD) levels of *Ccbe1*. The clones that were confirmed to have *Ccbe1* KD were then expanded and, two *Ccbe1* KD ESCs lines and one scrambled control ESCs line (Sh-control) were established.

2.11. Immunofluorescence in cryosections

Cultured EBs were embedded in optimal cutting temperature (OCT), frozen at -80 ºC, sectioned with a cryostat microtome (MARCA) (10 µm thick slices) onto glass slides. Frozen sections were fixed with 1% paraformaldehyde and blocked with 2% bovine serum albumin and 0.02% Tween20 in PBS. Samples were incubated with polyclonal rabbit anti- activate caspase-3 primary antibody (R&D Systems, AF835; 1:500 dilution) followed by incubation with goat anti-rabbit IgG antibody Alexa 488 (H+L; Invitrogen, A-11008; dilution 1:1000). Slides were mounted with Mowiol containing DAPI.

2.12. Immunolabelling

ESCs culture media was discarded and cells washed with 500 µL 1xPBS and dissociated with 200 µL 1x Trysin at 37ºC for 5 minutes. Then 500 µL of ESCs culture media was added, cells centrifuged for 5 minutes at 1000 rpm at RT and the supernatant discarded. Cells were resuspended in 500 µL ice-cold 1x PBS/3%FBS/10mM NaN_3 freshly prepared, after which were centrifuged at

1800 rpm for 15 minutes at 4ºC. The supernatant was carefully removed with a P1000 micropipette, and the cells resuspended and fixed with 200µL 0,5% PFA in 1xPBS for 15 minutes at RT. Additionally, for washing 500 µL of ice-cold 1xPBS/0,5% BSA/0,1% saponin were added and cells centrifuged at 1800 rpm for 15 minutes at 4ºC. This washing step was repeated twice after supernatant removal. Further on, cells were resuspended in 100 µL of 1xPBS/0,5% BSA/0,1% saponin containing the primary antibody and left incubating for 1 hour at 4ºC. Washing steps were repeated twice once more. Then cells were resuspended in 100 µL of 1xPBS/0,5% BSA/0,1% saponin containing the secondary antibody and left for incubation in the dark for 1 hour at RT. Cells were washed one final time as described above and prior to FACS analysis resuspended in 500 µL ice-cold 1x PBS. The dilutions used for each antibody are listed in Table 3.

2.13. Methylene Blue Diffusion Assay

Cultured EBs were collected and washed in PBS before being treated with 1 mg/mL solution of methylene blue in PBS for 10 minutes and then washed with PBS. Embryoid bodies were embedded in optimal cutting temperature (OCT), frozen at -80 ºC, sectioned with a cryostat microtome (MARCA) (10 µm thick slices) onto glass slides and mounted directly with DPX mountant for histological observation.

2.14. Cell Proliferation Assay with Dye eFluor® 670

Cultured ESCs were resuspended in PBS and counted. Cell Proliferation Dye eFluor® 670 (eBioscience) in PBS was added drop by drop to the cell suspension and incubated at 37 ºC in the dark. Labeling was stopped by adding cold growth medium and incubation on ice. Lastly, ESCs were washed several times with growth medium. ESCs were then cultured as embryoid bodies in hanging drops as described above.. Flow cytometry was carried out upon excitation with the red laser (633 nm) being the emission signals measured in

the FL4 channel (661/16 nm). The fluorescence intensity of the dye was measured along the time of culture at days 0, 2, 3, 4, 5 and 6 by flow cytometry and data analysed using the BD FACSDivaTM software (version 6.1.3, BD Biosciences). Graphics generated using FlowJo X 10.0.7r2, Tree Star, Inc. A total of 50.000 events were acquired for each ESC cell line.

2.15. Production of Recombinant human CCBE1 protein

HEK 293T cells grown in suspension in spinner flasks were transfected using standard Calcium Phosphate Method with the CCBE1 expression vector containing the full-length CCBE1 coding sequence followed by a 6x Histidine tag at the C-terminal. After 4 days of incubation, the conditioned media containing CCBE1 recombinant protein was loaded onto a 1 mL Histrap column (GE Healthcare). The bounded protein was eluted by discontinuous imidazole gradient and the fractions containing CCBE1 (more than 90% pure) were dialyzed overnight into 25 mM Tris/0.15 M NaCl/2m M CaCl2 pH7.5, then frozen in liquid nitrogen and kept at -80ºC until further use.

2.16. Statistics

Statistical analysis was performed using GraphPad Prism version 5.00 for Windows, GraphPad Software, San Diego California USA.

RESULTS

Chapter III

Ccbe1 is required for normal cardiac-specification and proliferation in differentiating mouse embryonic stem cells

Tiago Justo[1,2,5], Paulo N G Pereira[2,5], João Facucho-Oliveira[1,2], José M Inácio[5], Ibrahim Domian[3,4] and José António Belo[5*]

[1] Regenerative Medicine Program, Biomedical and Medicine Sciences Department and

[2] Institute for Biotechnology and Bioengineering, Center for Molecular e **Structural** Biomedicine, University of Algarve, Campus de Gambelas, 8005-139 Faro, Portugal

[3] PhD Program in Biomedical Sciences, UALG, Portugal.

[4] Cardiovascular Research Center, Massachusetts General Hospital, Boston MA 02114–2790, USA and Harvard Stem Cell Institute, Cambridge MA 02138, USA

[5] Stem Cells and Development Laboratory, CEDOC, NOVA Medical School / Faculdade de Ciências Médicas, Universidade Nova de Lisboa, 1150-082 Lisboa, Portugal

Manuscript submitted to: *Stem Cells and Development*

Author's contribution:

The majority of the work was performed by Tiago Justo with the exception of the isolation of cardiac progenitors at E.9.5 from mice embryos performed by Ibrahim Domian.

3.1. SUMMARY

Understanding the molecular pathways regulating cardiogenesis is crucial for early diagnosis of heart diseases and to improve therapeutic approaches. During normal mammalian cardiac development, collagen and calcium-binding EGF domain-1 (*Ccbe1*) is expressed in the first and second heart field progenitors as well as in the proepicardium, but its role in early cardiac commitment remained unknown. We demonstrate that during mouse embryonic stem cell (ESC) differentiation *Ccbe1* is upregulated upon emergence of *Isl1*- and *Nkx2.5*- positive cardiac progenitors. Furthermore, *Ccbe1* is markedly enriched in *Isl1*-positive cardiac progenitors derived from embryonic stem cell differentiating *in vitro* as well as progenitors isolated from embryos developing in vivo. *Ccbe1* knockdown during ESC differentiation resulted in impaired development of cardiac mesoderm. In addition, knockdown of *Ccbe1* leads to reduced cell proliferation resulting in smaller embryoid bodies. Accordingly, blockade of CCBE1 with antibody during ESCs differentiation causes the same defects. Interestingly, these defects seem to be independent of the role of CCBE1 in the regulation of VEGF-C signaling. Collectively, our results indicate that CCBE1 is essential for the formation of cardiac mesoderm and proliferation of differentiating mouse ESCs.

Keywords: Cardiac differentiation; Cell Proliferation; ESCs; *Ccbe1*; Cardiac Mesoderm.

3.2. INTRODUCTION

Identification of genes and the study of their role in cardiogenesis are important to elucidate the molecular events regulating cardiomyocyte lineage commitment. This is critical for the control of cardiac commitment from different stem cell sources and the use of mature cardiac cells in the context of regenerative medicine. In a differential screen designed to identify novel genes required for the correct development of the heart precursor lineages Bento et al. (2011), we identified *CCBE1*, a gene coding for a secreted protein that contains collagen domains and a calcium binding EGF-like domain. Expression analysis showed that *Ccbe1* is expressed in precursors of the first heart field (FHF), secondary heart field (SHF), and proepicardium in mice between embryonic day (E) 7.0 to E9.5 (Facucho-Oliveira et al., 2011). Similarly, *CCBE1* was similarly found to be expressed in FHF and SHF populations during early chick cardiac development (Furtado et al., 2014). These findings implicate CCBE1 in the control of early cardiac commitment, but its function in this context remains elusive. Previous work has also shown that *Ccbe1* is expressed in the pericardium between E11.0 and E12.5 (Bos et al., 2011), however, at these stages *Ccbe1* is deeply involved in the development of the lymphatic system. Indeed, *Ccbe1* loss-of-function in mice leads to prenatal death due to defective lymphatic vasculature (Bos et al., 2011). *Ccbe1* is required for the budding and migration of lymphatic endothelial cells (LECs) from the anterior cardinal veins to give rise to the lymphatic vasculature (Bos et al., 2011; Hagerling et al., 2013). Absence of proper lymphatic vessels results in generalized tissue edema by E14.5 and the mutants die shortly after. In a more recent report, it was demonstrated that absence of the collagen domains from CCBE1 in mice fully phenocopies the *null* mutant (Roukens et al., 2015). The mode of action of CCBE1 involves the recruitment of the metalloprotease ADAMTS3 extracellularly to promote the conversion of immature (Pro-)VEGF-C into its mature and fully active pro-lymphangiogenic form (Jeltsch et al., 2014; Le Guen et al., 2014).

In humans, mutations in CCBE1 have been associated with Hennekam syndrome (HS), a disorder characterized by abnormal lymphatic system

development. Interestingly, some patients also present with congenital heart defects including hypertrophic cardiomyopathy and ventricular septal defects (Alders et al., 2009; Connell et al., 2010; Connell et al., 2012), consistent with a role of CCBE1 during heart formation. Although two recent studies suggest that cardiac development is normal in *Ccbe1* mutant mice (Burger et al., 2015; Jakus et al., 2014), we showed that *CCBE1* is required for the migration of the cardiac precursor cells to form the heart tube during chicken heart development (Furtado et al., 2014). Modulation of *CCBE1* levels in the chick embryos leads to cardia bifida when the cardiac fields are exposed to high levels of *CCBE1.* Conversely, exposure to low levels of *CCBE1* result in incorrect fusion of the bilateral cardiac fields to form the heart tube.. Therefore, given those opposing observations about the role of CCBE1 in the development of the heart from different species, we sought to study the role of CCBE1 during cardiogenesis using an established model of cardiac differentiation *in vitro* using mouse ESCs.

Here, we analyzed the effect of *Ccbe1* loss-of-function during differentiation of mouse ESCs and identified a role in early cardiac mesoderm commitment as well as in cell proliferation. In addition, we examined *Ccbe1* expression in differentiating mouse ESCs and confirmed its expression in isolated cardiac progenitor populations derived from ESCs.

3.3. RESULTS

High *Ccbe1* expression coincides with the appearance of cardiac progenitors

To evaluate *Ccbe1* expression during cardiac differentiation, we exploited the double transgenic RG ESC line, wherein the red fluorescent protein dsRed is under the control of the second heart field (SHF) enhancer of *Mef2C*, and the enhanced green fluorescent protein (GFP) is under the control of the cardiac-specific enhancer of *Nkx2.5* (Domian et al., 2009). This ESC line allows for the isolation of pure populations of FHF (GFP^+/$dsRed^-$ or G^+R^-) and SHF (GFP^+/$dsRed^+$ or G^+R^+, and GFP^-/$dsRed^+$ or G^-R^+) progenitors by FACS.

Mouse RG ESCs were differentiated as embryoid bodies using the hanging droplets method and samples collected every 48 hours to examine *Ccbe1* expression by quantitative PCR (qPCR). Cardiac specific genes including *Mesp-1*, *Isl1, Nkx2.5, αMhc* and *cTnt,* were then assayed as markers of cardiac differentiation. Expression analysis showed high levels of *Mesp-1* at day 4 of differentiation (Figure 3.1B) consistent with the formation of cardiac mesoderm. The expression of the cardiac mesoderm and SHF marker *Isl1* increased considerably at day 4 and peaked at day 6, while the general cardiac progenitors and primitive cardiomyocyte marker *Nkx2.5* increased only at day 6 (Figure 3.1C-D). The appearance of *Isl1* expression at day 4, prior to *Nkx2.5,* is consistent with the *in vivo* situation where *Isl1* expression precedes *Nkx2.5* expression during the formation of cardiac mesoderm (Laugwitz et al., 2008). The cardiomyocyte markers *cTnt* and *αMhc* were first expressed at day 6 with peak expression at day 10 (Figure 3.1E-F). Analysis of *Ccbe1* expression revealed similar levels of *Ccbe1* in undifferentiated mouse ESCs and days 2 and 4 of differentiation (Figure 3.1A). Higher *Ccbe1* expression was detected at day 6 of differentiation, concurrent with peak of expression of cardiac progenitor

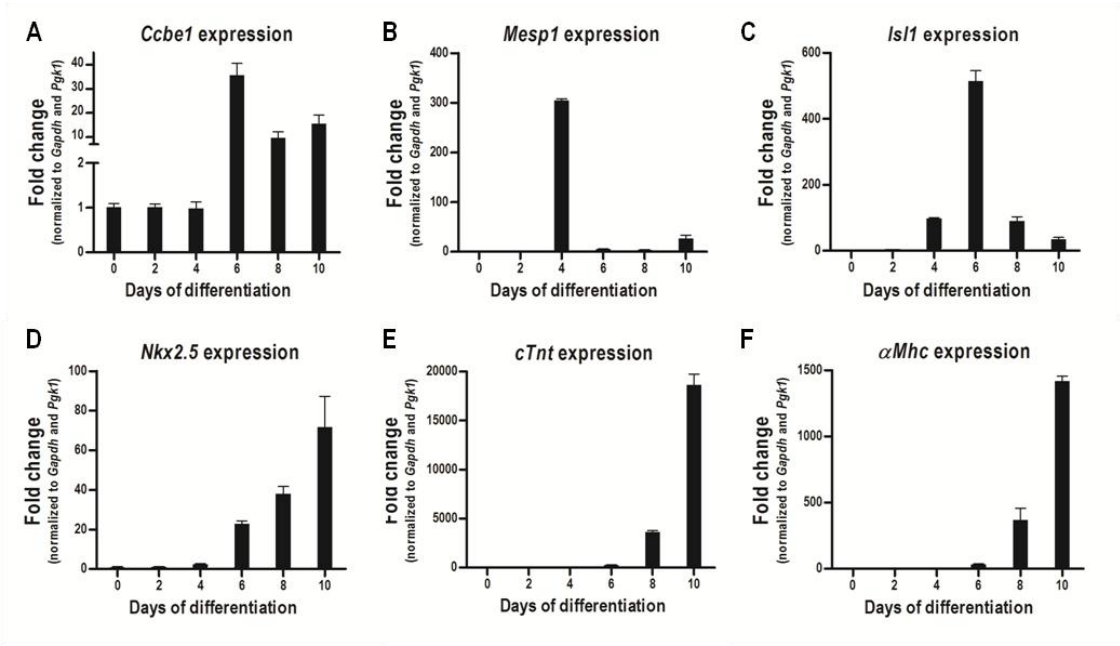


Figure 3.1 - Expression of (A) *Ccbe1*, (B) *Mesp-1*, (C) *Islet1*, (D) *Nkx2.5*, (E) *αMhc*, and (F) *cTnt* during differentiation of mouse ESCs. Samples were collected from undifferentiated cells (Und) and at days 2, 4, 6, 8 and 10 of differentiating RG mouse ESCs. Expression is presented as fold change relative to undifferentiated cells. Data represent the mean + SEM of two biological replicates in technical qPCR triplicates.

marker *Isl1* and the increasing expression of *Nkx2.5*. At days 8 and 10, *Ccbe1* was highly expressed but not as high as day 6. This suggests that in embryonic stem cells differentiating *in vitro*, like mouse embryos developing *in vivo* (Facucho-Oliveira et al., 2011), the appearance of high levels of *Ccbe1* expression coincide with the formation of FHF and SHF cardiac progenitors.

Ccbe1 is expressed in SHF and proepicardium cardiac progenitors

In order to confirm that the increase in *Ccbe1* expression is associated with the formation of FHF (G^+R^-) and SHF (G^+R^+ and G^-R^+) cardiac progenitor cells, we isolated these populations by FACS and analyzed *Ccbe1* expression along with *Isl1* and *Nkx2.5* expression to confirm the identity of the isolated populations. Expression analysis revealed *Ccbe1* transcripts in all cardiac progenitor populations (Figure 3. 2A-C). At day 6, *Ccbe1* expression was higher (3.8 fold) in SHF G^-R^+ cardiac progenitor cells or SHF-derived cells compared to

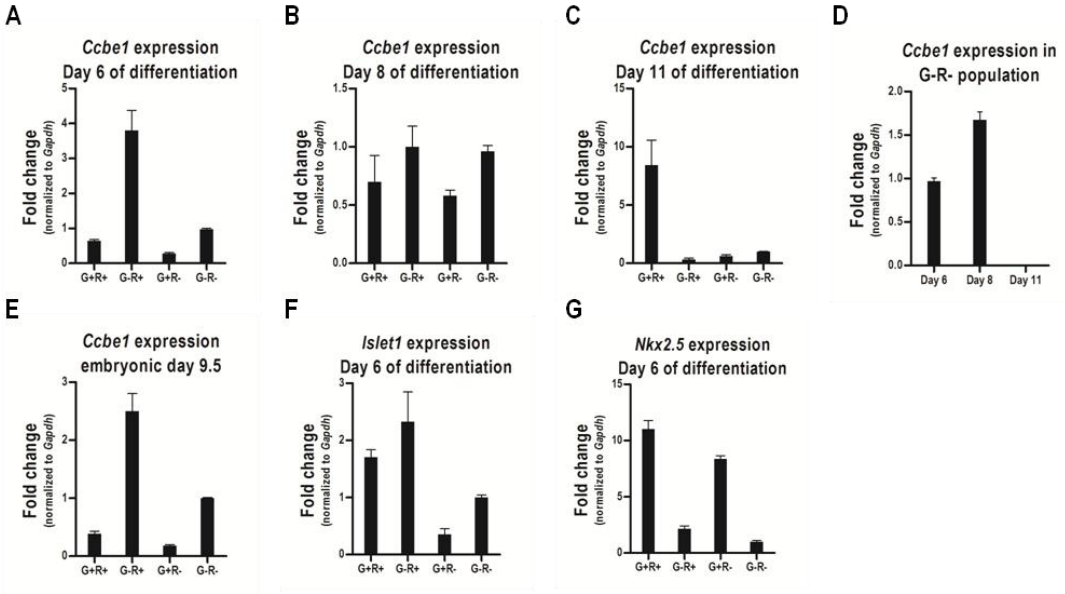


Figure 3.2 - ***Ccbe1*** **expression in cardiac progenitors isolated from differentiating mouse ESCs and embryos at E9.5.** Expression was analyzed in isolated populations, defined as FHF G^+R^- population, SHF G^+R^+ and G^-R^+ populations and control G^-R^- population at **(A)** day 6, **(B)** day 8 and **(C)** day 11 from differentiating mouse ESCs, and from **(E)** E9.5 mouse embryos. Expression is represented as fold change relative to the control G^-R^- population. **(D)** Analysis of *Ccbe1* in the control G^-R^- population at day 6, 8 and 11 relative to the expression at day 6. The identity of the sorted populations at day 6 was confirmed by analyzing the expression of **(F)** *Ist1* and **(G)** *Nkx2.5*. Mean + SEM of two biological replicates.

the double negative control (Figure 3.2A). In contrast, *Ccbe1* expression was down regulated in SHF G^+R^+ and FHF G^+R^- cardiac progenitors when comparing to the same control population.. Interestingly, a very similar *Ccbe1* expression profile was observed in FHF and SHF progenitors isolated from mouse embryos at E9.5 (Figure 3.2E). *Ccbe1* was also enriched in the equivalent SHF G^-R^+ population isolated from Nkx2.5-GFP/SHF-dsRed transgenic mouse embryos (2.5 fold). *Isl1* and *Nkx2.5* expression is consistent with the identity of the sorted populations from mouse RG ESCs at day 6 of differentiation (Figure 3.2F-G).

At day 8 of ESC differentiation, the G^-R^+ cells continue to have higher *Ccbe1* expression than the G^+R^+ and G^+R^- progenitors (Figure 3.2B). However, *Ccbe1* expression in the SHF G^-R^+ population was similar to the expression in control G^-R^- population, likely due to the emergence of non-cardiac cell types (G^-R^-) that also express high levels of *Ccbe1* (Figure 3.2E). *Ccbe1* has been previously shown to be expressed in the dermamyotome of the somites and in cells located in the vicinity of the anterior cardinal veins between stages E8.75 and E10.5 mouse development (Facucho-Oliveira et al., 2011). At day 11, *Ccbe1* was mainly observed in the G^+R^+ population (Figure 3.2C). This is likely related with the formation of proepicardium progenitors, which are known to derive from *Nkx2.5*$^+$/*Isl1*$^+$ cells at the lateral zone of the cardiogenic mesoderm (Mommersteeg et al., 2010; Zhou et al., 2008) and were also shown to express *Ccbe1* in mouse embryos (Facucho-Oliveira et al., 2011). Taken together, these results show that *Ccbe1* expression in differentiating mouse ESCs is concurrent with the appearance of specific populations of cardiac progenitors.

Ccbe1 knockdown leads to reduced cardiac mesoderm formation from differentiating ESCs

The presence of *Ccbe1* early on in both mouse ESC derived and mouse embryonic cardiac progenitors led us to hypothesize that *Ccbe1* could have a role during establishment or maintenance of cardiac progenitors. To test this, we generated transgenic mouse ESC lines expressing shRNAs targeting *Ccbe1* and evaluated the effect of *Ccbe1* loss-of-function in ESC-derived cardiac differentiation. *Ccbe1* KD ESC clonal lines were generated by lentiviral

transduction using RG as WT line. We carried out spontaneous differentiation of two individual *Ccbe1* KD ESC clones (Clone 1 and Clone 2) in comparison to the SH control ESC line to exclude off target effects of the shRNAs. Analysis of *Ccbe1* expression in undifferentiated cells confirmed significant *Ccbe1* downregulation in Clone 1 and Clone 2 when compared to the *Ccbe1* levels found in the control line (Figure 3.3A). At days 2 and 4 of differentiation, *Ccbe1* expression was also reduced when compared to that of the Control line, but was only statistically significant at day 4 (Figure 3.3A). Additionally, when *Ccbe1* expression normally peaks at day 6 of differentiation (Figure 3.1A), KD ESCs showed a very strong reduction in the expression of *Ccbe1* (Figure 3.3A). Together, this data confirm that knockdown of *Ccbe1* was maintained during the *in vitro* differentiation of both cell lines.

To understand if *Ccbe1* knockdown affects the pluripotency of mouse ESCs, we analyzed the expression of the pluripotency markers like *Sox2*, *Oct4* and *Nanog.* According to our analysis, expression of *Oct4* and *Sox2* was not affected in both KD ESC lines (Figure 3.3B). However, *Nanog* expression was significantly reduced in both clones. It is known that *Nanog* expression normally oscillates in pluripotent mouse ESC cultures and that ESCs with lower *Nanog* expression are more prone to differentiate but without any lineage bias (Abranches et al., 2014; Chambers et al., 2007). In agreement with this, expression of the early ectodermal marker *Fgf5,* of the endodermal marker *Afp* and of the mesoderm marker *BraT* were not significantly different in differentiating SH control ESCs and both *Ccbe1* KD ESC lines (Figure 3.3B). This suggests that, even though *Nanog* is downregulated, the *Ccbe1* KD ESCs retain the pluripotency and trilineage differentiation capacity.

We then showed that expression of cardiac mesoderm markers *Mesp1* and *Isl1* was markedly reduced both in Clone 1 and Clone 2 at day 4 of differentiation (Figure 3.3B). In addition, expression of *Nkx2.5* was also reduced in both differentiated *Ccbe1* KD clones. These data suggest that in the absence of *Ccbe1,* ESCs have reduced capacity to differentiate towards the *Mesp1* and *Isl1* cardiogenic mesoderm lineage. This in turn results in a reduction of Nkx2.5-expressing cardiac progenitor cells. Surprisingly, at day 6 of differentiation, expression of the cardiac precursor marker *Nkx2.5* was seemingly not affected,

and *Isl1* was downregulated only in clone 2 (Figure 3.3B). In the case of the early cardiomyocyte markers *αMhc* and *cTnt* there was a slight reduction in both *Ccbe1* KD clones, which could indicate that differentiation towards mature

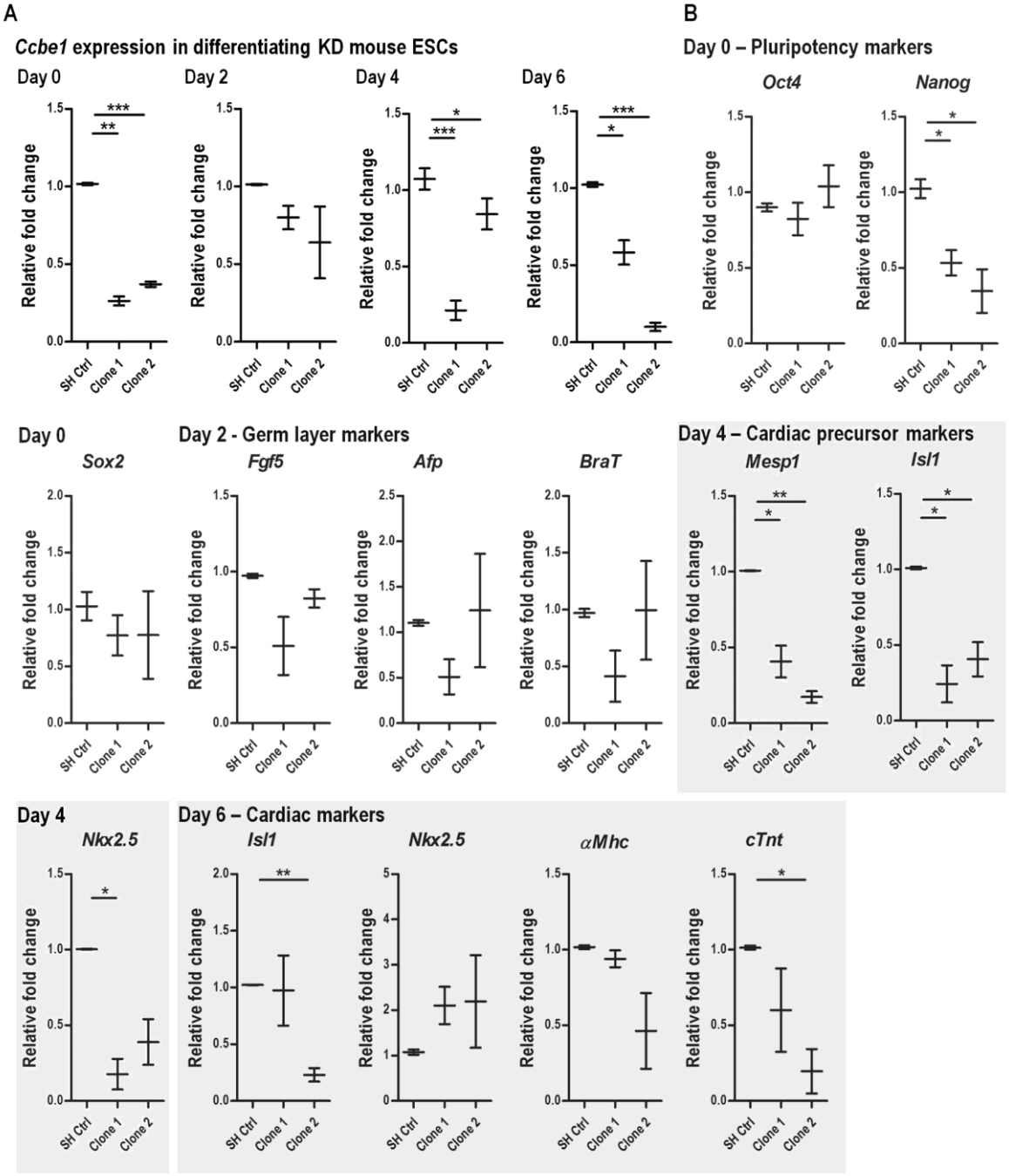


Figure 3.3 - *Ccbe1* knockdown leads to reduced cardiac mesoderm formation from differentiating mouse ESCs. (A) qPCR analysis of *Ccbe1* up to day 6 of differentiation; **(B)** qPCR analysis at **Day 0** of pluripotency markers *Oct4*, *Nanog* and *Sox2*, at **Day 2** of the germ layer markers *Fgf5, Afp* and *BraT;* at **Day 4** of the cardiac precursor markers *Mesp1* and *Isl1,* and *Nkx2.5; and* at **Day 6** of the cardiac markers *Isl1, Nkx2.5, cTnt* and *aMhc*. Analysis was performed in two individual *Ccbe1* KD ESC clones (Clone 1 and Clone 2) and compared to SH Control. Mean ± SEM of three biological replicates in technical qPCR triplicates; paired *t*-test relative to SH Control group; statistical significance * ***p***<0.05, ** ***p***<0.01, *** ***p***<0.001

cardiomyocytes is reduced in the absence of *Ccbe1*. Therefore, even though at day 4 of differentiation there was a decrease in the *Mesp1*- and *Isl1*-expressing cardiac mesoderm, at day 6 the *Nkx2.5*-expressing cardiac precursors were established in a seemingly normal proportion. However, morphological analysis of the embryoid bodies revealed that by the third day of differentiation the embryoid bodies derived from both *Ccbe1* KD ESC lines were slightly smaller when compared to the embryoid bodies derived from the parental line wild type (WT) RG or SH Control lines (Figure 3.4A). This phenotype was more evident at day 4 of differentiation. By day 5 that the embryoid bodies derived from *Ccbe1* KD ESCs lines were roughly half the size of the WT and SH controls (Figure 3.4A). Since the number of cells in the embryoid bodies can affect differentiation (Nakazawa et al., 2013; Zeng et al., 2013), we cannot exclude that the effect on the proportion of cardiac precursors by day 6 of differentiation in those clones is a consequence of the observed paucity in the embryoid bodies from derived from the KD ESC clones. Nevertheless, together these data suggest that *Ccbe1* is essential for the formation of ESC-derived *Mesp1*- and *Isl1*-expressing cardiac mesoderm and for proper growth of the embryoid bodies.

The defects induced by the knockdown of *Ccbe1* are CCBE1 specific

To confirm that it was indeed the absence of *Ccbe1* that led to the phenotype observed on the differentiating ESCs, two different strategies were conducted. On the one hand, we supplemented the WT RG ESCs during differentiation with CCBE1 antibody to function as a blocking antibody. We cultured the ESCs up to day 5 of differentiation in the presence of 100 ng/mL CCBE1 antibody in the differentiation medium or with the equivalent amount of buffer used as the control.

As shown in Figure 3.4B, supplementing the culture medium with CCBE1 antibody led to a significant decrease on the size of the embryoid bodies derived from WT ESCs by day 5 of differentiation. Furthermore, analysis of *Mesp-1 and Isl-1* expression in ESCs treated with CCBE1 antibody showed that their expression was also downregulated (Figure 3.4C), similarly to what was observed in the differentiating *Ccbe1* KD ESCs (Figure 3B, Day 4). These data are consistent with the phenotype observed in the differentiating ESCs being

specifically caused by *Ccbe1* loss-of-function. On the other hand, we supplemented the *Ccbe1* KD ESCs during differentiation with exogenous full length CCBE1 protein. Since CCBE1 is a secreted protein, ESCs could be cultured up to day 5 of differentiation in the presence of 200 ng/mL of

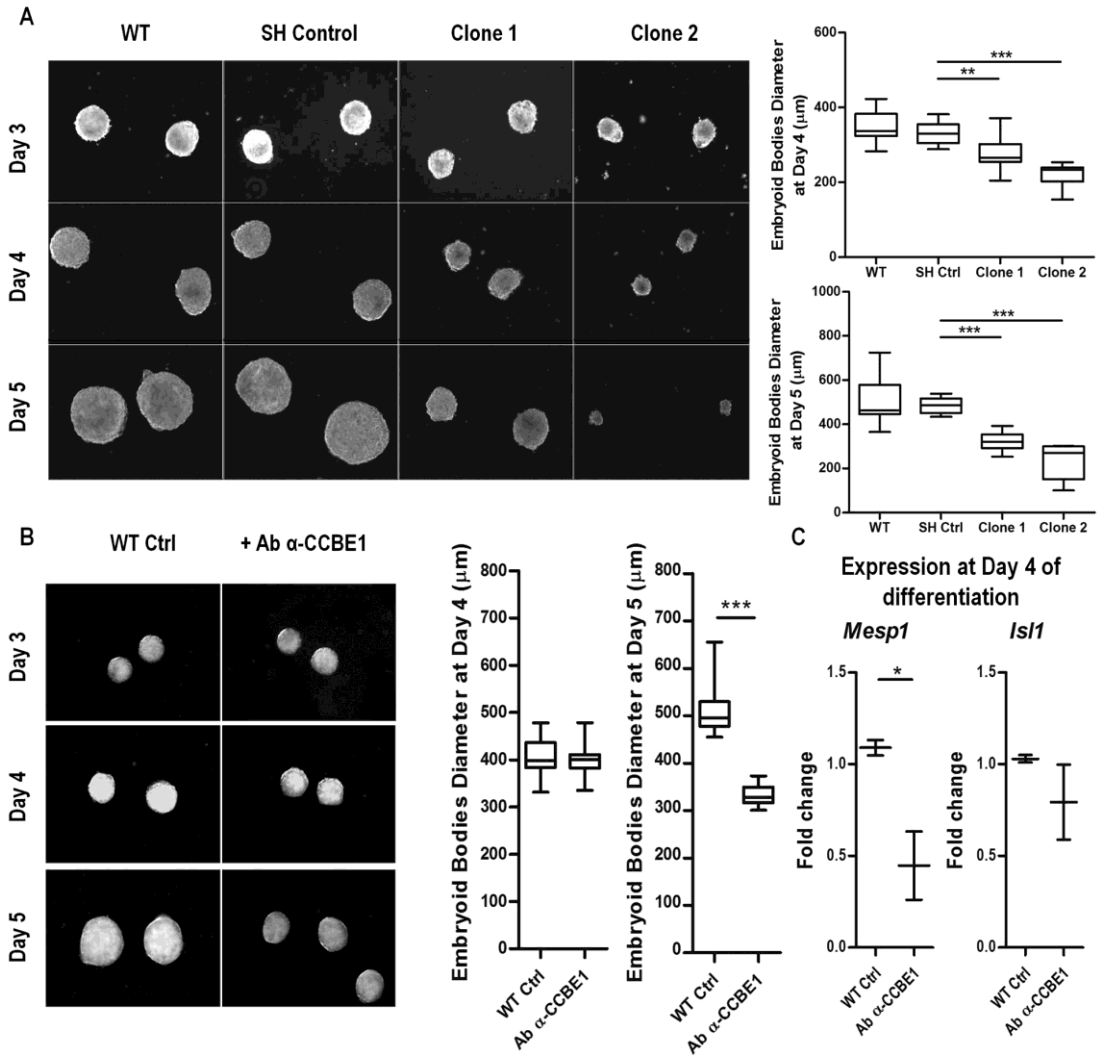


Figure 3.4 - *Ccbe1* loss-of-function leads to smaller embryoid bodies. Phase-contrast micrographs of *Ccbe1* KD ESC-derived embryoid bodies and controls **(A)** at days 3, 4 and 5 of differentiation. Graphs at the right side show the corresponding diameter measurements of embryoid bodies at days 4 and 5 of differentiation; **(B)** Phase-contrast micrographs of WT ESCs derived embryoid bodies at days 3, 4 and 5 of differentiation with and without supplementation of 100ng/mL of CCBE1 antibody. Graph on the right side shows the corresponding diameter measurements of the embryoid bodies at day 5 of differentiation; unpaired *t*-test relative to non supplemented group; statistical significance ***p***<0.001; Scale-bars: 200 µm; **(C)** qPCR analysis of cardiac precursor markers at day 4 of differentiation. Data represent the mean ± SEM of three biological replicates in technical qPCR triplicates; paired *t*-test relative to SH Control group; statistical significance * ***p***<0.05, ** ***p***<0.01, *** ***p***<0.001

recombinant CCBE1 protein in the differentiation medium or with the equivalent volume of buffer used as the control.

As shown in Figure 3.5A-B, supplementing the culture medium with recombinant CCBE1 protein led to a fair increase on the size of the embryoid bodies derived from the Ccbe1 KD ESC lines by day 5 of differentiation, more evident in the case of clone 2. This suggests that treatment with exogenous CCBE1 rescues the phenotype, but since the rescued embryoid bodies are still smaller than in the case of the WT and SH Control lines, and the fact that the

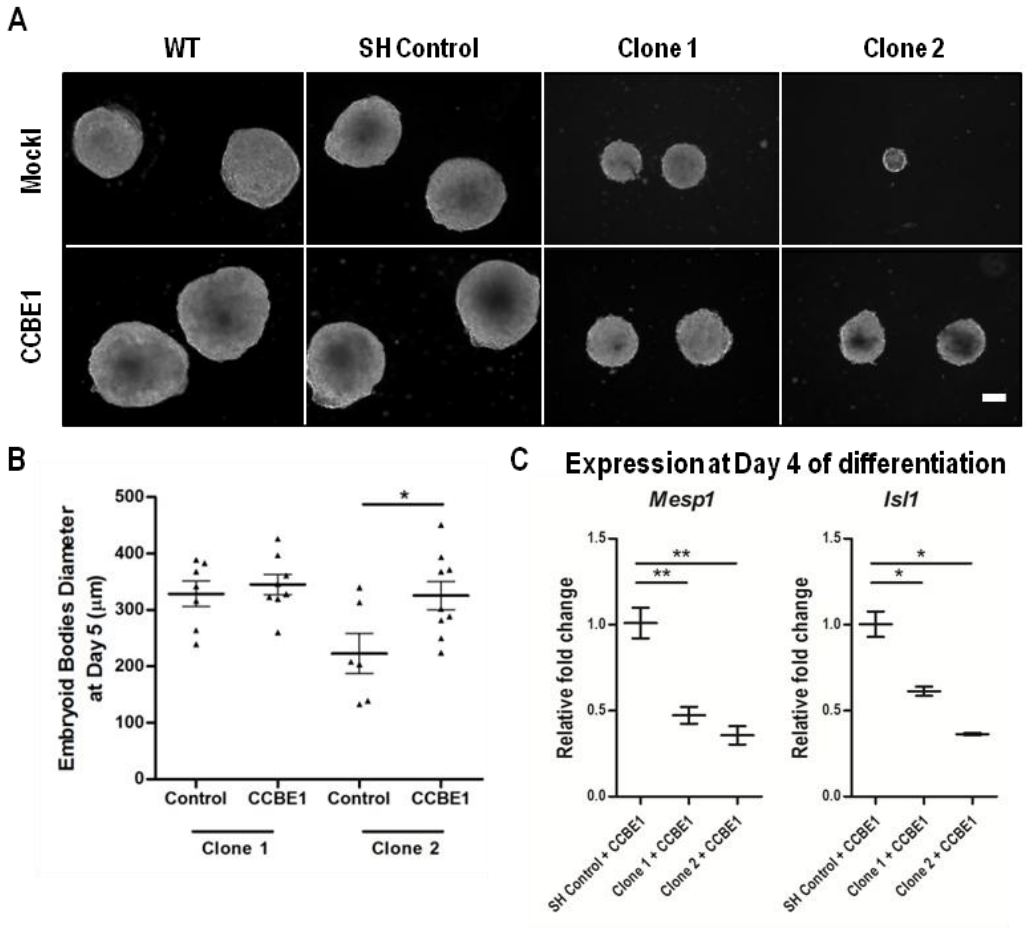


Figure 3.5 - Recombinant CCBE1 partially rescues the defects caused by the loss of *Ccbe1*. Phase-contrast **(A)** Phase-contrast micrographs of *Ccbe1* KD ESC-derived embryoid bodies and controls with supplementation of 200ng/mL recombinant CCBE1 or with buffer control, at day 5 of differentiation. Scale-bar: 200µm. **(B)** A partial rescue on the size of the *Ccbe1* KD embryoid bodies is observed at this stage of differentiation**. (C)** qPCR analysis of cardiac precursor markers at day 4 of differentiation. Data represent the mean ± SEM of two biological replicates in technical qPCR triplicates; paired *t*-test relative to SH Control group; statistical significance * ***p***<0.05, ** ***p***<0.01, *** ***p***<0.001.

cardiac mesoderm markers are still downregulated (Figure 3.5C), it is suggested that the rescue is only partial. This corroborates however that the phenotype observed in the differentiating *Ccbe1* KD ESCs is indeed caused by knockdown of *Ccbe1*.

Ccbe1 KD decreases the proliferation of differentiating ESCs

Next, we addressed the cellular mechanism causing the impaired growth of the *Ccbe1* KD ESC-derived embryoid bodies. While increased cell death and/or reduced proliferation are probable causes, another hypothesis is that the dense shell-like outer layer normally present in embryoid bodies, which secretes morphogens essential for mesoderm and cardiac mesoderm formation (Coucouvanis and Martin, 1999; Sachlos and Auguste, 2008), could be absent in *Ccbe1* KD embryoid bodies. It has been shown that absence of this dense shell-like layer, described as resembling the visceral endoderm of developing mouse embryos, caused by the blockade of collagen/β1-integrin interaction in mouse induced pluripotent stem cell (miPSC)- derived embryoid bodies affects both the size of embryoid bodies and cardiac differentiation (Zeng et al., 2013). Since CCBE1 is an extracellular protein with collagen domains, we decided to test if this dense outer layer was properly established in *Ccbe1* KD ESC-derived embryoid bodies. To test this, we performed a methylene blue (Mb) diffusion assay. Embryoid bodies were incubated with 1 mg/mL solution of Mb, and depending on the presence or absence of the visceral endoderm-like layer, the embryoid bodies should accumulate Mb at the surface or present a homogenous staining, respectively. According to our analysis and like in the controls, the staining accumulates at the surface of the *Ccbe1* KD ESC-derived embryoid bodies at all stages (Figure 3.6A). This indicates that the dense outer shell-like layer resembling the visceral endoderm is present in the absence of *Ccbe1*.

To understand whether the knockdown of *Ccbe1* affects cell viability in differentiating ESCs, we evaluated caspase-3 mediated apoptosis and cell proliferation. As shown in Figure 3.6B, immunofluorescence staining from cryosectioned embryoid bodies showed no obvious differences on the number of cleaved caspase-3 positive cells between control and *Ccbe1* KD ESC-derived embryoid bodies. Quantification of the percentage of apoptotic cells by

immunolabelling against cleaved caspase-3 followed by FACS at days 4 and 5 of differentiation confirmed that there were no differences between *Ccbe1* KD ESC-derived embryoid bodies and controls (Fig. 3.6C). This suggests that cell

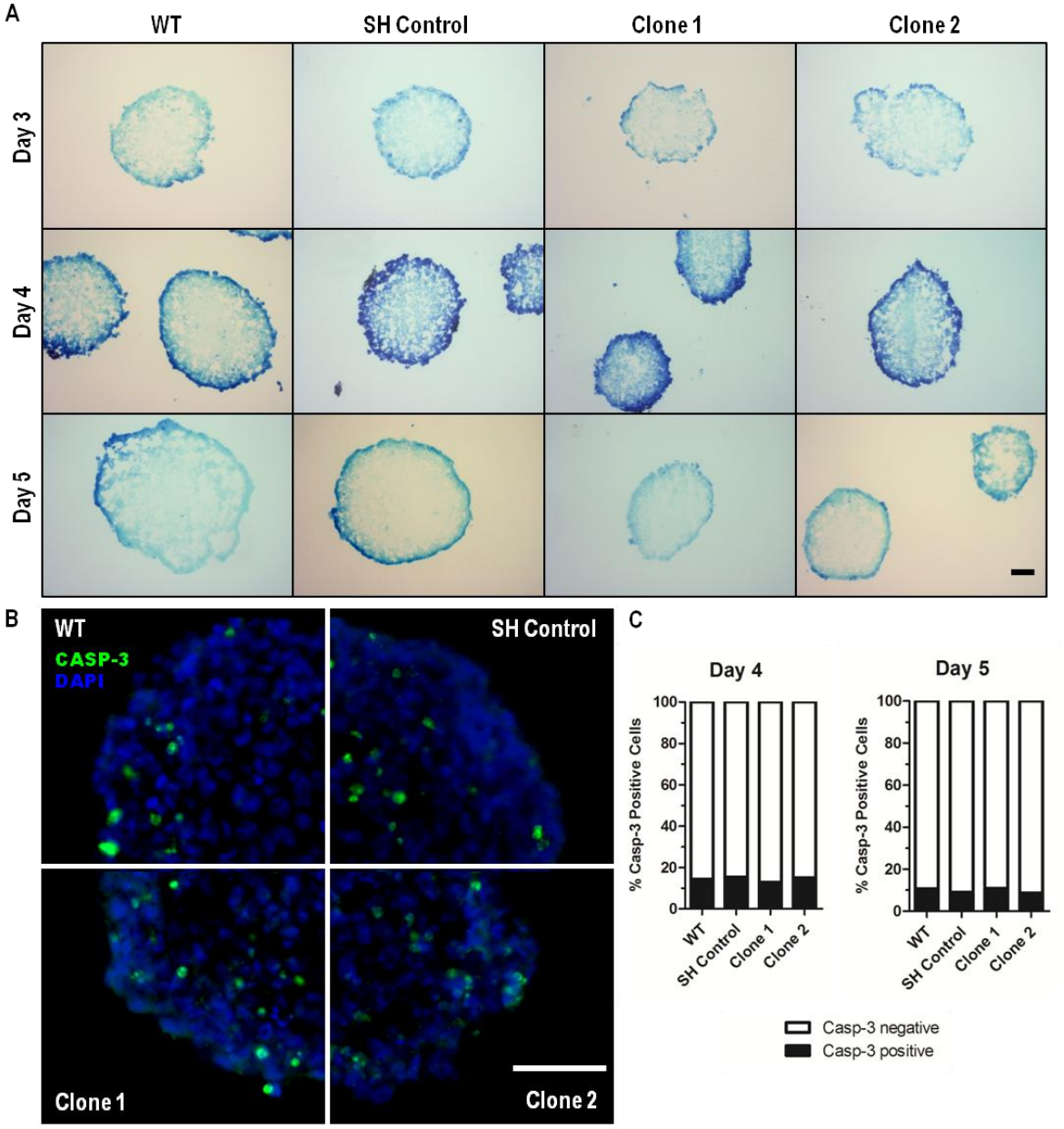


Figure 3.6 – Visceral endoderm-like layer is present and cell death is not affected in the absence of Ccbe1. (A) Methylene blue (Mb) diffusion assay. Bright-field micrographs of cryosectioned embryoid bodies at days 3, 4 and 5 of differentiation. The dark blue staining surrounding the surface confirms the presence of the dense outer shell-like layer in both controls and in the Ccbe1 KD ESC-derived embryoid bodies at all stages. Accumulation of Mb at the surface is caused by the presence of extracellular matrix proteins (ECM) synthesized by the visceral endoderm-like layer creating a physical barrier to the dye. **(B)** Micrographs of active caspase-3 immunofluorescence staining of cryosectioned embryoid bodies at day 4 of differentiation. The Green channel marks the active caspase3 positive cells, and the blue channel marks for DAPI. Scale-bar: 50µm. **(C)** Quantification of immunolabelled CASP-3 positive cells analyzed by FACS at days 4 and 5 of differentiation.

death cannot explain the smaller size of the embryoid bodies in the absence of *Ccbe1*.

Next, we performed a proliferation assay consisting of labelling the ESCs with nuclear proliferation dye eFluor® 670 and analyzing the fluorescence decay dependent on cell division for the following 6 days of differentiation. At day 0 the ESC were equally labeled, the cells presented the highest fluorescence and hence all cells locate at quadrant (Q)1 (Figure 3.7A). At the second day, all the cells located at Q2, indicating that the cells from all individual lines divided in a similar manner (Figure 3.7B). At day 3 of differentiation, while the fluorescence decay continued gradually in WT and SH control lines, the clone 2 presented a delay suggesting a decreased proliferation rate (Figure 3.7C). At day 4 both *Ccbe1* KD differentiating ESCs showed a higher percentage of cells in Q2 when compared to the controls (Figure 3.7D), indicating that the proliferation rate of both lines was decreased. At days 5 and 6 this difference was more pronounced (Figure 3.7E-F), where most of the differentiated ESCs from the control lines had lost all the fluorescence (Q4 – 73%), but the differentiating ESCs from Clone 1 and 2 had more cells remaining in Q2 35% and 43%, and in Q3 51% and 26% respectively. This data strongly suggests that knockdown of CCBE1 results in a reduction in the proliferation of progenitors by days 5 and 6 of differentiating *Ccbe1* KD ESC lines. This observation is consistent with the observed smaller size of the embryoid bodies derived from both *Ccbe1* KD clones (Figure 3.4A). In contrast, in pluripotency there were no differences in the proliferation between control and *Ccbe1* KD ESC lines (data not shown). Taken together, these data indicate that knockdown of *Ccbe1* leads to decreased proliferation of differentiating ESCs, which becomes striking between day 4 and day 6 of differentiation.

CCBE1 functions in a VEGF-C/VEGFR3 independent signaling pathway during cardiac differentiation.

During lymphangiogenesis, CCBE1 is required to facilitate the maturation of Pro-VEGF-C into its mature form (Jeltsch et al., 2014; Le Guen et al., 2014). Even though VEGF-C is indisputably involved in the migration of LECs during lymphangiogenesis, in other contexts it has also been shown to regulate cell proliferation (Dias et al., 2002; Shin et al., 2002). This prompted us to determine

if *Vegf-C* was normally expressed during ESC differentiation and, if the observed defects in *Ccbe1* KD ESCs lines can be rescued by mature recombinant VEGF-C. Gene expression analysis confirmed the expression of

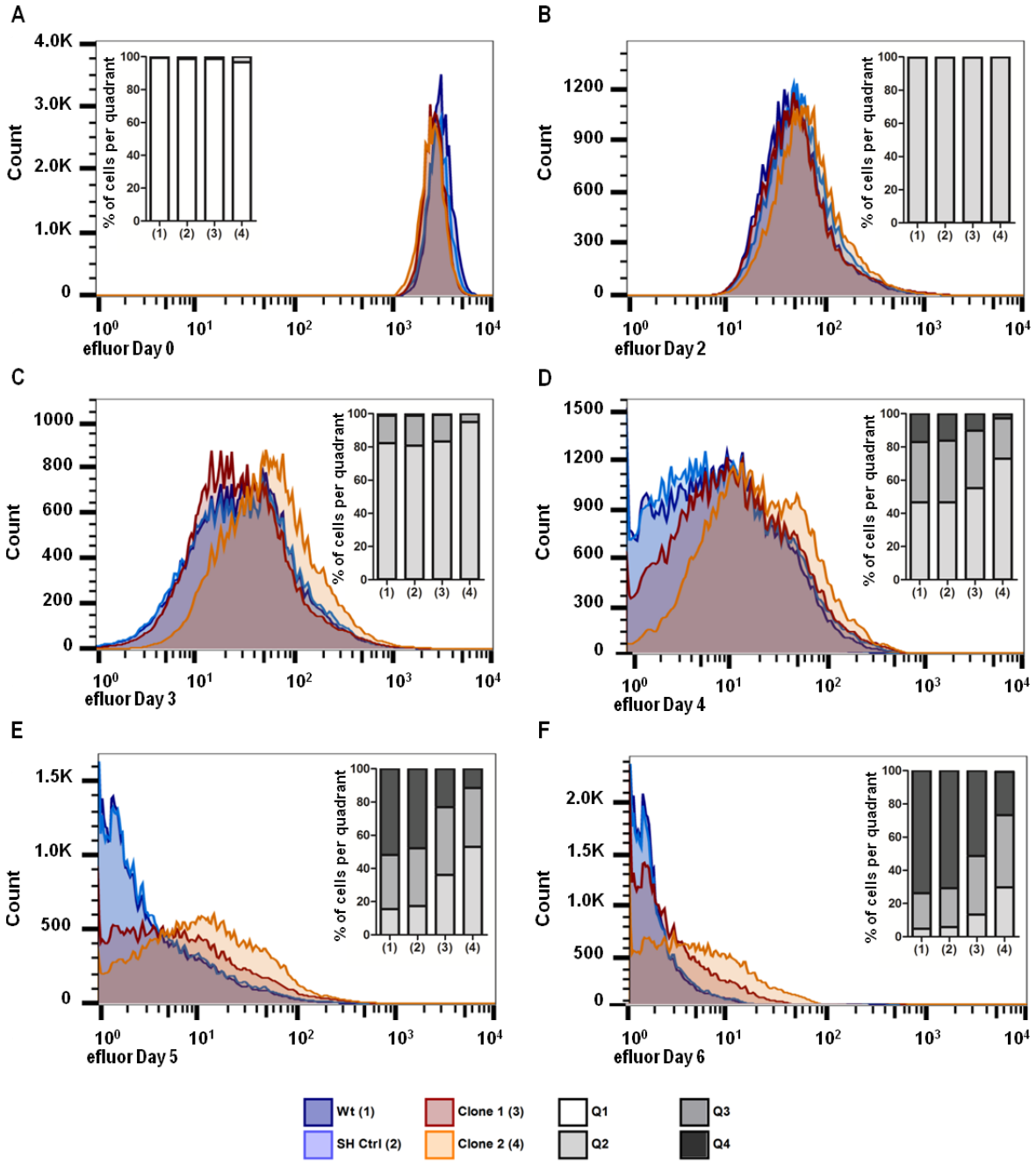


Figure 3.7 - Ccbe1 knockdown decreases the proliferation of differentiating ESCs. Cell Proliferation Assay with dye eFluor® 670 analysed by FACS. Colored histograms represent the cell count for each ESC line: (1) WT Control (dark blue), (2) SH Control (light blue), (3) Clone 1 (dark red) and (4) Clone 2 (orange). Histograms were divided into four quadrants (Q1-Q4) according to the decreasing fluorescence of the controls, along the six different time points: (A) day 0, (B) day 2, (C) day 3, (D) day 4, (E) day 5 and (F) day 6 of differentiation. The maximum fluorescence corresponds to Q1, while no fluorescence corresponds to Q4. The bar-graphs represent the percentage of cells distributed in each quadrant for each ESC line. Two biological replicates were performed.

Vegf-C and its receptor *Vegfr3* during WT mouse ESCs differentiation with higher expression levels of *Vegf-C* appearing at days 4 and 5 of differentiation (Figure 3.8A). This is in agreement with the identified expression of *Vegf-C* in mouse embryos as early as E7.0, when the different early mesoderm progenitors are being established (Kukk et al., 1996). Furthermore, this time point coincides with the onset of the phenotypes observed in the differentiating *Ccbe1* KD ESCs. Interestingly, at days 4 and 5 of differentiation, *Vegf-C* was downregulated in both *Ccbe1* KD clones and in ESCs supplemented with antibody against CCBE1 when compared to the respective controls (Figure 3.8B-C). This could be an indication that the phenotypes observed in the absence of *Ccbe1* were related to reduced VEGF-C signaling. This prompted us to supplement differentiating *Ccbe1* KD ESC with 100 ng/mL of mature recombinant VEGF-C. However, mature VEGF-C supplementation did not rescue the growth of EBs, nor the expression of *Mesp1* and *Isl1* cardiac mesoderm markers in the differentiation KD clones (Figure 3.8D-E). Moreover, *Esm1* and *Prox1,* two known targets of the VEGF-C signaling during lymphangiogenesis (Shin et al., 2008), were not significantly altered upon supplementation with mature recombinant VEGF-C (Figure 3.8E). Collectively, these results suggest that, despite the expression of *Vegf-C* and its receptor, VEGF-C signaling is not required for CCBE1 activity during cardiac differentiation. Thus, CCBE1 acts independently of VEGF-C signaling during cardiac mesoderm commitment and cell proliferation of differentiating ESCs

3.4. DISCUSSION

We here show that like mouse and chick embryonic cardiac development (Facucho-Oliveira et al., 2011; Furtado et al., 2014), *Ccbe1* is upregulated in SHF cardiac progenitors derived from ESC differentiating *in vitro*. Thus, high *Ccbe1* expression is correlated with the onset of cardiac specification both *in vivo* and *in vitro*. In addition, disruption of normal CCBE1 activity by shRNA knockdown or by blocking antibody results in a clear decrease in the expression of the early cardiac mesoderm markers *Mesp1* and *Isl1* at day 4. These results show that CCBE1 is required for the normal development of early cardiac precursors.

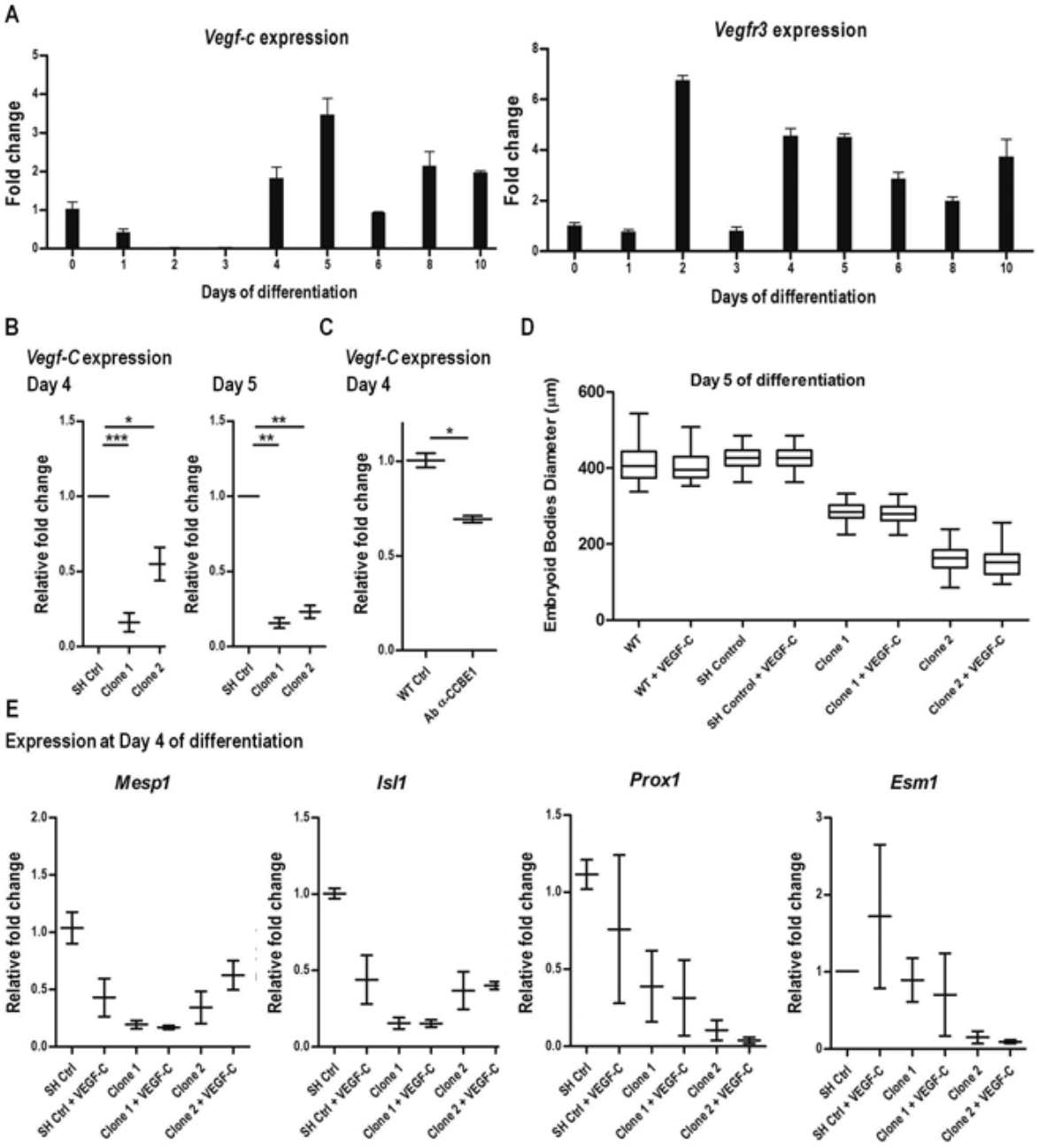


Figure 3.8 – Defects caused by the absence of *Ccbe1* seem unrelated to the role of CCBE1 in VEGF-C maturation. (A) Expression of *Vegf-c and Vegfr3 in* differentiating ESCs; **(B)** qPCR analysis of *Vegf-c* in *Ccbe1* KD ESC clones and SH Control at days 4 and 5 of differentiation; **(C)** qPCR analysis of *Vegf-c* in WT embryoid bodies supplemented with 100ng/mL of CCBE1 antibody; **(D)** Embryoid bodies diameter measurements at day 5 of differentiation after supplementation with 100ng/mL of mature VEGF-C; paired *t*-test relative to non supplemented groups; no statistical significance is present. **(E)** qPCR analysis of cardiac precursor markers *Mesp1* and *Isl1,* and of downstream targets of VEGF-C signaling *Prox1* and *Esm1* at day 4 of *Ccbe1* KD ESC differentiation upon supplementation or not with 100ng/mL of mature VEGF-C; Data represents the mean ± SEM of three biological replicates in technical qPCR triplicates; paired *t*-test relative to SH Control group; statistical significance * ***p***<0.05, ** ***p***<0.01, *** ***p***<0.001

The developmental fate of differentiating ESCs depends on embryoid body size, growth factor signaling, ECM proteins that constitute the developmental nich and how ESCs interact with this niche (Bratt-Leal et al., 2009; Czyz and Wobus, 2001; Goh et al., 2013; Higuchi et al., 2013; Taylor-Weiner et al., 2013; Zeng et al., 2013). We show here that in the absence of *Ccbe1*, the growth of differentiating embryoid bodies was arrested especially from day 4 onwards. This impaired embryoid body growth is caused by reduced cell proliferation and not by cell death, nor by the absence of the outer endoderm-like layer. Therefore, it is possible that the reduced proliferation rate of *Ccbe1* KD ESCs differentiating in vitro could influence cardiac mesoderm commitment. Nevertheless, the size of the embryoid bodies treated with CCBE1 blocking antibody was preserved at day 4 and was only slightly reduced at day 5. In contrast, the expression of the cardiac precursor markers was markedly diminished during the same time period. This indicates that the cardiac mesoderm commitment phenotype is *Ccbe1* dependent and is not primarily related to the size of the embryoid bodies. The effect of reduced embryoid body size seems to be more critical from day 5 onwards, coinciding with the time that cell proliferation is more severely affected and embryoid bodies are strikingly smaller. From this time point onwards most alterations in gene expression may be related to the reduced size of the embryoid bodies and consequent alterations in the microenvironment of the cells. Therefore, our data suggests that, besides its role during early cardiac commitment, *Ccbe1* may have an independent role in the proliferation of differentiating ESCs.

In mice, *Ccbe1* loss-of-function results in embryonic lethality at E14.5 due to lymphangiogenesis defects (Bos et al., 2011). Indeed, CCBE1 has been implicated in the modulation of VEGF-C signaling during mammalian lymphangiogenesis (Bos et al., 2011; Hagerling et al., 2013) and zebrafish lymphangiogenesis (Astin et al., 2014; Hogan et al., 2009; Le Guen et al., 2014). In that context, CCBE1 recruits the metalloprotease ADAMTS3 to promote the *in situ* maturation of pro-VEGF-C into a mature lymphangiogenic-promoting form, thereby activating the downstream signaling pathway (Jeltsch et al., 2014; Le Guen et al., 2014). Interestingly, *in vitro* studies have also shown that VEGF-C regulates cell proliferation in contexts other than lymphangiogenesis (Dias et al., 2002; Shin et al., 2002). The only function

described so far for CCBE1 is on the maturation of VEGF-C. Furthermore expression of *Vegf-C* was downregulated in the *Ccbe1* KD ESCs. We therefore hypothesized that the cell proliferation and cardiac mesoderm commitment defects caused by the silencing of *Ccbe1* could be related to Vegf-C maturation. However, mature recombinant VEGF-C did not rescue the *Ccbe1* KD phenotype with respect to embryoid body size or expression of cardiac mesoderm markers *Isl1* and *Mesp1.* In the experiments with the *Ccbe1* KD clones, this downregulation could be somewhat related to the reduced size of the embryoid bodies. However, as mentioned above, the WT embryoid bodies supplemented with the CCBE1 antibody had preserved size at day 4 of differentiation, but still showed the downregulation of the cardiac mesoderm markers and *Vegf-C*. Therefore, it is possible that the downregulation of *Vegf-C* is more directly related with the absence of *Ccbe1*, like in the case of the cardiac mesoderm markers, or with problems in the specification of a particular cell population that could express *Vegf-C*. Contrasting with this, supplementing the differentiating *Ccbe1* KD ESCs with full-length recombinant CCBE1 led to a fair rescue the size of the EBs. Therefore this suggests that the defects caused by the disruption of *Ccbe1* are likely independent of the molecular mechanism involving VEGF-C signaling. Nonetheless, since the supplementation with full-length recombinant CCBE1 does not rescue the expression of the cardiac mesoderm markers when *Ccbe1* is downregulated, one cannot exclude that in its role in early cardiac specification, CCBE1 can synergistically interact with additional growth factors which expression may also be affected.

During mouse development, the expression of *Vegf-C* can be detected in the embryo as early as E7.0 (Kukk et al., 1996), which corresponds to days 4 and 5 of ESCs differentiation, and correlates with our expression analysis. However, it is only around E8.5 that *Vegfr3* starts to be expressed in mice (Kukk et al., 1996), and lymphangiogenesis is triggered around E9.0 when the first LECs start to differentiate from blood endothelial cells (Francois et al., 2008). Prior to that stage, there is no obvious role appointed to VEGF-C/VEGFR3 signaling pathway. In agreement with this, when supplementing differentiating ESCs with mature recombinant VEGF-C, the two downstream targets of VEGF-C signaling *Esm1* and *Prox1* remained unaltered despite the presence of VEGFR3 at those stages. Therefore, this suggests that those target genes

induced during lymphangiogenesis are not activated by VEGF-C signaling during early ESC differentiation or, alternatively, that the ESCs at those differentiation stages may not respond to VEGF-C signaling. Reports where similar experiments with supplementation of mature VEGF-C during ESCs differentiation were performed showed that VEGF-C influenced differentiation much later during differentiation in a lymphangiogenesis-like process (Liersch et al., 2006; Mishima et al., 2007). Therefore, our data is in agreement with the defects in cell proliferation and cardiac mesoderm commitment being caused by the absence of CCBE1 and independent of its role in VEGF-C maturation.

In vivo, *Ccbe1* is expressed at later stages of mammalian cardiac development (Facucho-Oliveira et al., 2011), and both chick embryos and HS patients present cardiac defects when *CCBE1* is disrupted (Alders et al., 2009; Connell et al., 2010; Connell et al., 2012; Furtado et al., 2014). Therefore, it would be interesting to study if *Ccbe1* has a role at later stages of ESCs cardiac differentiation. One possibility would be to study the disruption of the gene at specific time points later during cardiac commitment in differentiating ESCs. This would allow us to evaluate whether the modulation of *Ccbe1* can contribute *in vitro* to the full maturation of cardiac progenitors and/or beating cardiomyocytes.

Taken together, our data indicates that during ESC-derived cardiac differentiation *Ccbe1* is required to promote cell proliferation and the formation of cardiac mesoderm precursor cells.

Chapter IV

From Stem Cells to Heart: Identification of novel cardiac genetic players by RNA-seq

Tiago Justo, Isabel Duarte, Paulo Pereira, José Belo, Matthias Futschik

[1] Regenerative Medicine Program, Biomedical and Medicine Sciences Department and

[2] Institute for Biotechnology and Bioengineering, Center for Molecular e **Structural** Biomedicine, University of Algarve, Campus de Gambelas, 8005-139 Faro, Portugal

[3] PhD Program in Biomedical Sciences, UALG, Portugal.

[4] Cardiovascular Research Center, Massachusetts General Hospital, Boston MA 02114–2790, USA and Harvard Stem Cell Institute, Cambridge MA 02138, USA

[5] Stem Cells and Development Laboratory, CEDOC, NOVA Medical School / Faculdade de Ciências Médicas, Universidade Nova de Lisboa, 1150-082 Lisboa, Portugal

Author's contribution:

All the laboratorial work was carried out by T. Justo (except the RNA sequencing, which was done by an external company). The bioinformatics analysis of the RNA-seq data was performed by I. Duarte. All authors have contributed to the interpretation of the results.

4.1. SUMMARY

The identification of genetic signatures of specific cells populations enables their isolation for transplantation in patients, and/or disease treatments in regenerative medicine applications. Here we aimed at investigating some of the genetic networks underlying embryonic cardiogenesis to identify novel genes or isoforms. Firstly, we isolated from differentiating ESCs a sub-population of cardiac progenitor cells and collected good quality RNA. After, we identified by RNA sequencing – a whole-transcriptome screening analysis – significant differential expression for 3 potentially novel cardiogenesis-associated genes, namely Asb2, Cck and 3632451O06Rik.

KEYWORDS: Stem cells; Cardiac progenitors; Heart development; First heart field; Second heart field; RNA-seq; Transcriptome; Functional enrichment analysis.

4.2. INTRODUCTION

Vast arrays of genes and signalling pathways have been reported to be crucial in all steps of cardiogenesis, whether we refer to *in vivo* or *in vitro* models. When the cardiac mesoderm specification occurs, different early cardiac markers start to be expressed, such as *Nkx2.5, Isl1, Gata4* and *Tbx18* (reviewed in Meilhac et al., 2014), and such markers are also found to be expressed in differentiating pluripotent stem cells (PSCs; Laflamme and Murry 2011).

During the embryonic development there are mainly two pools of progenitor cells that give rise to the heart – the first heart field (FHF) and the second heart field (SHF). These progenitors have only recently been described in mouse ESCs (Domian et al., 2009). These authors generated a double transgenic mouse ESC line, wherein the red fluorescent protein dsRed is under the control of the SHF enhancer of *Mef2C*, and the enhanced green fluorescent protein (eGFP) is under the control of the cardiac-specific enhancer of *Nkx2.5*. With the expression of these fluorescent markers it was possible to isolate, in developing hearts and differentiating ESCs, the cell populations expressing one, both or none of these reporters, corresponding to the sub-populations of the FHF, SHF and non-cardiac control cells.

This cell line is a good model for embryonic cardiogenesis and a powerful tool to study the divergent genetic origin of the different cardiac progenitor populations, which supports our choice to work with this double transgenic ESC line, where we performed RNA sequencing (or simply RNA-seq) in order to analyze the transcriptome of each cell sub-population. This allows the quantification of gene expression in a genome-wide fashion and in a given moment in time, representing an additional way of identifying novel genes or growth factors that regulate cardiogenesis. Additionally, it enables the discovery and identification of putative isoforms of well-known cardiac genetic players, as well as the analysis of how their expression changes with time within the same cell population, and between populations from the same differentiation time point.

4.3. RESULTS

Isolation of different cardiac progenitor populations and RNA collection, integrity control and quality analysis

Differentiating the double transgenic ESCs line, here on referred as RG, allowed the isolation of pure populations of FHF (G^+R^-) and SHF (G^+R^+ and G^-R^+) progenitors by fluorescence-activated cell sorting (FACS). ESCs were cultured as embryoid body aggregates (EBs) as described before, and cell suspensions were prepared at days 4 and 6 of differentiation to isolate the populations of cardiac progenitor cells under FACS. Prior to sorting, cell complexity and viability was taken into account to set the scatter gate, and the sorting gates established for cell sorting were adjusted as described in the literature (Figure 4., upper left panel; Domian et. al 2009). Indeed, this approach granted the successful isolation of the four individual cell populations. The upper right panel of Figure 4.1 indicates that according to the gates described, on day 4 of differentiation the G^+R^- isolated population (gate P4) represents around 1% of the total cell suspension, the G^-R^+ population (gate P5) represents 0,3% and G^+R^+ population (gate P3) represents 0,2%. The same panel depicts a final schematic representation with colored dots identifying each of the sorted populations. On day 6 of differentiation the percentage of cells in each population did not change considerably (data not shown). To assess the purity of the sorted populations, the preliminary sorted cells underwent a second round of sorting, using the same gate settings, hence confirming the identity of these cells, and verifying the that the mechanical stress exerted on them during the sorting process, did not have a significant impact on cell viability (data not shown).

After sorting the cells, total RNA collection was performed as described before, and RNA quality and integrity assays were performed using the Experion™ automated electrophoresis chips. This procedure is a crucial step in the isolation of mRNA for further sequencing analysis, since it detects potential contamination of our sample with ribosomal RNA (rRNA) and genomic DNA. During the electrophoresis the separation generates a broad mRNA peak, and

any contaminating rRNA peaks will be visible. rRNA peaks and RNA ratios, such as the ratio of the two large ribosomal RNA molecules (28S/18S rRNA), are calculated for total RNA samples providing a measure of RNA degradation since the 28S rRNA is more susceptible to degradation than the 18S fragment. An intact RNA sample has a 18S:28S ratio of around 2.0. In this analysis the results are displayed in an electropherogram and simulated gel view, which indicates if the sample has been degraded and if it contains genomic DNA or rRNA contaminations. The combination of these two parameters gives an RQI

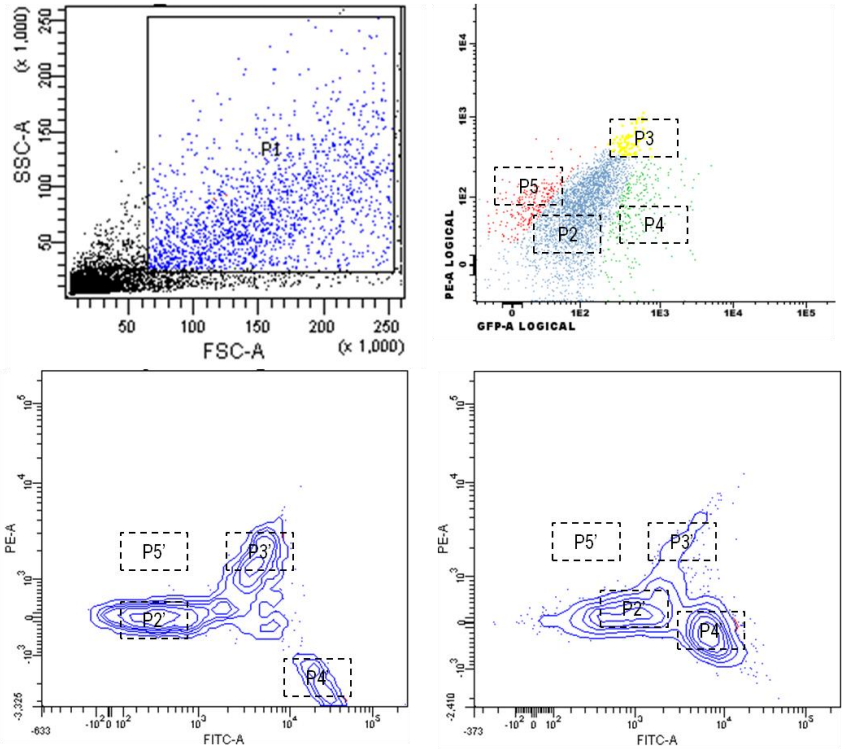
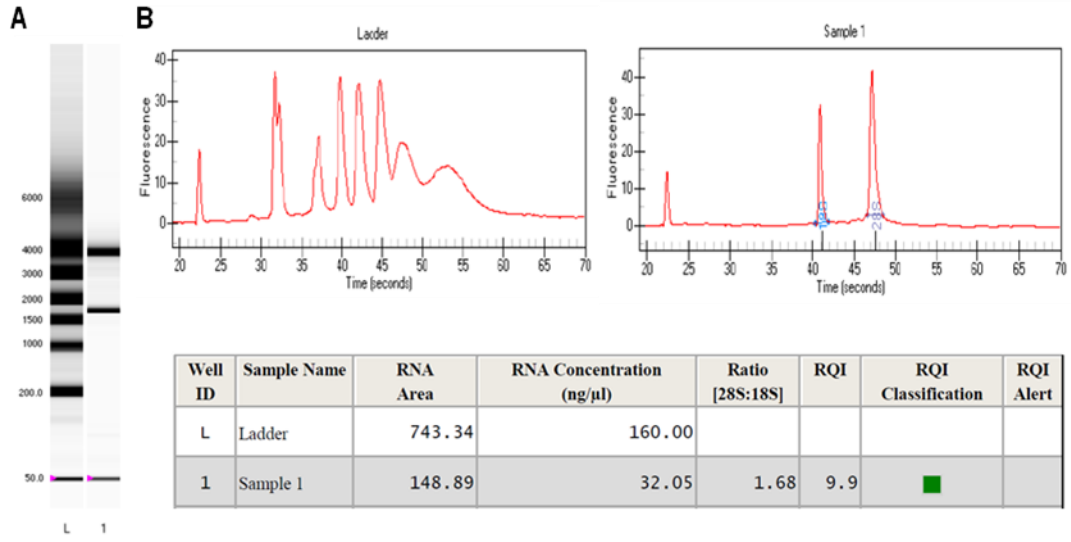


Figure 4.1 – Gate settings used for ESC-derived cardiac progenitor cell sorting. The **upper left panel** shows cell complexity and viability, which was taken into account to set the scatter gate P1. The **upper right panel** shows the gate settings established for the cell sorting used in our experiment, as described in the literature, here exemplified by one of the replicates from day 4 of differentiation. The red fluorescence is represented on the y-axis and the green fluorescence is represented on the x-axis. Both axes are in *Logicle* (bi-exponential) scale. Colored dots are represented as a visual aid to highlight each of the populations sorted, where the blue dots from P2 correspond to the G^-R^- population, the yellow P3 dots correspond to the G^+R^+ population, the P4 green dots, correspond to the G^+R^- population and the red P5 dots correspond to the G^-R^+ population. The **two lower panels** show a retrospective analysis of the FACS data of our cell sorting experiments, displaying a different visualization of the cell population density (left panel corresponds to day 4 of differentiation, and right corresponds to day 6 of differentiation). Gates P2'-P5' show the optimal settings to accurately sort the cell populations *de facto* present in our sample.

score with optimal values varying between 8-10. Figure 4.2 shows an example of such analysis in one RNA sample isolated from one of the sorted cell populations, indicating that no contaminants with DNA or rRNA were present. A RQI value of 9.9 was obtained for this sample indicating a biological material of excellent quality. The results for this sample are representative of all the 18 samples collected for RNA sequencing. In total 500ng of each RNA sample were used for sequencing. Please note that one of the replicates of the G^+R^+ population at day 6 was not sequenced since the RNA got degraded during its transportation to the sequencing lab located in Finland.

RNA sequence data quality analysis

The first step of RNAseq data analysis is checking the quality of the resulting dataset. There are several quality metrics that are routinely used, from which we chose two of the most informative ones: (i) hierarchical clustering of the samples (Figure 4.) and (ii) the distribution of the expression values (Figure 4.4). When we cluster the expression values for all mapped reads (mapped reads are the ones that confidently align to the mouse genome), we obtain the



Well ID	Sample Name	RNA Area	RNA Concentration (ng/µl)	Ratio [28S:18S]	RQI	RQI Classification	RQI Alert
L	Ladder	743.34	160.00				
1	Sample 1	148.89	32.05	1.68	9.9		

Figure 4.2 - Integrity and quality analysis of a representative RNA sample from the cell populations isolated by FACS. In **(A)** the RNA sample 1 does not contain any contaminations with genomic DNA nor with rRNA, and it kept its integrity, i.e. it is not degraded. In **(B)** the ratio between the18S and 28S peaks from sample 1 shows a good RQI value of 9.9. This indicates that the sample is of excellent quality and can proceed to RNA sequencing.

dendrogram shown in Figure 4., containing three main clusters corresponding to a separation of the samples according to time of differentiation. Noteworthy is the fact that replicate 1 from the $G^{-}R^{+}$ sample for day 6 clusters outside all other samples, suggesting that the values of expression measured for this replicate are significantly different from its replicate (RedDay6_0). Here, we hypothesize that this difference is due to the fact that the RNA from this sample was extracted via a different protocol (trizol). Additionally, for each time point we

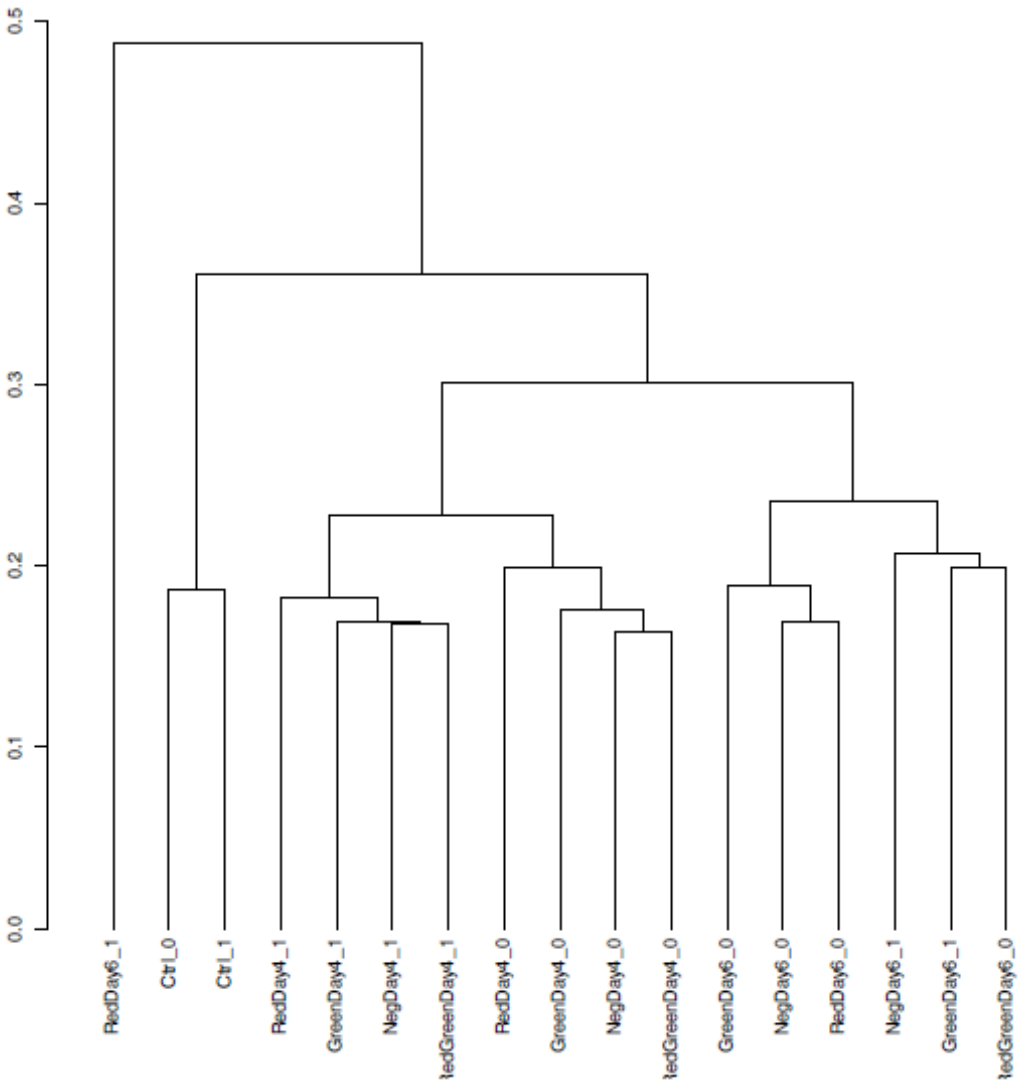


Figure 4.3 – Hierarchical clustering of the expression of all genes (showing the replicates). When we cluster the expression values for the mapped reads we obtain the following dendrogram where we can see that there are three main clusters separating the samples time wise. Noteworthy is the fact that the replicate sample for red day 6 clusters outside hinting that the values of expression measured are significantly different from its replicate. Since the RNA from this sample was extracted via a different protocol (trizol) we hypothesize that this fact explains this observation.

observe a tendency for all cell populations from the same biological replicate to cluster together instead of clustering with its population replicate. This suggests that the cell populations from the same differentiation assay are more similar between themselves, than they are between the same population from the other biological differentiation replicate. Accordingly, this hierarchical clustering analysis shows that time is the most significant variable in our experimental setting, i.e. time of differentiation explains most of the variability, while the subpopulations of cardiac progenitors is a secondary variable.

Figure 4.4 shows the distribution of the gene expression values in each population evaluated. As expected, the normalization procedure has generated samples with very similar distributions. However, the RedDay6 sample distribution shows a significantly different median, as highlighted by the red visual aid line. This observation corroborates the results from the hierarchical clustering analysis, which shows its replicate 1 clustering outside the main

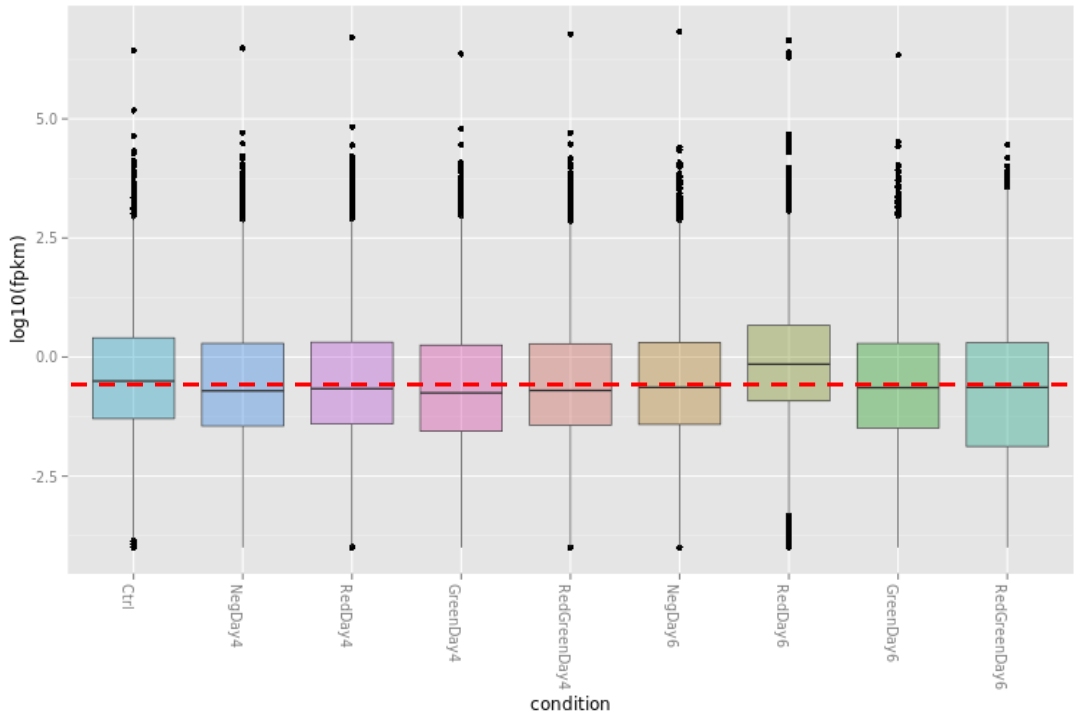


Figure 4.4 – Boxplot displaying the distribution of the expression values of the samples from our dataset. The distribution is similar between all samples. As highlighted by the red visual aid line, we can see that the median of the RedDay6 population deviates from those of the other populations, corroborating what was observed in the clustering shown in the previous figure for its replicate number 1 (RedDay6_1).

cluster.

As aforementioned, the preliminary analysis has shown that the individual cell populations were not clustering together (only time was effectively clustering the samples). Accordingly, to investigate the pureness of these cell populations we decided to verify the accuracy of the FAC Sorting by counting the total number of dsRed and eGFP transcripts present in the RNA-seq dataset. These results (Figure 4.) unequivocally show that there is no pure $G^{-}R^{+}$ populations in our samples. Consequently, downstream analysis were conducted exclusively using the populations for which there were congruent results, namely the populations $G^{+}R^{-}$ at day 4 of differentiation and $G^{+}R^{-}$, $G^{+}R^{+}$ at day 6 of differentiation.

Identification and functional network analysis of differentially expressed genes in different cardiac progenitor populations

To identify the genes differentially expressed in our cardiac progenitor populations ($G^{+}R^{-}$ day 4 and $G^{+}R^{-}$, $G^{+}R^{+}$ day 6) we used as baseline the

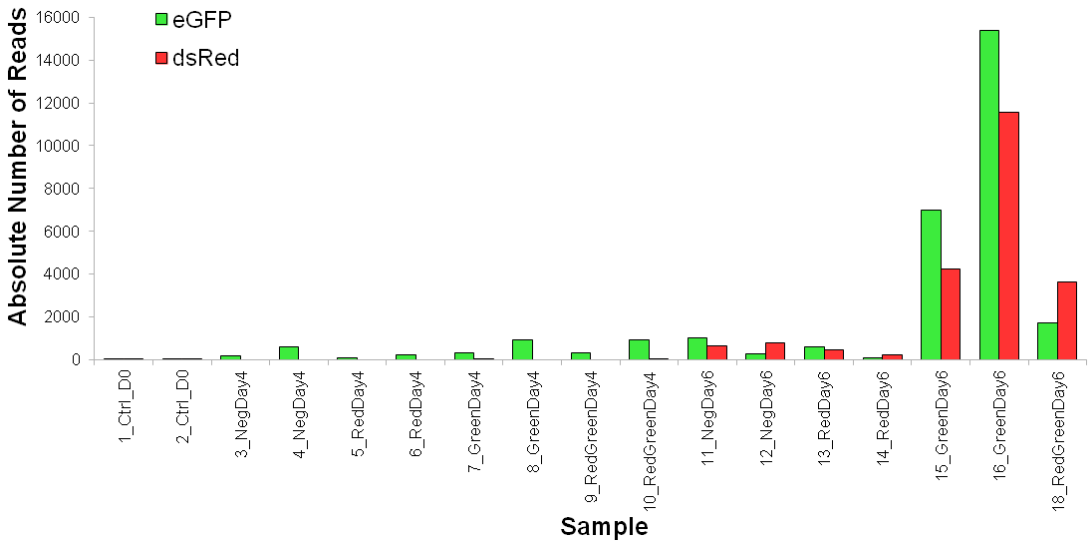


Figure 4.5 – Absolute dsRed and eGFP transcripts count from RNA-seq dataset. Counting the number of dsRed and eGFP transcripts present in each sorted population let us investigate its pureness. The numbers unequivocally show that there are no pure red populations at days 4 and 6 of differentiation. Also, although the number of eGFP transcripts in the $G^{+}R^{-}$ population at day 6 is high, dsRed transcripts seem to be also very enriched leading us to hypothesize that this population could be in fact $G^{+}R^{+*}$. Please note that the number of counts is absolute for each sample, directly depending on the number of cells sorted. This means that comparisons are only meaningful between red and green counts within the same sample, but not between the different samples.

expression levels of the control population G^-R^- for the corresponding differentiation day. The results are summarized in two different Venn diagrams: Figure 4.6 uses G^-R^- day 4 as baseline, and Figure 4.7 uses G^-R^- day 6 as baseline. These show the results from the intersections obtained from the set of exclusive genes up-regulated in the “all versus all" population analysis. These sets of genes represent the unique genetic signature of each sorted population, since all genes present in more than one comparison were excluded. The supplementary Table 4.4 - 4.6 in annex, present the full lists containing the genes that are exclusively up-regulated in each population. Table 4.4 shows the gene that is up-regulated in the G^+R^- population at day 4; Table 4.5 contains the genes up-regulated in the G^+R^- population at day 6; and Table 4.6 lists the genes up-regulated in the G^+R^+ population at day 6. This approach identified the exclusive set of up-regulated genes for each cell population (when compared to the control double negative population). Subsequently, we proceeded with the annotation of these genes, namely by finding (1) its attributed molecular functions (if any) described in public databases, (2) the biological processes they are involved in, and (3) the potential functional pathways where they play a role. After the annotation, we performed functional enrichment analysis on our gene lists in order to investigate the presence of functions enriched in our gene lists, i.e. to discover which biological functions and/or pathways are common to sub-groups of genes in our lists. For that, we used the online tool DAVID (http://david.abcc.ncifcrf.gov), using its resources to query the KEGG Pathway database (http://www.kegg.jp) and the Gene Ontology (GO) (http://www.geneontology.org) Biological Process and Molecular Function databases.

When looking at the G^+R^- population at day 4 we can identify only one exclusive gene - March6 - whose expression is up-regulated when compared to the baseline expression values of the G^-R^- population on the same day of differentiation (Table 4.4). Interestingly this gene has not been associated with any cardiac function. In the G^+R^- population at day 6 of differentiation one can identify 75 genes exclusively up-regulated compared to the baseline expression of the control population from the same day of differentiation.

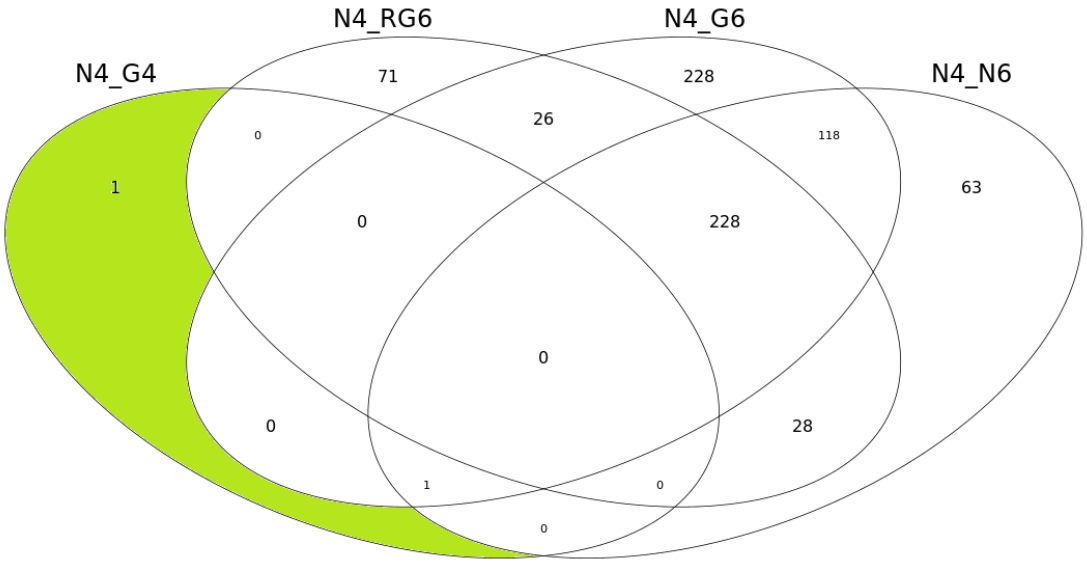


Figure 4.6 – Venn diagram highlighting the number of genes exclusively up-regulated in the G^+R^- population at day 4 of differentiation. Only gene March6 is exclusively up-regulated in the G^+R^- population at day 4 of differentiation, using as baseline the values of the control population from the same day of differentiation.

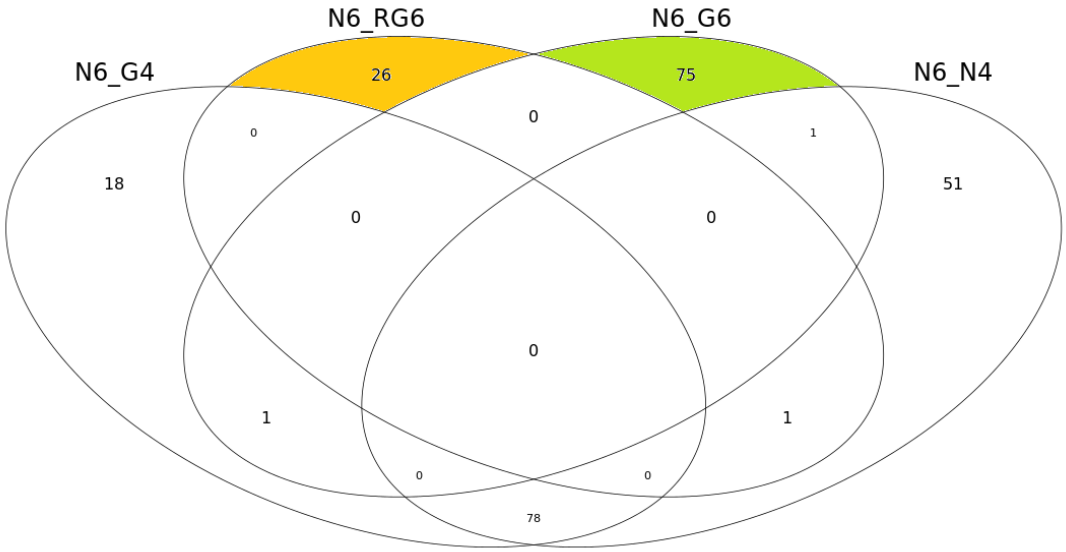


Figure 4.7 – Venn diagram highlighting the number of genes exclusively up-regulated in the G^+R^- and G^+R^+ populations at day 6 of differentiation. As depicted in this scheme there are 75 exclusive genes up-regulated in the G^+R^- population and 26 exclusive genes up-regulated in the G^+R^+ population at day 6 of differentiation, using as baseline the values of the control population from the same day of differentiation in both cases.

Through functional enrichment analysis the clusters with better q-value (which is the p-value adjusted for multiple testing) grouped several of these genes as being involved in cardiac functions. Table 4.2 shows the top 5 results for enriched biological functions in our gene list from the green day6 population. Accordingly, (1) heart development tops the list with 15 out of the 75 genes, i.e. 19,7% of the genes, followed by (2) cardiac muscle contraction with 10 genes, (3) hypertrophic cardiomyopathy with 10 genes, (4) myofibril assembly with 7 genes and (5) dilated cardiomyopathy with 10 genes. Figure 4.8 represents the cardiac muscle contraction pathway, which is the most significant biological process, grouping 10 genes from our list. Interestingly, among the 75 up-regulated genes we can identify *Nkx2.5*, corroborating the cardiac identity of these cells. Hence these results are the proof of principle that we were able to sort a population of cells with enriched cardiac biological functions. However we can also identify *Mef2c* as being up-regulated in this population of cells, which is consistent with the presence of dsRed transcripts described in Figure 4.. This suggests that this population of cells is not exclusively *Nkx2.5*$^{+}$ FHF progenitors but rather a mixture with double positive *Nkx2.5*$^{+}$/*Mef2c*$^{+}$ SHF progenitors, i.e. $G^{+}R^{-}$ with $G^{+}R^{+.}$

Database	Term	#genes	%genes	*p*-value	*q*-value
GO-BP	Heart development	15	19,7	2.0E-14	2.8E-11
KP	Cardiac muscle contraction	10	13,2	5.6E-10	1.5E-8
KP	Hypertrophic cardiomyopathy	10	13,2	5.6E-10	2.9E-8
GO-BP	Myofibril assembly	7	9,2	7.8E-9	5.2E-8
KP	Dilated cardiomyopathy	10	13,2	8.6E-10	6.8E-8

Table 4.2 – List of enriched biological functions found in the up-regulated gene list exclusively from the $G^{+}R^{-}$population at day 6 of differentiation. GO-BP: Gene Ontology Biological Process; KP: KEGG Pathway.

Database	Term	#genes	%genes	*p*-value	*q*-value
GO-MF	Carbohydrate binding	5	20	0,00164	0,0537

Table 4.1 – List of enriched biological functions found in the up-regulated exclusive genes from the $G^{+}R^{+}$population at day 6 of differentiation. GO-MF: Gene Ontology Molecular Function.

The same enrichment analysis was applied to investigate the functional annotation of the 26 genes exclusively up-regulated in the G^+R^+ population of cells at day 6 of differentiation. Contrasting with the previous results, none of these genes are significantly associated with a cardiac biological function. Interestingly, Table 4.1 lists carbohydrate binding as the only biological process found to be enriched in this gene list, grouping 5 of the identified genes, however with no statistical significance (*q*-value = 0,0537).

Given that the main aim of this study was the identification of novel genes potentially involved in cardiogenesis, it was important to isolate from differentiating mouse ESCs different populations of cardiac progenitor cells. Regarding the G^+R^+ population at day 6 of differentiation, on the one hand we can identify the presence of dsRed and eGFP transcripts in this population

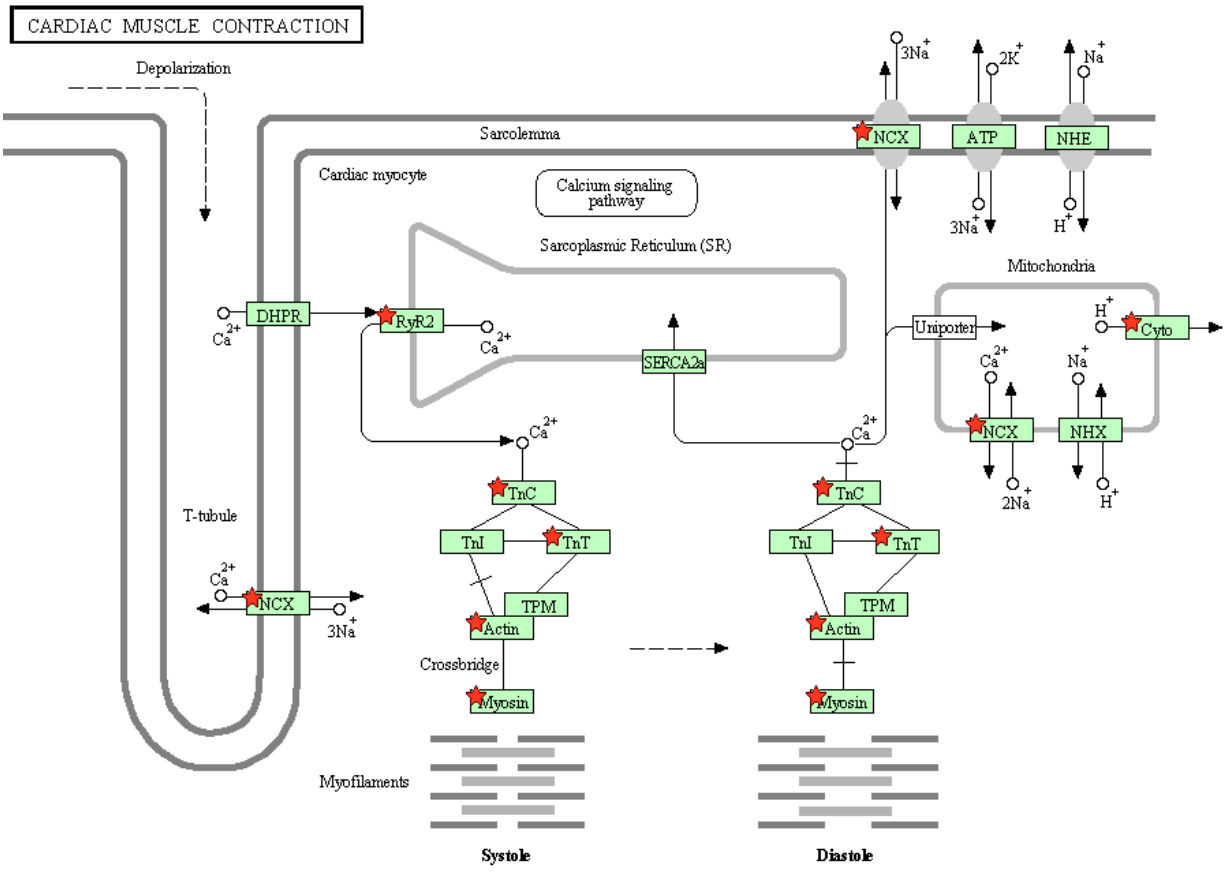


Figure 4.8 – Cardiac Muscle Contraction pathway, which is significantly enriched in the list of up-regulated genes in the G^+R^- population at day 6 of differentiation. The Cardiac Muscle Contraction is one of the biological processes which groups 13,2% of the genes present in our list. The red stars highlight the genes from our list, most notably TnC and TnT. *Scheme retrieved from the KEGG Pathway database.*

(Figure 4.5), but on the other hand neither *Nkx2.5* or *Mef2c* are found to be significantly up-regulated when compared to the control population, suggesting that these cells cannot be considered SHF progenitors. Indeed, the fact that all the significantly up-regulated genes do not have any cardiac functional annotation is consistent with this population not being representative of a cardiac progenitor population.

The G^+R^- population at day 4 of differentiation displays the same pattern, since there is a lack of up-regulated *Nkx2.5* and the only gene that is found to be significantly up-regulated is not annotated to have a cardiac function.

These findings led us to focus further analyses solely on the dataset corresponding to the G^+R^- population from day 6 of differentiation, which shows strong association with cardiac development, despite, as mentioned before it being most likely composed of a mixed population of cardiac progenitors.

Pre-validation of exclusively up-regulated candidate genes in the G^+R^- population at day 6 of differentiation

The detailed analysis of the up-regulated genes in the G^+R^- population from day 6 of differentiation showed 71 genes with known functional annotation, of which 49 are present in the Gene Ontology Biological Process (GO-BP) database. As aforementioned in Table 4.1 the biological process with best enrichment score is "Heart Development". Accordingly, when one analyses the overall biological processes to which the genes are associated, the results seem consistent with this finding, where 37 out of the total 75 genes are annotated to be involved in at least one of the following processes: cardiac or heart development, muscle development, blood and vasculature development and ion transport. From all the other genes with no association with any of these processes, we selected a number of candidates for experimental validation by prioritizing the gene list using the following criteria: log 2 fold change expression above 3, *q*-value below 0.05, and relevant molecular function keywords. The resulting candidate gene list is presented on Table 4.3. Next, we engaged in verifying which of these genes had published expression patterns either by RNA *in situ* hybridization and other RNA expression data. This analysis yielded relevant data for *Asb2,*

Gene	log2 fold change	*p-value*	*q-value*	Function Keywords
Asb2	5,30441	5,65283	5,00E-05	Proteolysis
Rpph1	4,8952	4,48402	5,00E-05	N/A
Cck	4,14957	3,42346	1,00E-04	Secreted hormone
Lca5l	3,41157	4,1799	5,00E-05	Coiled coil
Diras2	3,37906	3,86454	5,00E-05	Membrane associated
Synpo2l	3,13446	4,04153	5,00E-05	Protein binding
Lad1	3,08541	2,61011	5,00E-05	Extracellular matrix
3632451O06Rik	2,12134	3,05876	0,00015	Transmembrane

Table 4.3 – List of candidate genes selected from the significantly up-regulated genes in G^+R^- population at day 6 of differentiation, with log 2 fold change above 3 and not associated with cardiac, vascular or muscle development

Cck and 3632451O06Rik, whose functions are highlighted in the following paragraphs.

According to the Genecards database, *Asb2* encodes a protein member of the ankyrin repeat and SOCS box-containing (ASB) family, which has a role in protein degradation by coupling suppressor-of-cytokine-signalling (SOCS) proteins with the elongin BC complex. Functionally it is associated to retinoic acid-induced growth inhibition and differentiation of myeloid leukemia cells. This gene presents several alternatively spliced transcript variants, encoding multiple isoforms. Despite the fact that no function as yet been associated with cardiac development, its expression can be transiently found in mouse embryonic hearts at E9.5-10.5 (Terami et al., 2007). Accordingly, in the Genecards database is found RNA expression data for this gene with enriched expression in human heart and muscle tissues.

Cck is a gene which encodes for cholecystokinin, which is a neuropeptide secreted by intestinal cells and neurons in response to a meal. It acts as a gut hormone that regulates pancreatic enzyme secretion and gastrointestinal motility, and modulating the satiety signal. It can be cleaved in a variety of biologically active forms including CCK-33, CCK-8 (CCK octapeptide), CCK-39 and CCK-58. Its expression in the developing mouse embryo can be found as early as of E11.5 in the brain, however its physiological role is yet unknown (Giacobini et al., 2008). Interestingly, even though the available RNA *in situ*

data does not identify its expression in the developing heart region, RNA-seq data found on the Genecards database, shows that it is expressed in human heart and skeletal muscle, consistent with our findings of being enriched in ESC-derived cardiac progenitor cells.

We also selected the gene *3632451O06Rik* as a very promising candidate to study since it encodes an uncharacterized protein, with a strong expression in the developing heart at E9.5 – 10.5 (Terami et al., 2007), whose potential role in cardiogenesis has never been identified. One of its features is that the protein has a transmembrane and extracellular domain that could be very promising as a biomarker, enabling the isolation of cardiac progenitor cells.

For the other genes present in Table 4.4, the available data is not clear regarding its expression in the developing heart. Nonetheless, to validate whether any of these genes, including *Asb2, Cck* and *3632451O06Rik* have specific roles in early cardiogenesis, functional *in vitro* and *in vivo* analysis are required. The first approach would be the experimental quantitative PCR validation of its expression levels in embryonic hearts and cardiac progenitor cells. Henceforth one could engage in developing loss and gain of function assays to modulate the gene expression in these models, clarifying their potential involvement in cardiogenesis.

4.4. DISCUSSION AND CONCLUSION

It is a primary goal for translational medicine applications to improve the use of the genetic properties that characterize specific cell types found in the human body. One efficient and high-throughput approach to do so is by performing genome-wide analysis for the fast identification of genetic signatures of some of these cells. This technology intends in part to ease the isolation of specific cell populations that can be used for transplantation in patients and disease treatments.

In our analysis we aimed at investigating some of the genetic networks underlying embryonic cardiogenesis to identify novel genes or isoforms potentially involved in this important developmental process. By using a double

transgenic ESCs line we prompted to isolate ESC-derived FHF and SHF cardiac progenitors at days 4 and 6 of differentiation, at which a whole transcriptome analysis was performed to assess the expression of all genes and find potentially novel genetic players involved.

One major goal of this task was to probe the existence of a genetic signature that characterizes the different populations of cardiac progenitor cells at different time points. From the quality control analysis performed on our RNA-seq dataset, the hierarchical clustering grouped all the isolated cell populations in three major branches, separated time wise (Figure 4.3). Since the ESCs differentiation seems to recapitulate the early steps of *in vivo* developmental processes, these results might be related to the fact that during development many genes have a short and specific time window of expression. Additionally, the fact that the lineages' replicates do not cluster together suggests that the difference between lineages is very subtle and not fully grasped by an overall clustering analysis. However, when we were led to count the number of dsRed and eGFP transcripts (Figure 4.5) in the G^-R^+ population at day 6 due to its contrasting clustering, we could unequivocally confirm that there was not a pure red population in our samples. This prompted us to retrospectively analyze the data collected from the sorting experiments. From this analysis one can contrast the gates *de facto used* for cell sorting in the upper right panel of Figure 4.1 (P2-P5) with the gates that *could have* been applied to sort the cells (bottom two panels of Figure4.1 gates P2'-P5'). This 'logicle' representation follows the recommendations found in the literature (Herzenberg et al., 2006) and enables the user to objectively detect the presence of different cell populations within a sample. By doing so it allows the collection gates to be set around the cell population mass centre, discarding the probability of collecting rather mixed cell populations. Indeed, this representation clearly does not detect the existence of a pure G^-R^+ either at days 4 or 6 of ESCs differentiation as we do not observe the presence of a strong cell population on gate P5'. The discrepancies found in setting the sorting gates help to explain the fact that we cannot detect higher number of exclusively enriched expressed genes in each of the sorted populations, as most likely, the populations were not pure.

However it is now clear how these limitations on the FAC Sorting step have impacted our results. However, our analyses delivered a set of up-regulated genes enriched for development and cardiac functions in the mixed cell population of FHF and SHF isolated at day 6 of differentiation. This clearly corroborates their cardiac identity as many of the up-regulated genes are key cardiac transcription factors such as *Nkx2.5, Mef2c, Myocd, Lbh, Csrp3* and *Smyd1* (Table 4.5). At the same time, a set of candidate genes with no known cardiac annotation was also identified with having up-regulated expression in these ESC-derived cardiac progenitors, of which the most interesting are found in Table 4.3. Remarkably, when searching for gene expression data for our candidate genes in the literature and on the available databases, we found that screening experiments performed by independent labs, had identified some of our selected genes to be enriched in other populations of ESC-derived cardiac cells. For example, Terami and colleagues had previously identified through microarrays *Nebl, Asb2, Diras2, 3632451O06Rik, Rcsd1* and *Popdc2* to be richly expressed in ESC-derived cardiomyocytes (Terami et al., 2007). In this experiment the authors used a transgenic Nkx2.5/eGFP mouse ESC line, from which they isolated cardiomyocytes from differentiating EBs and searched for cardiogenesis-associated genes. In agreement with this approach we were also able to detect by RNA-seq – a more recent technology for whole-transcriptome screening analysis – some of the genes they have highlighted as interesting. Moreover, we detected significant differential expression for 3 potentially novel cardiogenesis-associated genes, namely *Asb2, Cck* and *3632451O06Rik*. These candidates hold great promise, and their functional involvement in cardiogenesis should be further investigated experimentally.

Chapter V

General Discussion and Future Perspectives

GENERAL DISCUSSION AND FUTURE PERSPECTIVES

Due to the complexity of heart disease and to its increasing incidence and prevalence in developed countries, it is urgent to fully understand all the molecular pathways involved in heart development in order to generate effective therapies. The scope of the present work aimed at exploring the role of *Ccbe1* in cardiac differentiation of mouse ESCs. The investigation of the role of *Ccbe1* had been mainly focusing on lymphangiogenesis, but a possible role in cardiogenesis deserved further investigation. A report from our lab has described that *Ccbe1* is expressed in the early cardiogenic fields of the developing heart in mouse embryos as early as E7.0 (Facucho-Oliveira et al., 2011). Later, we demonstrated that modulating *Ccbe1* expression in chicken embryos led to incorrect fusion of the developing heart tube, which was incompatible with life due to the severity of the resulting defects (Furtado et al., 2014). These data are in agreement with what has been described in some HS patients, which show congenital heart defects including hypertrophic cardiomyopathy and ventricular septal defects (Alders et al., 2009; Connell et al., 2010; Frosk et al., 2015). This syndrome is however caused by different mutations in the human *CCBE1* gene, which may have different impact on the function of the protein and therefore be compatible with life. Moreover, CCBE1 was found to interact with proteins with no associated function in lymphangiogenesis, suggesting additional roles in other biological processes (Jeltsch et al., 2014). Here, we demonstrated that *Ccbe1* is required for the early cardiac commitment of differentiating mouse ESCs.

As described earlier, a report from our lab has associated *CCBE1* with cell proliferation on cardiac progenitors during avian heart development. In this study, our lab has shown that *CCBE1* morphants present a decreased proliferation rate in the splanchnic and pharyngeal mesoderm regions (Furtado et al., 2014). Other experiments performed in our lab using MEFs isolated from KO *Ccbe1* embryos revealed that the disruption of this gene also leads to a decrease in the proliferation when compared to MEFs isolated from wild type littermates (*unpublished data*). In line with this, one of the defects that stood out during the differentiation of the generated *Ccbe1* KD ESCs lines presented here, besides the impaired commitment of differentiating ESCs in cardiac

mesoderm progenitors, was a strikingly reduced proliferation of the differentiating cells leading to smaller EBs. At first, this proliferation defect could not be clearly dissociated from being the cause of the impairment in cardiac mesoderm formation, but experiments with supplementation of *Ccbe1* blocking antibody and recombinant CCBE1 suggested that these defects are independent from each other. Indeed, supplementing differentiating WT EBs with α-*Ccbe1* antibody led to a clear reduction in the expression of cardiac mesoderm markers at day 4 of differentiation, even though the size of the EBs was preserved. In contrast, when supplementing the differentiating *Ccbe1* KD ESCs lines with recombinant CCBE1, we were able to partially rescue the size of the KD EBs, but the expression of the cardiac mesoderm markers remained downregulated. These data suggest that those defects are independent from each other, but are intimately related to the disruption of *Ccbe1*, placing CCBE1 as a direct regulator of cell proliferation and cardiac mesoderm specification during ESC differentiation.

As mentioned above, supplementation of recombinant CCBE1 in differentiating *Ccbe1* KD lines was unable to rescue the expression of the cardiac markers *Mesp1* and *Isl1*, but could partially rescue the size of the EBs. This may perhaps be explained by the fact that the concentrations of purified human recombinant CCBE1 that were supplemented were too low, or even that the purified protein was not fully functional. This latter hypothesis brings to light some of the possible limitations that are presently being optimized on the CCBE1 protein production and purification protocols. One big challenge of working with uncharacterized proteins is that it is hard to establish functional assays to assure both the quality and functionality of the produced protein. This may be major limitation as it could result that dysfunctional forms of a protein are being used in experimental assays and lead to false negative or biased results. In this particular case of the CCBE1 protein, a recent report showing a physical interaction with ADAMTS3, leading to the processing of pro-VEGF-C in VEGF-C, was a major breakthrough (Jeltsch et al., 2014). This knowledge will allow now the development of functional assays that will serve as optimal functional assays and quality controls in the production of functional CCBE1 protein. Nonetheless, prior to these recent advances, we used the *Ccbe1* KD

ESCs clones to supplement with the purified CCBE1 protein produced by our lab, to assess if the purified protein was functional and could rescue the phenotypes caused by the absence of *Ccbe1*. As aforementioned, the protein was able to partially rescue the phenotype caused by the absence of *Ccbe1*, suggesting that the purified protein preserved its function at least to some extent. Another challenge related to the production of proteins is the choice of appropriate production systems, i.e. cell types and culture protocols, which will meet the requisites allowing all needed post-translation modifications on the produced proteins. Indeed, post-translation modifications such as glycosylation or enzymatic processing are known to affect the function of the proteins (Parekh, 1991; Creighton, 1993), and as predicted for the mouse and human CCBE1 forms (Figure 1.), there are two different glycosylation sites. Therefore, it is possible that if the purified CCBE1 lacks the normal glycosylation pattern its function can be somewhat affected. Alternatively, since it has been suggested that CCBE1 can be proteolytically processed by ADAMTS3 (Jeltsch et al., 2014), it is possible that the full length protein that we provided to the differentiating ESCs still had to be processed, or cleaved to become fully active, and consequently rescue all the resulting defects caused by the disruption *of Ccbe1.*

Despite the protein production limitations discussed above, the production and purification of the full recombinant human CCBE1 obtained for this work set unprecedented advances in the study of the function of this protein. So far, the literature involving the production and use of CCBE1 protein had stressed many limitations on the production and purification protocols (Bos et al., 2011) This led to the production and analysis in an lymphangiogenesis assay of a truncanted form of CCBE1 that lacks the collagen domains (Bos et al., 2011), which have recently been shown to be absolutely pivotal during lymphangiogenesis (Roukens et al., 2015). Generation of mutant mice and zebrafish with targeted deletions of the different predicted functional domains of CCBE1, actually allowed to discriminate the roles of the different domains of the CCBE1 protein during lymphangiogenesis (Roukens et al., 2015). Nonetheless, final protocol optimizations and scaling-up the production and purification of CCBE1 in our lab, will allow us to take even further the study of this protein. For

example, it will possibly allow us to determine the role of the full-length protein as an ECM component, the mechanisms underlying its involvement on cell proliferation and, more interestingly, its potential as a cardiogenic inducing factor during ESC differentiation.

The *in vitro* experiments performed throughout this work highlighted the need of *Ccbe1* during early stages of cardiac commitment from differentiating ESCs. With impaired formation of early cardiac mesoderm progenitors upon constitutive KD of *Ccbe1,* it becomes impossible to determine the role of *Ccbe1* at the later stages of cardiac differentiation using this model. As described in the literature, however, *Ccbe1* is expressed at later stages of mammalian cardiac development (Facucho-Oliveira et al., 2011), and both chicken embryos and HS patients present cardiac defects upon disruption of *CCBE1* (Alders et al., 2009; Connell et al., 2010; Connell et al., 2012; Furtado et al., 2014). This is a strong indication that *Ccbe1* has a role at later stages of cardiogenesis. To study this possible role of *Ccbe1* at later stages of ESCs cardiac differentiation, we could e.g. generate ESCs lines that would allow the disruption of this gene at specific time points during ESC differentiation. One way of doing this could be by generating inducible *Ccbe1* KD ESCs lines. This would allow us to understand whether *Ccbe1* affects the full maturation *in vitro* of existing cardiac progenitors and/or beating cardiomyocytes past its role in cardiac mesoderm formation.

From the analysis of our data, we postulated that the role of CCBE1 during cardiac commitment of differentiating ESCs seems to be independent of its known function in promoting the maturation of pro-VEGF-C. This brings into perspective a whole new set of possible signaling pathways being activated and working together with CCBE1 to promote cardiac mesoderm formation. However, whether during cardiac differentiation CCBE1 acts alone or has other interacting partners that function to promote cardiac mesoderm formation remains unknown. Moreover, it is also not understood how the downregulation of *Ccbe1* affects the expression of *Vegf-C* itself. To address these questions it would be interesting to perform a co-immunoprecipitation assay to identify novel interacting partners for CCBE1 to better understand what other signaling pathways or signaling partners couple with CCBE1 in the cardiac commitment context.

Ultimately, the goal of our lab is to understand if CCBE1 could be used to promote cardiomyocyte commitment from differentiating ESCs, to use them in the context of regenerating myocardial tissue. The use of cardiomyocytes as a source of cells for transplantation to regenerate damaged tissue is reaching unprecedented potential. However, to keep growing this potential it is necessary to create and optimize safe and ethically acceptable protocols to produce large-scale human mature cardiomyocytes. Ideally, genetic manipulation of the implanted cells should be avoided, and the best alternative is to use secreted factors that are able to effectively drive cardiomyocyte differentiation form ESCs. CCBE1 is a secreted protein, which makes it an interesting candidate to study. To test whether CCBE1 promotes cardiomyocyte commitment, we could generate an ESC line with inducible *Ccbe1* overexpression to evaluate the effects of *Ccbe1* gain-of-function. In contrast to the experiments with the supplementation of recombinant CCBE1, this strategy could possibly circumvent any dosage limitations as we would have CCBE1 being directly overexpressed by the differentiating ESCs.

One big limitation of using cell-based therapies in cardiovascular regenerative medicine is the lack of a long-lasting homing of the cells on the injured areas (Taghavi and George, 2013). However, recent developments on the use of such therapies have shown that pre-treating cells with several factors prior to their transplantation can greatly improve the cell-homing to the injured areas, therefore enhancing their therapeutic efficacy. One of such factors is Col IV which is shown to extend the paracrine beneficial effects of the transplanted cells and to help the maturation of the transplanted PSCs into cardiomyocytes (Li et al., 2015). Another factor, identified as Ro-31-8425, is shown to increase firm adhesion to intercellular adhesion molecule 1 (ICAM-1) enabling targeted delivery of systemically infused cells to the injured areas (Levy et al., 2015). Together, these advances may improve the clinical outcomes of cell-based therapies. Whether CCBE1 could also act as well as such a molecule is unknown and hence it would be of great interest to explore this matter further.

Other promising breakthroughs in this field are related to long term culture of PSCs-derived cardiomyocytes. Indeed a report from Lundy and colleagues shows how maintaining long term cultures of cardiomyocytes derived from

PSCs contributes to the acquisition of an adult-like cardiomyocyte phenotype due to the structural and contractile changes that these cells undergo using their protocols (Lundy et al., 2014). To be able to control cardiac commitment from different PSC sources and to use fully mature cardiac committed cells in the context of regenerative medicine it is pivotal that key growth and cardiogenic factors are identified and studied. Differential screenings are a valuable tool for the identification of potentially important genes required for the correct development of the cardiac lineages. *Ccbe1* was one of such genes, but there are still other recently identified gene candidates that also seem to be very promising, like *Asb2, Cck* and *3632451O06Rik*. These candidates hold great promise, they have been already shown to be expressed in the heart and their functional involvement in cardiogenesis should be investigated further. This brings into consideration that more important than just identifying genes expressed in cardiogenic populations, is to understand and clarify how the different cardiogenic factors already identified interact with each other, and how these can be manipulated to optimize the existing protocols used to derive cardiac cells from PSCs, and ultimately improve the available cardiac therapies in the regenerative medicine.

References

References

Abranches, E., Guedes, A.M., Moravec, M., Maamar, H., Svoboda, P., Raj, A. and Henrique, D. (2014) Stochastic NANOG fluctuations allow mouse embryonic stem cells to explore pluripotency. Development *141,* 2770-2779.

Abu-Issa, R., Waldo, K. and Kirby, M.L. (2004) Heart fields: one, two or more? Dev. Biol. *272,* 81–85.

Andrée, B., Duprez, D., Vorbusch, B., Arnold, H.H. and Brand, T. (1998) BMP-2 induces ectopic expression of cardiac lineage markers and interferes with somite formation in chicken embryos Mech. Dev. *70*, 119–131.

Aguirre A., Sancho-Martinez I. and Izpisua Belmonte J.C. (2013) Reprogramming toward heart regeneration: stem cells and beyond. Cell Stem Cell. *12,* 275–284

Alders, M., Hogan, B.M., Gjini, E., Salehi, F., Al-Gazali, L., Hennekam, E.A., Holmberg, E.E., Mannens, M.M., Mulder, M.F., Offerhaus, G.J.*, et al.* (2009). Mutations in CCBE1 cause generalized lymph vessel dysplasia in humans. Nature genetics *41*, 1272-1274.

Astin, J.W., Haggerty, M.J., Okuda, K.S., Le Guen, L., Misa, J.P., Tromp, A., Hogan, B.M., Crosier, K.E. and Crosier, P.S. (2014) Vegfd can compensate for loss of Vegfc in zebrafish facial lymphatic sprouting. Development. *141(13)*, 2680-2690.

Barton, C.A., Gloss, B.S., Qu, W., Statham, A.L., Hacker, N.F., Sutherland, R.L., Clark, S.J. and O'Brien, P.M. (2010) Collagen and calcium-binding EGF domains 1 is frequently inactivated in ovarian cancer by aberrant promoter hypermethylation and modulates cell migration and survival. Br. J. Cancer *102(1)*, 87-96.

Battista, S., Guarnieri, D., Borselli, C., Zeppetelli, S., Borzacchiello, A., Mayol, L., Gerbasio, D., Keene, D.R., Ambrosio, L., and Netti, P.A. (2005). The effect of matrix composition of 3D constructs on embryonic stem cell differentiation. Biomaterials *26*, 6194-6207.

Bento, M., Correia, E., Tavares, A.T., Becker, J.D., and Belo, J.A. (2011). Identification of differentially expressed genes in the heart precursor cells of the chick embryo. Gene expression patterns : GEP *11*, 437-447.

Black, D.L. (2003) Mechanisms of alternative pre-messenger RNA splicing. Annu. Rev. Biochem. *72*, 291-336.

Bos, F.L., Caunt, M., Peterson-Maduro, J., Planas-Paz, L., Kowalski, J., Karpanen, T., van Impel, A., Tong, R., Ernst, J.A., Korving, J., van Es, J.H., Lammert, E., Duckers, H.R. and Schulte-Merker, S. (2011). CCBE1 is essential for mammalian lymphatic vascular development and enhances the lymphangiogenic effect of vascular endothelial growth factor-C in vivo. Circ. Res. *109*, 486-491.

Bowers, S.L. and Baudino, T.A. (2010) Laying the groundwork for growth: Cell-cell and cell-ECM interactions in cardiovascular development. Birth Defects Res C Embryo Today *90*, 1-7.

Boyle, A.J., Schulman, S.P. and Hare, J.M. (2006) Stem Cell Therapy for Cardiac Repair - Ready for the Next Step. Circulation *114,* 339-352.

Brade, T., Pane, L.S., Moretti, A., Chien, K.R., and Laugwitz, K.L. (2013) Embryonic heart progenitors and cardiogenesis. Cold Spring Harb. Perspec. Med. *3,* a013847.

Bratt-Leal, A.M., Carpenedo, R.L., and McDevitt, T.C. (2009). Engineering the embryoid body microenvironment to direct embryonic stem cell differentiation. Biotechnology progress *25*, 43-51.

Buckingham, M., Meilhac, S. and Zaffran, S. (2005) Building the mammalian heart from two sources of myocardial cells. Nature Reviews Genetics *6*, 826-837.

Bumgarner, R. (2013) Overview of DNA microarrays: types, applications, and their future. Curr. Protoc. Mol. Biol.Chapter 22:Unit 22.1.

Burger, N.B., Bekker, M.N., Kok, E., De Groot, C.J., Martin, J.F., Shou, W., Scambler, P.J., Lee, Y., Christoffels, V.M., and Haak, M.C. (2015). Increased nuchal translucency origins from abnormal lymphatic development and is independent of the presence of a cardiac defect. Prenatal diagnosis. *35(6*):517-528.

Burggren, W.W. and Keller, B. (1997) Development of Cardiovascular Systems: Molecules to Organisms. New York: Cambridge University Press.

Cao, F., Wagner, R.A., Wilson, K.D., Xie, X., Fu, J.D., Drukker, M., Lee, A., Li, R.A., Gambhir, S.S., Weissman, I.L.*, et al.* (2008). Transcriptional and functional profiling of human embryonic stem cell-derived cardiomyocytes. PloS one *3*, e3474.

Carlson, B.M. (2014) Human embryology and developmental biology, Fifth Edition. Philadelphia: Elsevier Saunders.

Chambers, I., Silva, J., Colby, D., Nichols, J., Nijmeijer, B., Robertson, M., Vrana, J., Jones, K., Grotewold, L. and Smith, A. (2007) Nanog safeguards pluripotency and mediates germline development. Nature *450*,1230-1234.

Connell, F., Kalidas, K., Ostergaard, P., Brice, G., Homfray, T., Roberts, L., Bunyan, D.J., Mitton, S., Mansour, S., Mortimer, P.*, et al.* (2010). Linkage and sequence analysis indicate that CCBE1 is mutated in recessively inherited generalised lymphatic dysplasia. Human genetics *127*, 231-241.

Correia, C., Serra, M., Espinha, N., Sousa, M., Brito, C., Burkert, K., Zheng, Y., Hescheler, J., Carrondo, M.J.T., Šarić, T. and Alves, P.M. (2014) Combining hypoxia and bioreactor hydrodynamics boosts induced pluripotent stem cell differentiation towards cardiomyocytes. Stem Cell Rev. and Rep. *10,* 786–801.

Coucouvanis, E. and Martin, G.R. (1999) BMP signaling plays a role in visceral endoderm differentiation and cavitation in the early mouse embryo. Development *126*, 535-546.

Creighton, E.T. (1993). Proteins: Structures and Molecular Properties, second edition. New York: W H Freeman and Company.

Czyz, J. and Wobus, A. (2001). Embryonic stem cell differentiation: the role of extracellular factors. Differentiation; research in biological diversity *68*, 167-174.

Davis, D.R. and Stewart, D.J. (2011) 'Autologous cell therapy for cardiac repair'. Expert. Opin. Biol. Ther. *11,* 489-508.

DeHaan R. L. (1963). Migration patterns of the precardiac mesoderm in the early chick embryo. Exp. Cell Res. *29*, 544–560.

Dias, S., Choy, M., Alitalo, K., and Rafii, S. (2002). Vascular endothelial growth factor (VEGF)-C signaling through FLT-4 (VEGFR-3) mediates leukemic cell proliferation, survival, and resistance to chemotherapy. Blood *99*, 2179-2184.

Doetschman, T.C., Eistetter, H., Katz, M., Schmidt, W. and Kemler, R. (1985) The in vitro development of blastocyst-derived embryonic stem cell lines: formation of visceral yolk sac, blood islands and myocardium. Journal of Embryology and Experimental Morphology *87*, 27-45.

Domian, I.J., Chiravuri, M., van der Meer, P., Feinberg, A.W., Shi, X., Shao, Y., Wu, S.M., Parker, K.K., and Chien, K.R. (2009). Generation of functional ventricular heart muscle from mouse ventricular progenitor cells. Science *326*, 426-429.

Facucho-Oliveira, J., Bento, M., and Belo, J.A. (2011). Ccbe1 expression marks the cardiac and lymphatic progenitor lineages during early stages of mouse development. Int J Dev Biol *55*, 1007-1014.

Foldes, G., Harding, S.E., and Ali, N.N. (2008). Cardiomyocytes from embryonic stem cells: towards human therapy. Expert opinion on biological therapy *8*, 1473-1483.

Francois, M., Caprini, A., Hosking, B., Orsenigo, F., Wilhelm, D., Browne, C., Paavonen, K., Karnezis, T., Shayan, R., Downes, M., et al. (2008). Sox18 induces development of the lymphatic vasculature in mice. Nature *456*, 643-647.

Frosk, P., Chodirker, B., Simard, L., El-Matary, W., Hanlon-Dearman, A., Schwartzentruber, J., Majewski, J.; FORGE Canada Consortium, Rockman-Greenberg, C. (2015) A novel CCBE1 mutation leading to a mild form of hennekam syndrome: case report and review of the literature. BMC Med. Genet.*16*, 28.

Furtado, J., Bento, M., Correia, E., Inácio, J.M. and Belo, J.A. (2014) Expression and function of Ccbe1 in the chick early cardiogenic regions are required for correct heart development. PLoS One *9(12)*, e115481.

Galvagni, F., Pennacchini, S., Salameh, A., Rocchigiani, M., Neri, F., Orlandini, M., Petraglia, F., Gotta, S., Sardone, G.L., Matteucci, G.*, et al.* (2010). Endothelial cell adhesion to the extracellular matrix induces c-Src-dependent VEGFR-3 phosphorylation without the activation of the receptor intrinsic kinase activity. Circ Res *106*, 1839-1848.

Garbern, J.C., Mummery, C.L. and Lee, R.T: (2013) Model Systems for Cardiovascular Regenerative Biology. Cold Spring Harb Perspect Med *3*, a014019.

George, E.L., Georges-Labouesse, E.N., Patel-King, R.S., Rayburn, H. and Hynes, R.O. (1993) Defects in mesoderm, neural tube and vascular development in mouse embryos lacking fibronectin. Development (Camb.) *119*, 1079-1091.

Giacobini, P. and Wray, S. (2008) Prenatal expression of cholecystokinin (CCK) in the central nervous system (CNS) of mouse. Neurosci. Lett. *438*, 96-101.

Gnecchi, M. and, Melo, L.G. (2009) Bone marrow-derived mesenchymal stem cells: isolation, expansion, characterization, viral transduction, and production of conditioned medium. Methods Mol Biol. *482*, 281-94.

Goh, S.K., Olsen, P. and Banerjee, I. (2013) Extracellular matrix aggregates from differentiating embryoid bodies as a scaffold to support ESC proliferation and differentiation. PLoS One *8*:e61856.

Hagerling, R., Pollmann, C., Andreas, M., Schmidt, C., Nurmi, H., Adams, R.H., Alitalo, K., Andresen, V., Schulte-Merker, S., and Kiefer, F. (2013). A novel multistep mechanism for initial lymphangiogenesis in mouse embryos based on ultramicroscopy. The EMBO journal *32*, 629-644.

Hansson, E.M. and Lendahl, U. (2013). Regenerative medicine for the treatment of heart disease. Journal of internal medicine *273*, 235-245.

Hashimoto, H. and Yuasa, S. (2013). Testosterone induces cardiomyocyte differentiation from embryonic stem cells. Journal of molecular and cellular cardiology.

Hastings, C.L., Roche, E.T., Ruiz-Hernandez, E., Schenke-Layland, K., Walsh., C.J., and Duffy, G.P. (2014) Drug and cell delivery for cardiac regeneration. Adv. Drug Deliv. Rev. *84*, 85-106.

Higuchi, S., Lin, Q., Wang, J., Lim, T.K., Joshi, S.B., Anand, G.S., Chung, M.C., Sheetz, M.P., and Fujita, H. (2013). Heart extracellular matrix supports cardiomyocyte differentiation of mouse embryonic stem cells. Journal of bioscience and bioengineering *115*, 320-325.

Hogan, B.M., Bos, F.L., Bussmann, J., Witte, M., Chi, N.C., Duckers, H.J., and Schulte-Merker, S. (2009). Ccbe1 is required for embryonic lymphangiogenesis and venous sprouting. Nature genetics *41*, 396-398.

Holubec, T., Caliskan, E., Sündermann, S.H., Starck, C.T., Plass, A., Bettex, D., Falk, V. and Maisano, F. (2014) The use of extracellular matrix patches in cardiac surgery. J. Card. Surg. *2, 1*45-148.

Ichihara, Y., Shinoka, T., Matsumura, G., Ikada, Y. and Yamazaki, K. (2015) A new tissue-engineered biodegradable surgical patch for high-pressure systems. Interact.. Cardiovasc. Thorac. Surg. *6,* 768-776.

INE – Instituto Nacional de Estatística, I.P. (2013). Portugal – Doenças cérebro-cardiovasculares em números – 2013. Lisboa: Autor.

Jakus, Z., Gleghorn, J.P., Enis, D.R., Sen, A., Chia, S., Liu, X., Rawnsley, D.R., Yang, Y., Hess, P.R., Zou, Z., et al. (2014). Lymphatic function is required prenatally for lung inflation at birth. The Journal of experimental medicine *211*, 815-826.

Jeevanantham V., Butler M., Saad A., Abdel-Latif A., Zuba-Surma E.K. and Dawn B. (2012) Adult bone marrow cell therapy improves survival and induces long-term improvement in cardiac parameters: a systematic review and meta-analysis. Circulation *126(5)* 551-568.

Jeltsch, M., Jha, S.K., Tvorogov, D., Anisimov, A., Leppänen, V.M., Holopainen, T., Kivelä, R., Ortega, S., Kärpanen, T. and Alitalo, K. (2014) CCBE1 enhances lymphangiogenesis via A disintegrin and metalloprotease with thrombospondin motifs-3-mediated vascular endothelial growth factor-C activation. Circulation *129(19)*, 1962-1971.

Jessup, M., and Brozena, S. (2003). Heart failure. The New England journal of medicine *348*, 2007-2018.

Jing, D., Parikh, A., Canty, J.M., Jr., and Tzanakakis, E.S. (2008). Stem cells for heart cell therapies. Tissue engineering Part B, Reviews *14*, 393-406.

Keller, G. (2005). Embryonic stem cell differentiation: emergence of a new era in biology and medicine. Genes & development *19*, 1129-1155.

Keller, G.M. (1995). In vitro differentiation of embryonic stem cells. Current opinion in cell biology *7*, 862-869.

Kelly, R.G., Buckingham, M.E. and Moorman, A.F.(2001) Heart fields and cardiac morphogenesis. Cold Spring Harb. Perspect. Med. *3,* a015750.

Khezri, S., Valojerdi, M.R., Sepehri, H., and Baharvand, H. (2007). Effect of basic fibroblast growth factor on cardiomyocyte differentiation from mouse embryonic stem cells. Saudi medical journal *28*, 181-186.

Kim, J., Shapiro, L. and Flynn, A. (2015) The Clinical Application of Mesenchymal and Cardiac stem cells as a therapy for cardiovascular disease. Pharmacology & Therapeutics *151,* 8-15.

Kreuger, J., Nilsson, I., Kerjaschki, D., Petrova, T., Alitalo, K., and Claesson-Welsh, L. (2006). Early lymph vessel development from embryonic stem cells. Arteriosclerosis, thrombosis, and vascular biology *26*, 1073-1078.

Kukk, E., Lymboussaki, A., Taira, S., Kaipainen, A., Jeltsch, M., Joukov, V., and Alitalo, K. (1996). VEGF-C receptor binding and pattern of expression with VEGFR-3 suggests a role in lymphatic vascular development. Development *122*, 3829-3837.

Laflamme, M.A. and Murry, C.E. (2011) Heart regeneration. Nature. *473,* 326–335.

Laugwitz, K.L., Moretti, A., Caron, L., Nakano, A., and Chien, K.R. (2008). Islet1 cardiovascular progenitors: a single source for heart lineages? Development 135, 193-205.

Le Guen, L., Karpanen, T., Schulte, D., Harris, N.C., Koltowska, K., Roukens, G., Bower, N.I., van Impel, A., Stacker, S.A., Achen, M.G., et al. (2014). Ccbe1 regulates Vegfc-mediated induction of Vegfr3 signaling during embryonic lymphangiogenesis. Development 141, 1239–1249.

Levy, O., Mortensen, L.J., Boquet, G., Tong, Z., Perrault, C., Benhamou, B., Zhang, J., Stratton, T., Han, E., Safaee, H., Musabeyezu, J., Yang, Z., Multon, MC., Rothblatt, J., Deleuze, JF., Lin, C.P. and Karp. J.M. (2015) A Small-Molecule Screen for Enhanced Homing of Systemically Infused Cells. Cell Reports. *In press.*

Lewis, S.L. and Tam, P.P. (2006) Definitive endoderm of the mouse embryo: formation, cell fates, and morphogenetic function. Dev Dyn. *235,* 2315-2329.

Li, R.K., Jia, Z.Q., Weisel, R.D., Mickle, D.A.G., Zhang, J., Mohabeer, M.K., Rao, V. and Ivanov, J. (1996) Cardiomyocyte transplantation improves heart function. Ann. Thorac. Surg. *62,* 654-661.

Li, X., Chen, Y., Schéele, S., Arman, E., Haffner-Krausz, R., Ekblom, P., Lonai, P. (2001) Fibroblast growth factor signaling and basement membrane assembly are connected during epithelial morphogenesis of the embryoid body. J. Cell Biol. *153*, 811-822.

Liersch, R., Nay, F., Lu, L., and Detmar, M. (2006). Induction of lymphatic endothelial cell differentiation in embryoid bodies. Blood *107*, 1214-1216.

Lundy, S.D., Gantz, J.A., Pagan, C.M., Filice, D. and Laflamme M.A. (2014) Pluripotent stem cell derived cardiomyocytes for cardiac repair. Curr Treat Options Cardiovasc Med. *Jul;16(7)*:319.

Makkar, R.R., Smith, R.R., Cheng, K., Malliaras, K., Thomson, L.E., Berman, D., Czer, L.S., Marbán, L., Mendizabal, A., Johnston, P.V., Russell, S.D., Schuleri, K.H., Lardo, A.C., Gerstenblith, G. And Marbán, E. (2012) Intracoronary cardiosphere-derived cells for heart regeneration after myocardial infarction (CADUCEUS): a prospective, randomised phase 1 trial. Lancet. *379*, 895-904.

Martínez-Ramos, C., Rodríguez-Pérez, E., Garnes, M.P., Chachques, J.C., Moratal, D., Vallés-Lluch, A. and Pradas M.M. (2014) Design and assembly procedures for large-sized biohybrid scaffolds as patches for myocardial infarct. Tissue Eng. Part C Methods *20*, 817-827.

Meilhac, S.M., Lescroart, F., Blanpain, C. and Buckingham, M.E. (2014) Cardiac cell lineages that form the heart. Cold Spring Harb. Perspect. Med. *3,* a013888.

Mishima, K., Watabe, T., Saito, A., Yoshimatsu, Y., Imaizumi, N., Masui, S., Hirashima, M., Morisada, T., Oike, Y., Araie, M., et al. (2007). Prox1 induces lymphatic endothelial differentiation via integrin alpha9 and other signaling cascades. Molecular biology of the cell *18*, 1421-1429.

Mommersteeg, M.T., Dominguez, J.N., Wiese, C., Norden, J., de Gier-de Vries, C., Burch, J.B., Kispert, A., Brown, N.A., Moorman, A.F., and Christoffels, V.M. (2010). The sinus venosus progenitors separate and diversify from the first and second heart fields early in development. Cardiovascular research *87*, 92-101.

Nakazawa, K., Yoshiura, Y., Koga, H. and Sakai, Y. (2013) Characterization of mouse embryoid bodies cultured on microwell chips with different well sizes. J. Biosci. Bioeng. *116*, 628-33.

Nardi, N.B. and Meirelles, L.S. (2006) Mesenchymal stem cells: isolation, in vitro expansion and characterization. Handb Exp Pharmacol. *174*, 249-282.

Nemer, G., and Nemer, M. (2001). Regulation of heart development and function through combinatorial interactions of transcription factors. Annals of medicine *33*, 604-610.

NHBLI – National Heart, Lung and Blood Institute (2012). Retrieved from: http://www.nhlbi.nih.gov/health/health-topics/topics/ht/risks

Parekh, R.B. (1991) Mammalian cell gene expression: protein glycosylation. Current Opinion in Biotechnology. *Oct;2(5),* 730-734.

Puceat, M. (2008). Protocols for cardiac differentiation of embryonic stem cells. Methods *45*, 168-171.

Radisic, M. and Christman, K.L. (2013) Materials science and tissue engineering: repairing the heart. Mayo Clin. Proc. *88(8),* 884-898.

Roukens, M.G., Peterson-Maduro, J., Padberg, Y., Jeltsch, M., Leppänen, V.M., Bos, F.L., Alitalo, K., Schulte-Merker, S. and Schulte, D. (2015) Functional Dissection of the CCBE1 Protein: A Crucial Requirement for the Collagen Repeat Domain. Circ. Res. *116(10)*, 1660-1669.

Sachlos, E. and Auguste, D.T. (2013) Embryoid body morphology influences diffusive transport of inductive biochemicals: a strategy for stem cell differentiation. Biomaterials. *29*, 4471–4480.

Sakakibara, Y., Nishimura, K., Tambara, K., Yamamoto, M., Lu, F., Tabata, Y. and Komeda, M. (2002) Prevascularization with gelatin microspheres containing basic fibroblast growth factor enhances the benefits of cardiomyocyte transplantation. J. Thorac. Cardiovasc. Surg. *124,* 50-56.

Sanganalmath, S.K. and Bolli, R. (2013) Cell therapy for heart failure - A comprehensive overview of experimental and clinical studies, current challenges, and future directions. Circ Res. *113*:810-834.

Sato, H., Takahashi, M., Ise, H., Yamada, A., Hirose, S., Tagawa, Y., Morimoto, H., Izawa, A., and Ikeda, U. (2006). Collagen synthesis is required for ascorbic acid-enhanced differentiation of mouse embryonic stem cells into cardiomyocytes. Biochemical and biophysical research communications *342*, 107-112.

Schenke-Layland, K., Nsair, A., Van Handel, B., Angelis, E., Gluck, J.M., Votteler, M., Goldhaber, J.I., Mikkola, H.K., Kahn, M. And Maclellan, W.R. (2011) Recapitulation of the embryonic cardiovascular progenitor cell niche. Biomaterials. *32(11)*, 2748-2756.

Schoenwolf, G.C., Bleyl, S.B., Brauer, P.R. and Francis-West, P.H.(2015) Larsen's human embryology, fifth edition. Philadelphia: Elsevier Saunders.

Schultheiss, T.M., Burch, J.B.E. and Lassar, A,B. (1997). A role for bone morphogenetic proteins in the induction of cardiac myogenesis. Genes Dev. *11*, 451– 462.

Senyo, S.E., Steinhauser, M.L., Pizzimenti, C.L., Yang, V.K., Cai, L., Wang, M., Wu, T.D., Guerquin-Kern, J.L., Lechene, C.P. and Lee RT. (2013) Mammalian heart renewal by pre-existing cardiomyocytes. Nature *493(7432),* 433-436.

Shin, J.W., Huggenberger, R., and Detmar, M. (2008). Transcriptional profiling of VEGF-A and VEGF-C target genes in lymphatic endothelium reveals endothelial-specific molecule-1 as a novel mediator of lymphangiogenesis. Blood *112*, 2318-2326.

Taghavi, S. And George, J.C. (2013) Homing of stem cells to ischemic myocardium. Am. J. Transl. Res. *4,* 404–411.

Takahashi, T., Lord, B., Schulze, P.C., Fryer, R.M., Sarang, S.S., Gullans, S.R., and Lee, R.T. (2003). Ascorbic acid enhances differentiation of embryonic stem cells into cardiac myocytes. Circulation *107*, 1912-1916.

Takahashi, K.and Yamanaka, S. (2006) Induction of pluripotent stem cells from mouse embryonic and adult fibroblast cultures by defined factors. Cell *126(4)*:663-676.

Taylor-Weiner, H., Schwarzbauer, J.E., and Engler, A.J. (2013). Defined extracellular matrix components are necessary for definitive endoderm induction. Stem Cells *31*, 2084-2094.

Terami, H., Hidaka, K., Shirai, M., Narumiya, H., Kuroyanagi, T., Arai, Y., Aburatani, H. and Morisaki, T. (2007) Efficient capture of cardiogenesis-associated genes expressed in ES cells. Biochem. Biophys. Res. Commun. *355,* 47-53.

Valente, M., Nascimento,D.S., Cumano, A. and Pinto-do-Ó, P. (2014) Sca-1+ Cardiac Progenitor Cells and Heart-Making: A Critical Synopsis. Stem Cells Dev. *19,* 2263–2273.

Weitze, G. (2006) Embryonic stem cell-derived embryoid bodies: an in vitro model of eutherian pregastrulation development and early gastrulation. Handb. Exp. Pharmacol. *174*, 21-51.

Yamamoto, F. and Yamamoto, M. (2007) Scanning copy number and gene expression on the 18q21-qter chromosomal region by the systematic multiplex PCR and reverse transcription-PCR methods. Electrophoresis. *28(12)*,1882-1895.

Zaffran, S., Kelly, R.G., Meilhac, S.M., Buckingham, M.E. and Brown, N.A. (2004) Right ventricular myocardium derives from the anterior heart field. Circ. Res. *95*, 261–268.

Zeng, D., Ou, D.B., Wei, T., Ding, L., Liu, X.T., Hu, X.L., Li, X., and Zheng, Q.S. (2013). Collagen/beta(1) integrin interaction is required for embryoid body formation during cardiogenesis from murine induced pluripotent stem cells. BMC cell biology *14*, 5.

Zhou, B., von Gise, A., Ma, Q., Rivera-Feliciano, J., and Pu, W.T. (2008). Nkx2-5- and Isl1-expressing cardiac progenitors contribute to proepicardium. Biochemical and biophysical research communications *375*, 450-453.

Zuk, P.A., Zhu, M., Mizuno, H., Huang, J., Futrell, J.W., Katz, A.J., Benhaim, P., Lorenz, H.P. and Hedrick, M.H. (2001) Multilineage cells from human adipose tissue: implications for cell-based therapies. Tissue Eng. *2*, 211-28

Annexes

Table 4.4 - The only up-regulated exclusive gene in the G^+R^- cardiac progenitor population at day 4 of differentiation, using as baseline the gene expression values of the G^-R^- population at day 4.

Official Gene Symbol	Name
March6	SPHK1 interactor, AKAP domain containing

Table 4.5 – List of up-regulated exclusive genes in the G^+R^- cardiac progenitor population at day 6 of differentiation, using as baseline the gene expression values of the G^-R^- population at day 6.

Official Gene Symbol	Name
Sphkap	SPHK1 interactor, AKAP domain containing
Klhl30	kelch-like 30 (Drosophila)
Rpph1	ribonuclease P RNA-like 3; ribonuclease P RNA component H1; ribonuclease P RNA-like 2; predicted gene 6093
Lbh	limb-bud and heart
Erbb4	v-erb-a erythroblastic leukemia viral oncogene homolog 4 (avian)
Amph	amphiphysin
Plcxd3	phosphatidylinositol-specific phospholipase C, X domain containing 3
Hspb7	heat shock protein family, member 7 (cardiovascular)
Sh3bgr	similar to putative SH3BGR protein; SH3-binding domain glutamic acid-rich protein
Nrxn3	neurexin III
Ap1s2	adaptor-related protein complex 1, sigma 2 subunit
Nebl	nebulette
Tnnt2	troponin T2, cardiac
Synpo2l	synaptopodin 2-like
Cdkn1c	cyclin-dependent kinase inhibitor 1C (P57)
Lca5l	Leber congenital amaurosis 5-like
Myom1	myomesin 1
Ccnd2	cyclin D2
Mef2c	myocyte enhancer factor 2C
Myot	myotilin
Popdc2	popeye domain containing 2
Mybpc3	myosin binding protein C, cardiac
Ppargc1a	peroxisome proliferative activated receptor, gamma, coactivator 1 alpha
Smpx	small muscle protein, X-linked
Gyg	glycogenin
Diras2	DIRAS family, GTP-binding RAS-like 2
Vcan	versican

Akap2	A kinase (PRKA) anchor protein 2; paralemmin 2
Palm2	A kinase (PRKA) anchor protein 2; paralemmin 2
Myocd	myocardin
Myl7	myosin, light polypeptide 7, regulatory
Myh7	myosin, heavy polypeptide 7, cardiac muscle, beta
Rcsd1	RCSD domain containing 1
Ldb3	LIM domain binding 3
Med12l	mediator of RNA polymerase II transcription, subunit 12 homolog (yeast)-like
3632451O06Rik	RIKEN cDNA 3632451O06 gene
Actn2	actinin alpha 2
8430429K09Rik	RIKEN cDNA 8430429K09 gene
Tcap	titin-cap
Xirp1	xin actin-binding repeat containing 1
Lad1	ladinin
Cox6a2	cytochrome c oxidase, subunit VI a, polypeptide 2
Asb2	ankyrin repeat and SOCS box-containing 2
Hspb3	heat shock protein 3
Nkx2-5	NK2 transcription factor related, locus 5 (Drosophila)
Filip1	filamin A interacting protein 1
Smyd1	SET and MYND domain containing 1
Myl3	myosin, light polypeptide 3
Trpc7	transient receptor potential cation channel, subfamily C, member 7
Kcnh7	potassium voltage-gated channel, subfamily H (eag-related), member 7
Cck	cholecystokinin
Cpeb2	cytoplasmic polyadenylation element binding protein 2
Tnni1	troponin I, skeletal, slow 1
Cap2	CAP, adenylate cyclase-associated protein, 2 (yeast)
Rhoj	ras homolog gene family, member J
Clec9a	C-type lectin domain family 9, member a
Sh2d2a	SH2 domain protein 2A
Slc8a1	solute carrier family 8 (sodium/calcium exchanger), member 1
Prkg1	protein kinase, cGMP-dependent, type I
Unc45b	unc-45 homolog B (C. elegans)
Cdh5	cadherin 5
Lmod1	leiomodin 1 (smooth muscle)
Myh6	myosin, heavy polypeptide 6, cardiac muscle, alpha
Trim55	tripartite motif-containing 55
Tnnc1	troponin C, cardiac/slow skeletal
Gm5779	predicted gene 5779
Csrp3	cysteine and glycine-rich protein 3
Rbm24	RNA binding motif protein 24
Actc1	actin, alpha, cardiac muscle 1; similar to alpha-actin (AA 27-375)
Jph2	junctophilin 2

Sox7	SRY-box containing gene 7
Sorbs2	sorbin and SH3 domain containing 2
Ryr2	ryanodine receptor 2, cardiac
Gimap6	GTPase, IMAP family member 6
Mpped2	metallophosphoesterase domain containing 2
Myl2	myosin, light polypeptide 2, regulatory, cardiac, slow
Car8	carbonic anhydrase 8; similar to Carbonic anhydrase-related protein (CARP) (CA-VIII)

Table 4.6 – List of up-regulated exclusive genes in the G^+R^+ cardiac progenitor population at day 6 of differentiation, using as baseline the gene expression values of the G^-R^- population at day 6**.**

Official Gene Symbol	Name
Emp1	epithelial membrane protein 1
Mmp3	matrix metallopeptidase 3
Kcne4	potassium voltage-gated channel, Isk-related subfamily, gene 4
Cyp1b1	cytochrome P450, family 1, subfamily b, polypeptide 1
Ptgs2	prostaglandin-endoperoxide synthase 2
Nov	nephroblastoma overexpressed gene
Bgn	biglycan
Gpnmb	glycoprotein (transmembrane) nmb
Hoxd12	homeo box D12
Lgals1	lectin, galactose binding, soluble 1
Gldn	gliomedin
Serpinb2	serine (or cysteine) peptidase inhibitor, clade B, member 2
Serpine2	serine (or cysteine) peptidase inhibitor, clade E, member 2
Cp	ceruloplasmin
Agtr2	angiotensin II receptor, type 2
Cbr2	carbonyl reductase 2
Eif2s3y	eukaryotic translation initiation factor 2, subunit 3, structural gene Y-linked
Mmp10	matrix metallopeptidase 10
Cpe	carboxypeptidase E; similar to carboxypeptidase E
Anxa1	annexin A1
Col5a3	collagen, type V, alpha 3
Inhbb	inhibin beta-B
S100a6	S100 calcium binding protein A6 (calcyclin)
Grem1	gremlin 1
Evx2	even skipped homeotic gene 2 homolog