UAlg
UNIVERSIDADE DO ALGARVE
DEPARTAMENTO DE CIÊNCIAS BIOMÉDICAS E MEDICINA

Study of the requirement of the CCBE1 growth factor in the generation of cardiac myocytes from hES cells

MS.c Thesis in Biomedical Science

Rita Catarina Vaz Drago

UNIVERSIDADE DO ALGARVE
DEPARTAMENTO DE CIÊNCIAS BIOMÉDICAS E MEDICINA

Study of the requirement of the CCBE1 growth factor in the generation of cardiac myocytes from hES cells

MS.c Thesis in Biomedical Science

Rita Catarina Vaz Drago

Orientador:
Professor Doutor José António Belo

Co-orientador:
Doutora Andreia Bernardo

Faro, 2012

MS.c Thesis proposal in Biomedical Science
Area of Developmental Biology by the Universidade do Algarve
Study of the requirement of the CCBE1 growth factor in the generation of cardiac myocytes from hES cells.

Dissertação para obtenção do Grau de Mestre em Ciências Biomédicas
Área de Biologia do Desenvolvimento pela Universidade do Algarve
Estudo da função do factor de crescimento CCBE1 na diferenciação de miócitos cardíacos a partir de células estaminais embrionárias humanas

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

ACKNOWLEDGEMENTS

I thank my supervisor, Prof. José Belo, for the support he provided me during this long journey and for the opportunity of working at his laboratory.

I would like to thank Prof. Roger Pedersen as well for welcoming me in his team during my Erasmus visit.

I want to leave a message to my co-supervisors, Andreia Bernardo and João Facucho-Oliveira, for the patience, for the trust and of course for everything I have learned from them.

To my colleagues and friends, thank you for the good moments we passed in these last two years.

I want to thank to all my teachers and scientists who influenced me to pursue this amazing journey.

Last but not least, I want to thank to my sister, brother, father, and especially to my mom for being the most important part of my life. Obrigada.

LIST OF CONTENTS

ABBREVIATIONS

3’ - 3 prime
5’ - 5 prime
µM - Micro molar
µl – Microliter
°C - Degrees Cesius
a.a. - Aminoacid
Ab - Antibody
A, C, T, G, U – Adenine, Cytosine, Thymine, Guanine, Uracil
Bp – Base pairs
BMP4 - Bone Morphogenetic Protein 4
BRA - Brachyury
BSA - Bovine serum albumine
CCBE1 - Collagen and calcium-binding *Epidermal growth factor* (EGF) domain-containing protein 1
cDNA - coding DNA
CDM – Chemically defined medium
ChIP – Chromatin immune precipitation
ChIP-seq - chromatin immunoprecipitation and sequencing
Da - Dalton
DAG - Diacylglycerol
DNA – Dehoxyribonucleic acid
DNase - Deoxyribonuclease
DTT - Dithiothreitol
ECM - Extracellular Matrix
EGF - Epidermal growth factor
EGF - Epidermal growth factor receptor
EMT - Epithelial-mesenchymal transition
ERK - Extracellular-signal-Regulated Kinase
FHF – First heart field
DAPI - 4',6-diamidino-2-phenylindole
FB50 – FGF2 (50ng/ml) + BMP4 (50ng/ml)
FGF2 – Fibroblast Growth Factor 2
FLyB – FGF2(20ng/ml) + Ly (uM)+ BMP4(ng/ml)
GAPDH - Glyceraldehyde-3-phosphate dehydrogenase
GATA4 - GATA-binding protein 4
GFP - Green fluorescent protein
hESCs – human embryonic stem cells
H/HPC – Heart/ hemangioblast precursor lineage
ICM - Inner cell mass
IgG – Immunoglobulin G
IP – Immunoprecititation
ISL1 - Insulin gene enhancer
JAK - Janus kinase
KD – Knock down
KDR - Kinase insert domain receptor
KO – Knock out
LY 294002 - 2-(4-Morpholinyl)-8-phenyl-4H-1- benzopyran-4-one
MAPK - Mitogen activated protein kinase

MEK - mitogen-activated protein kinase kinase
mESCs mouse embryonic stem cells
MESP1/2 – Mesoderm posterior 1/2
min - Minutes
mL - Milliliter
MMP - metalloproteases
mRNA – messenger ribonucleic acid
NaAc – Sodium acetate
NaCl – Sodium chloride
ng – Nanogram
NKX2.5 - NK2 transcription-factor related, locus 5
OCT4 - Octamer-binding protein 4
OE – Over expression
PBS - Phosphate buffer saline
PBST - Phosphate buffered saline +0.01 % Triton X100
PI3K - Phosphatidylinositol-3-OH kinase
PLCγ - Phospholipase c-γ
PKC - Protein kinase-C
PMSF -Phenylmethylsulfonyl fluoride
PVA – Polyvinyl alcohol
RNase - Ribonuclease
rpm - Rotations per minute
RT - Room temperature
RTK – receptor tyrosine kinase
qPCR – semi-quantitative polymerase chain reaction
secs - Seconds
SHF - Secondary heart field
shRNA – short hairpin RNA
SOX2 - SRY (sex-determining region Y) box 2
STAT - Signal transducer and activator of transcription
TGF-β - Transforming growth factor β
TE – Tris and EDTA (Ethylenediaminetetraacetic acid)
WHO - World Health Organization

LIST OF FIGURES

LIST OF TABLES

ABSTRACT

In normal development, cells from the totipotent zygote give rise to the pluripotent inner cell mass and epiblast cells, which subsequently become more and more restricted, and ultimately, originate all the differentiated cells of the human body. In this Master thesis a chemically defined protocol based on signalling present *in vivo* was used to differentiate *in vitro* a human embryonic stem cell (hESCs) line, with the intent of recapitulating cardiac development.

Along the differentiation protocol, the role of CCBE1 (Collagen and Calcium-Binding EGF-like domain 1) was studied. CCBE1 is a growth factor which has already been reported in mouse as being expressed and potentially required for the generation of cardiac precursor cells. As such, this study focused on three main topics: the characterization, function and regulation of CCBE1 in hES cells.

Regarding characterization, CCBE1 shows an expression pattern similar to the one previously reported in mouse ES cells, i.e. it is highly expressed in pluripotency, when cells start to differentiate it is downregulated, and its mRNA and protein levels rise progressively as cells differentiate towards cardiac precursors. There is a strong association between mRNA and protein expressions, as determined by quantitative PCR analysis and immunocytochemistry assays.

As for CCBE1 function, this was assessed by knocking down (KD) CCBE1 in hESCs using shRNA. CCBE1 KD cells lost their ability to differentiate properly into cardiac precursors, as evidenced by the downregulation of cardiac precursors genes. Additionally, pluripotency markers genes were also deregulated upon CCBE1 KD.

The pattern of CCBE1 expression raised some questions about its regulation in pluripotency and cardiac precursor cells. To investigate this, Chromatin Immunoprecipitation (ChIP) assays were performed with an antibody against a pluripotency protein (NANOG), one against an early mesoderm protein (BRACHYURY), and one against a cardiac protein (ISL1). The results revealed binding of NANOG, BRA and ISL1 to distinct CCBE1 regulatory regions and at different differentiation timings.

This work contributed to a further understanding of the complex process of human heart development by studying CCBE1, which is required for differentiation of cardiac mesoderm precursors.

RESUMO

O estudo da biologia do desenvolvimento, tema central desta dissertação, estende-se desde o momento da fecundação até à completa constituição do organismo adulto.

A formação dos diferentes tecidos e órgãos, dos mais simples aos mais complexos, implica a ocorrência de diversos comportamentos celulares, tais como, divisão, crescimento, diferenciação, migração, adesão, arranjo espacial e apoptose.

No que toca ao desenvolvimento dos vertebrados, e em particular do homem, o primeiro órgão a estabelecer-se, pelas três semanas de gestação, é aquele que tem como função distribuir oxigénio e nutrientes a todas as partes do organismo, o coração.

As células mesodérmicas precursoras cardíacas, que detêm o potencial para gerar todos os tipos celulares que constituem o coração, são originadas durante a gastrulação, o primeiro processo embrionário de remodelação celular a partir do qual são originados os três folhetos germinativos, endoderme, mesoderme e ectoderme. O desenvolvimento do coração pode ser dividido em quatro etapas principais: formação do crescente cardíaco, no qual as células mesodérmicas se organizam em forma de crescente na região anterior do embrião; formação do coração linear, resultado do movimento ventral desempenhado pelas células precursoras; *looping* cardíaco, que consiste na reestruturação da forma linear do coração, desenhando um movimento em espiral; por fim, dá-se a remodelação do coração, em que por septação, se originam as aurículas e ventrículos.

Muito do que se sabe sobre o desenvolvimento cardíaco resulta principalmente de investigação realizada recorrendo ao uso em animais modelo, particularmente galinha e ratinho. No entanto, existem diferenças genéticas e de desenvolvimento que distinguem o homem das referidas espécies. Deste modo, as células estaminais embrionárias humanas, obtidas a partir de embriões excendentários, resultantes de tratamento de fertilização *in vitro*, são um óptimo modelo para o estudo do desenvolvimento, fisiologia e doenças cardíacos.

As células estaminais embrionárias derivam de embriões em estados de desenvolvimento muito precoces (morula ou blastocisto), são caracterizadas por terem o mesmo padrão genético que o daquelas que lhes deram origem, possuindo elevados níveis de expressão de OCT4, SOX2 e NANOG, a tríade que assegura o seu estado pluripotente. A pluripotência é definida como a capacidade detida por células individuais de se multiplicarem indefinidamente para originar outra célula pluripotente ou, em resposta a sinais provenientes do embrião (*in vivo*) ou da cultura celular (*in vitro*), se dividirem para originar todos os precursores celulares, que mais tarde originarão todos os tecidos do animal adulto.

A especificação das células precursoras cardíacas a partir das células pluripotentes é realizada essencialmente pela acção sinérgica de dois factores parácrinos, BMP (*Bone Morphogenetic Protein*) e FGF (*Fibroblast Growth Factor*).

Ainda que muitas questões acerca do desenvolvimento do coração já tenham sido respondidas, muitas outras persistem. Na tentativa de diminuir esta diferença, foi

desenvolvido um estudo focado na função de um factor de crescimento, já descrito em ratinho como importante na especificação cardíaca, ao longo da diferenciação de precursores cardíacos a partir de células embrionárias estaminais humanas.

Este factor de crescimento, designado por Ccbe1, foi identificado num *screening* genético realizado em galinha. CCBE1 deve o seu nome aos diferentes domínios que o constituem, sendo eles, domínio de ligação a colagénio e cálcio e a um receptor membranar do tipo EGF (*Epitelial Growth Factor*).

Reconhecendo o potencial, ainda pouco explorado, deste factor no desenvolvimento cardíaco precoce e das células estaminais como modelo experimental, foi desenhado um estudo que abordou essencialmente três vertentes; caracterização da expressão genética e proteica, função ao longo da diferenciação celular e regulação da expressão de CCBE1 por factores de transcrição já descritos.

As células estaminais humanas utilizadas neste trabalho foram isoladas em 1998 e pertencem à linha celular H9. Foram cultivadas em meio que favorece a pluripotência e diferenciadas em precursores cardíacos num meio quimicamente definido suplementado com FGF2 e BMP4.

Através de métodos de quantificação de mRNA (qPCR) e proteína (imunocitoquimica), foi descrito um padrão que sugere que CCBE1 desempenha uma função não apenas no processo de diferenciação cardíaca, mas também na manutenção da pluripotência, uma vez que tanto o transcrito como o produto proteico são grandemente expressos em células pluripotentes e precursores cardíacos.

No que toca ao estudo da função, foram geradas duas linhas *knock down* usando um vector que contém um shRNA, afectando a expressão de CCBE1 em cerca de 70%. Estas células foram usadas em ensaios de diferenciação celular, tendo sido medidos os níveis de mRNA dos genes com expressão característica de precursores cardíacos, tais como NKX2.5, BRA, GATA4, KDR. Os resultados mostram que a expressão desses genes é afectada, sendo negativamente regulados quando a expressão de CCBE1 é reduzida. Portanto, CCBE1 regula positivamente a diferenciação em precursores mesodérmicos cardíacos, comprometendo a sua eficiente especificação quando não está presente nos níveis fisiológicos. Além destas análises realizadas no contexto de diferenciação celular, foram também medidos os níveis de expressão de genes particularmente transcritos em pluripotência, nomeadamente, NANOG, OCT4 e SOX2. Também a expressão destes genes foi afectada, positivamente, para NANOG e OCT4, e negativamente em relação a SOX2.

De forma a explicar em termos de regulação ao nível do ADN o padrão de expressão apresentado por CCBE1, foram realizados diversos ensaios de imunoprecipitação da cromatina (ChIP). Por existir uma relação inversa entre o padrão de expressão de CCBE1 e BRA, colocou-se a hipótese do segundo ter algum efeito na regulação do primeiro. Uma base de dados de ChIP-seq (ChIP e sequenciação) revelou que há ligação de BRA a CCBE1 após 36h de diferenciação na linhagem cardíaca. Por não ser conhecida acção repressiva de BRA, a hipótese formulada propunha que BRA teria uma acção indutora sobre CCBE1, quando na presença de um co-activador. Deste modo, os factores de transcrição escolhidos para a realização de ChIP foram BRA, NANOG e ISL1.

Os resultados de ChIP revelam que há ligação de BRA e NANOG em pluripotência e ligação de BRA juntamente com ISL1 em células precursoras cardíacas. A ligação destes hipotéticos dímeros é feita em regiões reguladoras distintas, de acordo com o contexto celular. Com o objectivo de complementar este resultado, linhas *knock down* de BRA e ISL1 foram geradas. Os níveis de transcrição de CCBE1 nas células com expressão de BRA reduzida suportam a hipótese colocada, de que BRA terá um efeito indutor sobre CCBE1. No entanto, o mesmo não pode ser afirmado no caso de ISL1, já que não há alteração da expressão de CCBE1 em células *knock down* para aquele factor de transcrição. Por conseguinte, é necessário testar o efeito activador sobre CCBE1 de outros factores de transcrição que possuam uma função importante na caracterização de precursores cardíacos.

Concluindo, ficou provado pelos dados preliminares aqui apresentados que CCBE1 desempenha uma função essencial no processo de diferenciação cardíaca, nos seus estádios mais primários, assim como na manutenção do equilíbrio da pluripotência.

1. INTRODUCTION

1.1. Developmental Biology of the Heart

One of the most basic questions in developmental biology is how multicellular organisms develop from a single cell. During embryonic development, the egg is divided to give rise to many millions of cells that form structures such complex and varied as eyes, kidneys, heart or brain.

The first scientific approach to explain the process of development was introduced by Hippocrates in the fifth century BC. His thesis explained this process in terms of physical principles of heat and moisture.

Later on, another Greek philosopher named Aristoteles proposed two theories about development: pre-formation and epigenesis, but only the last one would be supported by the cellular theory developed by Mathias Schleiden and Theodor Schwann, in 1839

During the twentieth century, another concept was introduced to the developmental biology discipline, the genetics. It was proposed that genes control cell behaviour by regulating where and when proteins are synthesized, and thus together with the environmental signalling determine the main processes involved in development, namely, cell division, pattern formation, morphogenesis, cell differentiation, migration, death and growth.

This concept has been developed and further studied and it is now considered the basis of developmental biology research. Thousands of genes control the development of animals and a full understanding of the processes involved is far from being achieved. This Master's Thesis is an attempt to take a little step towards the full understanding of the development process and focuses, particularly in the development of the heart, studying the role of a gene in this complex and multistep process.

The heart is derived from the mesoderm, one of the three germ layers formed during gastrulation.

In mammalians, after fertilization, the initiating step in development, the egg goes through several mitotic divisions, a process called cleavage, giving rise to the morula, and later to the blastocyst (Fig.1.1.1). The morula cells are capable of forming all the embryonic and extraembryonic structures of the body and are thus said to be totipotent (Gilbert, 2003; Geens *et al.*, 2009). The blastocyst is composed of inner and outer cells; the inner cells are denominated the inner cell mass (ICM) and these cells can generate any cell type of the body. As such, these cells are no longer totipotent and are said to be pluripotent. (Fig. 1.1.2).

Due to its potential, pluripotent cells from the ICM are largely used in developmental biology research. These cells can be removed from the embryo in a manner that lets them retain their pluripotent characteristics, and thenceforth they are called embryonic stem cells (Zhang *et al.* 2006). This is an important topic that will be further discussed.

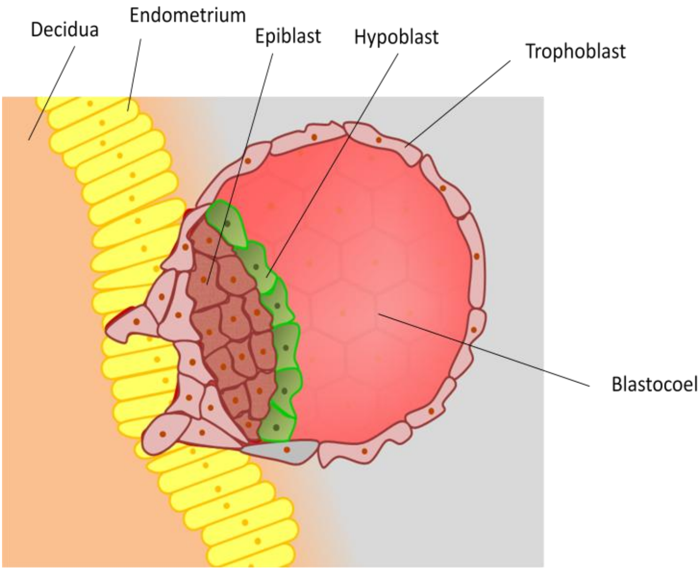


Fig. 1.1.1 – **Human embryo, 7 days of gestation.** Blastocyst, immediately prior to gastrulation, composed by the epiblast, hypoblast, trophoblast and blastocoel (adapted from Gilbert, 2003).

During the developmental process, the totipotent morula divides asymmetrically to form the blastocyst, which is composed of extraembryonic outer cells called trophoblast cells and the cells of the ICM. Once the blastocoel is formed, the resulting structure is called blastocyst. After the blastocyst is bound to the uterus, the first segregation of cells within the ICM begins. This rearrangement will give rise to two layers; the hypoblast (the lower layer) and the epiblast (the upper layer).

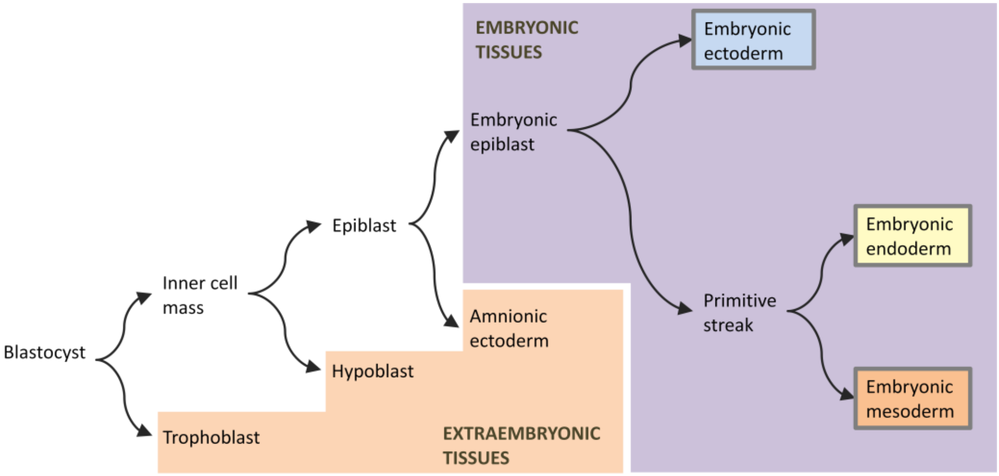


Fig. 1.1.2 – **Derivation of tissues in human embryo** (adapted from Gilbert, 2003).

Then, epiblast and hypoblast cells are reorganized, through a process called gastrulation, leading to the formation of the three germs layers: endoderm, mesoderm and ectoderm. Each of these layers constitutes the group of primary progenitors that will give rise to all tissues and organs of the embryo. This highly organised process of cell movement involves the whole embryo and it is first denoted by the formation of the Hensen's Node and the Primitive Streak (Gilbert, 2003; Keller, 2005; Ohta *et al.* 2010; Yang *et al*., 2002).

The primitive streak and the Hensen's node are transitory structures through which cells migrate, formed by accumulation followed by thickening of epiblast cells at the posterior region of the embryo. The primitive streak appears as a depression in the midline of the epiblast and the node is a regional thickening of cells at the anterior end of the streak. The primitive streak defines the axes of the embryo. It extends from posterior to anterior sides, separating the left to the right portion of the embryo (Gilbert, 2003; Keller, 2005; Ohta *et al.* 2010; Yang *et al*., 2002).

When the streak is reaching half-maximal extension, epiblast cells in the streak undergo an epithelial to mesenchymal transition (EMT) and start to move as individual cells into the space between the epiblast and the hypoblast to form mesoderm as well as definitive endoderm. As gastrulation progresses, the primitive streak elongates distally, with a continuously ingression of cells. Eventually, while cells of the anterior portion of the embryo are undergoing gastrulation, cells at the posterior end are already starting to form organs. Consequently the embryo exhibits a posterior-to-anterior gradient of development. At the end of gastrulation, all the endoderm and mesoderm cells have moved inwards and the remaining epiblast gives rise to the third germ layer, named ectoderm. (Lawson and Pedersen, 1991; Keller, 2005; Ohta *et al.* 2010; Yang *et al*., 2002). The three germ layers are the genesis for distinct parts of the body: the endoderm cells generating the foregut, liver and pharinge; the mesoderm cells giving rise to the kidneys, muscles, heart, skeleton and blood and lastly, the ectoderm forming the nervous system and the skin (Wolpert, 2007).

For the purpose of this work, the several steps of heart development, starting from mesoderm specification, will further be detailed bellow.

After migrating through the primitive streak, the mesoderm cells are arranged in a mediolateral axis by a BMP gradient (Gilbert, 2003, Tonegawa *et al.*, 1997). This different concentrations of BMP lead to the formation of different types of mesoderm, namely, the precordal plate, chordamesoderm, paraxial, intermediate and lateral plate (Fig. 1.1.3) (Gilbert, 2003).

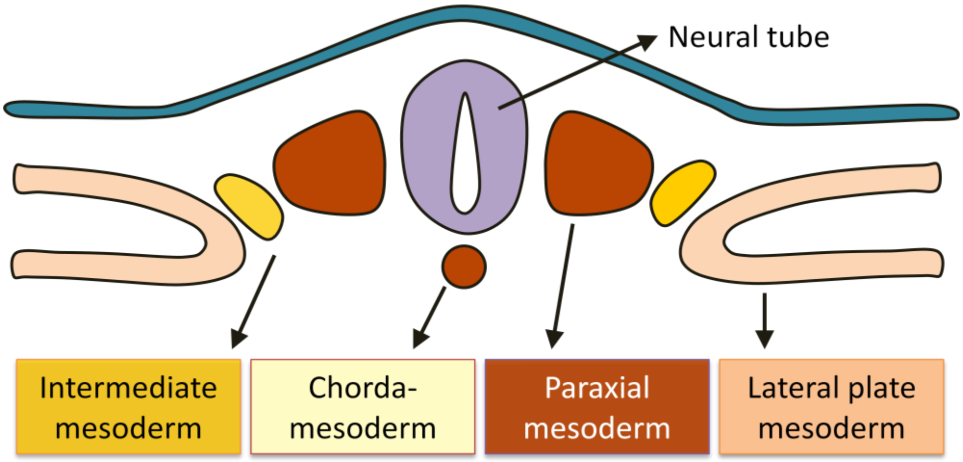


Fig. 1.1.3 – **The major lineages of amniote mesoderm**: intermediate, chorda, paraxial and lateral plate mesoderm. Precordal mesoderm is not represented because it is presented in a more anterior region (adapted from Gilbert, 2003).

The precordal plate mesoderm is located more anteriorly and forms much of the connective tissues and musculature of the face (Wolpert, 2007).

The chordomesoderm, which is positionated right in the middle of the embryo, establishes the anterior-posterior body axis and induces the formation of the neural tube, giving rise to the notochord at the future dorsal side of the embryo (Wolpert, 2007).

The paraxial mesoderm, also called somatic dorsal mesoderm, forms the somites, which produce the cartilage of the vertebrae and ribs, the muscle of the rib cage, limbs, abdominal wall, back and tongue, and the dermis of the dorsal skin (Wolpert, 2007).

The urogenital system has origin in the intermediate mesoderm (Wolpert, 2007).

Finally, with higher levels of BMP expression there is the lateral plate mesoderm, which once again is subdivided in three mesoderm types: splanchnic, somatic and extraembryonic. It is the splanchnic mesoderm with a more ventral position that gives rise to the heart, blood vessels and blood cells of the circulatory system. The somatic and extraembryonic mesoderms form the lining of the body cavities and extraembryonic membranes important for transporting nutrients to the embryo, respectively (Gilbert, 2003).

In vertebrates, the cardiogenic mesoderm arises from the anterior splanchnic lateral plate mesoderm. As mesodermal cells migrate through the streak to originate the presumptive heart mesoderm, they form two groups of cells at either sides of the Hensen's node (Redkar *et al.*, 2001). These two clusters take a lateral migratory path towards the cranio-lateral part of the embryo to form the cardiac crescent, which begins to express cardiac transcription factors (Harvey, 2002) (Fig. 1.1.4). This structure contains the precursors of the endothelial lining of the heart, cushions cells of the valves, atrial and ventricular myocytes, and Purkinje fiber (Gilbert, 2003) (Fig. 1.1.5).

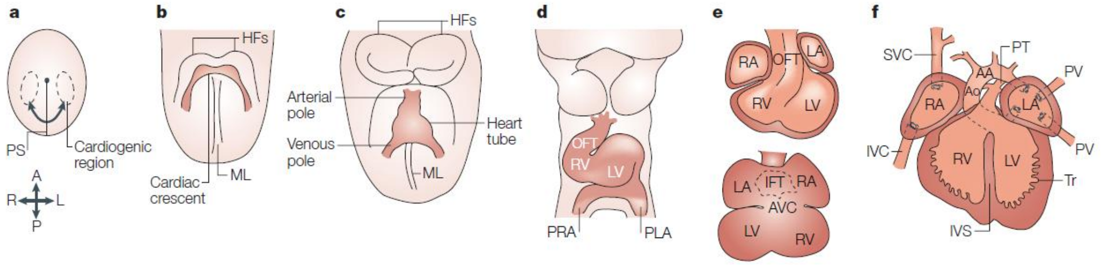


Fig. 1.1.4 – **Heart development of the mouse. a)** Cardiac progenitor cells originate in the primitive streak (PS), from where they migrate to the anterior region of the embryo at about embryonic day E6.5. **b)** In most vertebrates, the heart progenitor cells lie adjacent to the progenitors of the head (HF – head folds) and form the cardiac crescent, where differentiated myocardial cells are now observed (E7.5). **c)** The early cardiac tube forms through fusion of the cardiac crescent at the midline (ML) (E8). **d)** It subsequently undergoes looping (E8.5). **e)** By E10.5 the heart has acquired well-defined chambers, but is still a tube (upper panel, ventral view; lower panel, dorsal view). **f)** In the fetal heart (E14.5) the chambers are now separated as a result of septation and are connected to the pulmonary trunk (PT) and aorta (Ao), which ensure the separate pulmonary and systemic circulation of the blood, respectively, after birth.
Anterior (A)–posterior (P) and right (R)–left (L) axes are indicated. Right atrium (RA), right ventricle (RV), left atrium (LA), left ventricle (LV), AA (aortic arch), AVC (atrioventricular canal), IFT (inflow tract), IVC (inferior vena cava), IVS (interventricular septum), OFT (outflow tract), PLA (primitive left atrium), PRA (primitive right atrium), PV (pulmonary vein), SVC (superior vena cava), Tr (trabeculae) (from Buckingham *et al.*, 2005).

At the primitive streak stage there are transcription factors expressed transiently, such as Mesp1 and Mesp2 (mesoderm posterior 1 and 2) and Bra (Brachyury) and which are required for the movement of cells towards the anterior region of the embryo (Buckingham *et al.*, 2005).

Once in contact with the endoderm, specification of the cardiogenic mesoderm is induced through the BMP (bone morphogenic protein) and FGF (fibroblast growth factor) signaling pathways, which are produced by that ventral layer. BMP stimulates heart or blood development depending on its interaction with the Wnt pathway. When in presence of Wnt proteins at posterior region of the embryo, BMP promotes blood and inhibits heart formation. On the other hand, Wnt inhibitors, such as Cerberus, Dickkopf and Crescent, together with BMPs, both produced by the anterior endoderm, promote heart formation (Fig. 1.1.5). At the anterior region, Wnt and BMP signaling promote head formation (Harvey, 2002). Moreover, it is known that BMP also induces Fgf8 synthesis in the endoderm (Alsan and Schultheiss, 2002), which appears to be an inducer of the expression of cardiac proteins, including Gata4, Gata5, Gata6, NKX2.5 (NK2 transcription-factor related, locus 5), Mef2b and Mef2c (myocyte enhancer factor), Hand1 and Hand2 (heart and neural crest derivatives expressed transcript 1 and 2), Tbx5 (T-box 5) (Harvey, 2002 and Buckingham *et al.*, 2005). It is known as well, that BMP might also act in a concentration-dependent manner to induce or repress cardiogenisis (Alsan and Schultheiss, 2002). Working together, these transcription factors activate the expression of genes encoding cardiac muscle-specific proteins, such as cardiac actin, ANF (atrial natriuretic factor) and α-MHC (α-myosin heavy chains).

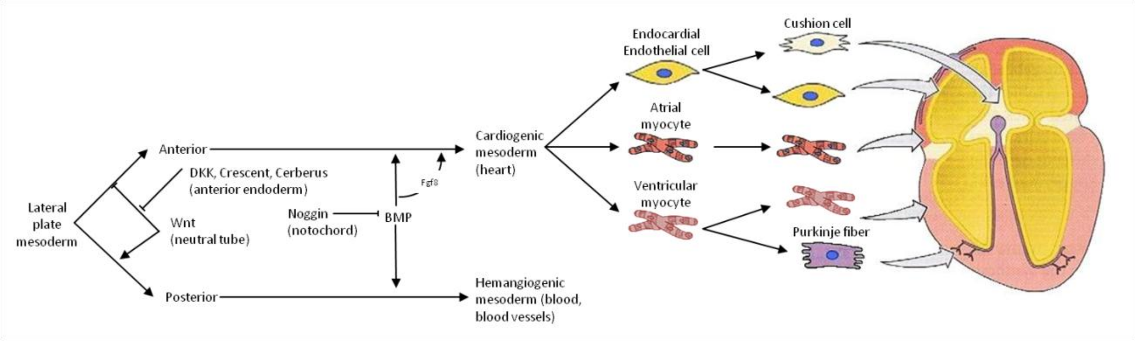


Fig. 1.1.5 – **Signaling pathways implicated into cardiogenic induction.** Inductive interaction involving Wnt and BMP signalling enables the generation of heart and blood lineages from the lateral plate mesoderm (adapted from Gilbert, 2003).

Once the cardiac crescent zone has been established, two cardiac progenitors regions can be distinguish, heart mesoderm ventrally and pericardial mesoderm dorsally, which are separated by the coelomic cavity. Heart mesoderm goes through another splitting that originates the muscular layer of the heart (myocardium) and the endocardium, whose cells originate the lining of blood vessels, heart valves, secrete the proteins that regulate myocardial growth and regulate the placement of nervous tissue in the heart (Gilbert, 2003). The pericardial mesoderm originates the pro-epicardial organ, from which the outer layer of the heart (epicardium), the coronary circulation and the interstitial fibroblasts are derived (Harvey, 2002).

The next step in heart development is the migration of the cardiac progenitor cells ventrally for the formation of the linear heart tube, in the midline of the embryo. While the fusion elapses, pulsations of the heart start being performed by the myocardial cells that develop their own ability to contract. Later, the linear heart tube acquires a spiral shape in consequence of a looping movement, in which its outer surface sweeps rightwards (Harvey, 2002). Looping of the heart converts the original anterior-posterior polarity of the heart tube into the right-left polarity seen in the adult organism.

While the heart is going through the looping, another morphological modification is taking place in its interior, the septation. Firstly, the heart is divided into one ventricle and one atrium by the endocardial cushions, which are structures composed by cells derived from the endocardium. The endocardial cushions are the precursors of the tricuspid and mitral valves, and aortic and pulmonary valves. Together with these endocardial cells, neural crest cells (migratory cell population that arises at the embryonic neural plate) also contribute for this septation (Gilbert, 2003). Another important function of the endocardial cushion is its participation in the formation of the right and left atrioventricular channels. From the bottom of the ventricle and the top of the atrium, inter-ventricular and inter-atrial septae are formed, respectively. Muscular cells from these septae start to proliferate and eventually fuse with the, already cited, endocardial cushions, therefore giving rise to a four-chambered heart (Gilbert, 2003; Harvey, 2002).

Going back in the developmental time line, it is import to point that there are two sources of myocardial cells for the building of the mammalian heart (Buckingham *et al.*, 2005; Musunuru *et al.* 2010), which will contribute to different parts of the heart. These distinct cells sources are organized into two fields, which are morphoregulatory and dynamic regions of developmental potency for the formation of the heart (Harvey, 2002) (Fig. 1.1.6). They are identified at the crescent cardiac stage as the first heart field (FHF) (more ventrally) and secondary heart field (SHF) (more dorsally). Gene markers such as Nkx2.5, Tbx5 and Hand1 are expressed in the first heart field that contributes to the formation of both the atria and the ventricles. On the other hand, the secondary heart field is characterized by the expression of Nkx2.5, Fgf10 (fibroblast growth factor 10), Isl1 (insulin gene enhancer protein 1), Hand1 and Hand2, and its descendent cells will populate the right ventricle and both atria (Buckingham *et al.*, 2005).

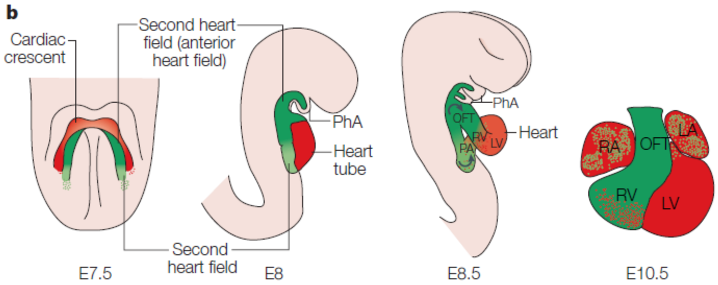


Fig. 1.1.6 – **Cardiac crescent organization in the mouse embryo**. The location and contribution of the second heart field (green), and is compared with the myocardial cells that are derived from the first heart field (red). Frontal views are shown for embryonic day 7.5 (E7.5) and E10.5, and lateral views for stages E8 and E8.5. LA (left atria), LV (left ventricle), OFT (outflow tract), PhA (pharyngeal arches), RA (right atrium), RV (right ventricle) (from Buckingham *et al.*, 2005).

Throughout the developmental process, the heart must maintain its rhythmical contractions, since their appearance during heart tube formation, through coordinated activation of the myocardium. These rhythmical contractions are carried out by the cardiac conduction system (CCS), which is composed by sinoatrial (SA) node, the atrioventricular (AV) node and a group of conducting fibers. (Boullin and Morgan *et al.*, 2005).

SA node is the first element to function in the CCS, it is found next to the wall of the right atrium, working as the dominant pacemaker, once its cells have the most rapid inherent rhythm. After, it passes the impulse to the AV node via the atrial myocardium.

When formed, the AV node delays the impulses passage from atria to ventricular myocardium.

The conduction pathway, which ends with the Purkinje fibers, spreads the electric signal by the all heart.

The nodal cells of the AV and SA nodes are similar to embryonic cardiomyocytes, nevertheless they are small with a poorly developed sarcoplasmic reticulum and lack a functional contractile unit due to poorly organized actin and myosin filaments.

The formation of the heart is a complex morphogenetic process that depends on the spatiotemporally regulated contribution of cardiac progenitor cells. These mainly derive from the splanchnic mesoderm of the first and second heart field (SHF), with an additional contribution of neurectodermally derived neural crest cells that are critical for the maturation of the arterial pole of the heart (Vincent and Buckingham, 2010). All together, this developmental process culminates in the formation of the “sun” of the body, as William Harvey described the heart in the seventieth century.

1.2. Cell-Cell Communication in Development

A body is more than a collection of randomly distributed cell types. Development involves not only the differentiation of cells, but also their organization into multicellular arrangements such as tissues and organs. Many of the answers to questions about morphogenesis involve the properties of the cell surface. Each cell type has different sets of proteins at its surface, and some of these proteins are responsible for forming the structure of the tissues and organs during development.

The extracellular matrix (ECM) appears as a critical region for much of animal development. It is composed by macromolecules secreted by cells into their immediate environment and it is a source of developmental signals. Cell adhesion and cell migration, essential behaviors on the developmental process, depend on the ability of cells to form attachments to extracellular matrices. ECM molecules can interact with their receptors on the plasma membrane triggering a cascade of events such as changes in gene expression, which in turn induces changes in cell’ behavior.

Extracellular matrices are made up of collagen, proteoglycans and glycoproteins, such as fibronectin and laminin. Fibronectin, for example, plays an important role in heart cell precursors’ migration to the midline and, when its signaling is interrupted, heart-forming cells fail to fuse and two separate hearts develop (Gilbert, 2003). Flectin is another protein of the extracellular matrix involved in heart development, particularly in the looping. When chick embryos are treated with monoclonal antibodies against flectin, heart looping is randomized (Brand, 2003).

One of the most important receptors that allows extracellular matrix to transduce its signal to the cell is the integrin. It binds to fibronectin on the outside of the cell and on the inside of the cell it serves as an anchorage site for actin microfilaments, which provide cell movement.

Another important mechanism of cellular communication uses paracrine factors. When proteins synthesized by one cell can diffuse over the ECM to induce changes in neighboring cells, the event is called a paracrine interaction, and the diffusible proteins are called paracrine factors. Many of these factors can be grouped into four major families on the basis of their structure, and all of them are somehow involved in heart development. These families are fibroblast growth factor (FGF), transforming growth factor β (TGF-β), Hedgehog and Wingless (Wnt) (Gilbert, 2003).

The FGF family has about two dozen structurally related members and is associated with several developmental functions, including angiogenesis and mesoderm formation (FGF2), and axon formation. FGFs can activate, by fosforilation, a set of receptor tyrosine kinases called the fibroblast growth factor receptors, which in turn can activate, namely, Ras/MEK/ERK signal transduction pathway and modulate the expression of, for example, serum response factor transcription factor, which enhances transcriptional activity of cardiac differentiation genes. (Alsan and Schultheiss, 2002; Brand, 2003; Gilbert, 2003).

The TGF-β superfamily comprises the TGF-β family, the activin family, the bone morphogenetic proteins (BMPs), the Vg1 family, glial-derived neurotrophic factor and the Mullerian inhibitory factor. From all these molecules, the one that plays an important role in heart development is the BMP, as described above, particularly BMP4, which has been found to regulate cell division, apoptosis, cell migration and differentiation. Members of this family bind to a TGF-β receptor dimer (type II and type I), which, when activated, can phosphorilate the Smad proteins 1 and 5 that, together with the Smad 4, are translocated to the nucleus to regulate gene expression (Gilbert, 2003).

BMP positively regulates FGF expression and, together their signals cooperate during cardiac induction. The FGF8 knockout mouse develops aberrantly and exhibits left–right asymmetry and various other cardiac malformations. However, early heart formation is not affected in this mutant mouse. In posterior mesoderm explants, a combination of BMP2 and FGF4 can stimulate heart formation, while neither factor alone can do so. Thus, FGF and BMP signaling seem to synergize to drive mesodermal cells into myocardial differentiation (Brand *et al.*, 2003).

The Hedgehog proteins are involved in many pathways, but the one relevant for cardiogenic fate is Indian hedgehog, which is expresses during gastrulation in the endodermal layer that will make contact with the cardiac mesoderm, inducing expression of the Bmp4 gene.

The Wnt family constituents act through the Frizzled transmembrane receptors. To provide a normal heart development, these molecules need to be blocked by specific inhibitors at the anterior region of the embryo to induce heart development.

Although FGF and BMP are the most important paracrine factors in heart development, there is another family of molecules that plays an important role in this process, named epidermal growth factor domain proteins (EGF-domain proteins) (Lee *et al.*, 1995; Brand, 2003; Nanba *et al.*, 2006). This protein family is characterized by extracellular domains that are thought to be involved in protein-protein interactions and that show homology to epidermal growth factor (EGF) (Beckingham *et al.*, 1998).

Intercellular signalling mediated by EGF-containing ligands and their cognate receptors are important regulators of growth and development. EGF domains are modular protein subunits found singly or in tandem, mostly in the extracellular milieu, where they are involved in a diverse array of functions (Wouters *et al.*, 2004). Structurally, the EGF domain is typically described as a small domain of 30–40 amino acids primarily stabilized by three disulfides bridges. Like FGF molecules, EGF uses the receptor tyrosine kinase (RTK) to transmit its signal to the cell interior.

These receptors all present kinase activity directed against tyrosine residues located both within the receptor itself (autophosphorylation) and on target downstream molecules. Ligand binding activates the kinase which is required for cellular responses (Fig. 1.2.1) (Wells, 1999). EGFR is a 170 kDa transmembrane glycoprotein and it is present on all epithelial and stromal cells as well as glial and smooth muscle cells (Wells, 1999).

The integrated biological response to EGFR activation varies from mitogenesis to apoptosis, migration to differentiation or dedifferentiation even in the same cell depending on the context, which includes cell density, type of matrix, other cytokines, and even the position within the tissue (Wells, 1999).

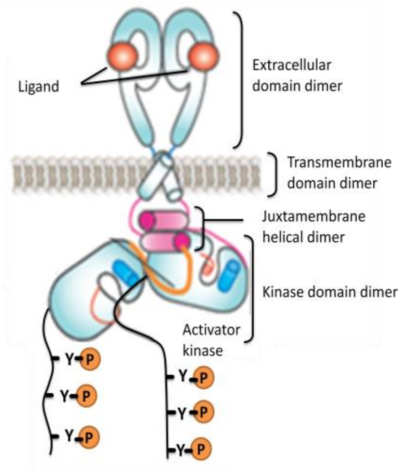


Fig. 1.2.1 – **The model for ligand-dependent activation of the EGF receptor** at the plasma membrane, which depicts contribution of the extracellular, transmembrane and the juxtamembrane domains to receptor dimerization (adapted from Jura *et al*, 2011).

Moreover, outside the cell, several classes of cell surface developmental signalling proteins with roles in cell-cell, or cell-extracellular matrix adhesion have proved to be Ca^{2+}-binding proteins. One of these molecules is the EGF-domain protein, together with integrins and cadherins (Beckingham *et al*., 1998).

It is believed that binding of Ca^{2+} by EGF-like domains functions to establish and stabilize the relative orientation of the EGF-like domains and thereby determine the overall shape of the molecule (Beckingham *et al*., 1998). The role of Ca^{2+} as a key and pivotal second messenger in cells depends largely on a wide number of heterogeneous so-called calcium binding proteins (CBP), which have the ability to bind this ion in specific domains and act as Ca^{2+}-modulated sensors, decoding Ca^{2+} signals (Yáñez *et al*., 2012).

Extracellular matrix plays an important role along different stages of heart development, when cellular movement and reshuffle take place. One of this crucial moments is gastrulation, in which cardiac mesoderm is specified. An integral component of gastrulation in all organisms is epithelial to mesenchymal transition, a fundamental morphogenetic event through which epithelial cells transform into mesenchymal cells (Ohta *et al.*, 2010).

An epithelium is composed of a sheet of epithelial cells that are closely associated with one another, like hipoblast and epiblast. By the other hand, mesenchymal cells generally adhere to their neighboring cells less tightly, being loosely associated and completely surrounded by extracellular matrix. Thus, EMT is a coordinated series of events involving loss of cell-cell and cell-extracellular matrix interactions and increased cell motility of epithelial cells.

All these diffusible molecules, particularly, BMP, FGF, EGF and calcium; membrane receptors, namely TGF-β, FGFR and EGFR; and signalling cascades, principally the Ras/MEK/ERK pathway, are involved in central steps of heart development, particularly in the earliest, as it will be further discussed.

1.3. Human Embryonic Stem Cells

The foundation of mammalian development and embryonic stem cell biology is pluripotency. Pluripotency is defined as the capacity of individual cells to self-renewal or divide to originate all lineages of the mature organism in response to signals from the embryo or the cell culture environment (Wray *et al.*, 2010). Embrionic stem cell lines are derived from pre-implantation embryos and retain these characteristics.

Many studies performed in mouse embryonic stem cells describe them 1. as tumorigenic, because they produce teracarcinomas when injected into adult mice; 2. as capable of complete integration into a developing embryo after being reintroduced into the blastocyst; 3. as being able to colonize the germ line and as such of generating chimaeric animals; 4. as competent to undergo multilineage differentiation and produce endoderm, mesoderm, and ectoderm; 5. as able to proliferate indefinitely; and finally, 6. As being able to produce two identical stem cell daughters when they divide (symmetrical self-renewal) or one stem cell daughter and one differentiated cell (asymmetrical self-renewal) (Burdon *et al.*, 2002).

These characteristics, which are largely shared by Human Embryonic Stem Cells (hESCs), confer them the potential to generate valuable differentiated cell types for drug development, for cell replacement therapies, and for gene delivery therapies. Moreover, stem cells can be used for gene targeting in order to create mutant cell lines to study gene function or to create mutant mouse models for studying diseases.

As such, embryonic stem cells represent a unique alternative model to investigate the basic principles of *in vivo* mammalian heart development (Kehat *et al*., 2001). Great similarity in genetic and epigenetic programs during embryonic development and embryonic stem cells differentiation have been described including the activation of the expression of transcription factors, cell receptors, tissue specific proteins and ion channels.

The gene-expression signature of pluripotent ES cells is relatively well known. Genetic studies identified three transcription factors present in the inner cell mass, which expression is remained and sometimes enhanced in embryonic stem cells. These are: Oct4 (octamer-binding protein 4), Sox2 (SRY (sex-determining region Y) box 2) and Nanog (derivation from the mythical Celtic land of perpetual youth). (Gilbert, S., 2003; Boyer *et al.*, 2005; Chambers and Tomlinson, 2009). Oct4, Sox2, and Nanog all bind to their own promoters, as well as to each other's promoters (Boyer *et al.,* 2005) and together they regulate the transcriptional regulatory hierarchy that specifies embryonic stem cell identity. Furthermore, Oct4, Sox2, and Nanog are thought to maintain the undifferentiated state of ES cells also by contributing to the repression of lineage specification factors (Jaenisch and Young, 2008).

Differentiation of ES cells towards cardiogenesis is characterized by a temporally and spatially ordered cascade of gene expression. Along this cascade, *Brachyury* expression, which specifies mesoderm, is followed by *Gata4* and *Nkx2.5* expression, which lead to the expression of *Mhc*, a cardiomyocite maturation specific gene (Kouskoff *et al*., 2005).

During heart development *in vitro*, multipotent mesoderm cells give rise to cardiac progenitor cells. These cells can originate all the lineages of the heart, namely myocites (Gilbert, 2003; Jaenisch and Young, 2008) (Fig. 1.3.1).

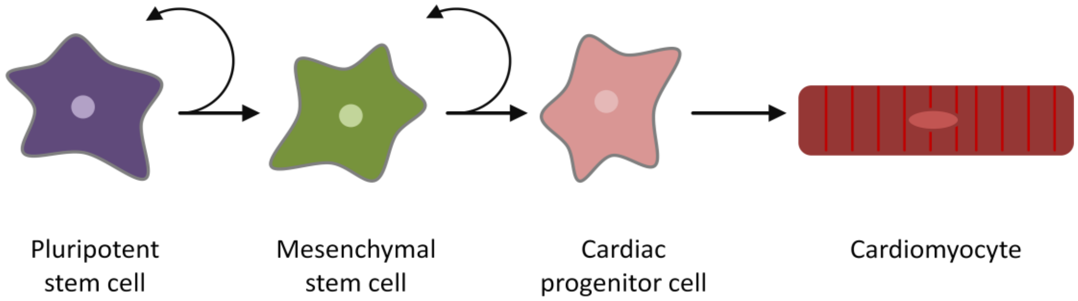


Fig. 1.3.1 – **The stem cell concept.** Cascade from pluripotent stem cell through mesenchymal stem cell and cardiac progenitor cell, to the well differentiated cell (cardiomyocyte) (adapted from Gilbert, 2003).

In order for hESCs to be of use for tissue replacement therapies it must be possible to direct their differentiation along specific pathways to produce a pure population of the desired. No contamination with undifferentiated ES cells can be tolerated because of the risk of subsequent tumour formation. (Nichols, 2001)

The human ES cells H9 line, used in this study, was derived by Thomson and colleagues in 1998 and characteristically retains a normal XX karyotype and expresses high levels of telomerase activity (Thomson *et al.,* 1998).

To obtain mesodermal cells from hESCs a two step protocol was performed. A combination of three different factors: FGF2, phosphoinositide 3-kinase (PI3K) inhibitor (LY294002) and BMP4, herein referred to as FLyB, was used to obtain early mesoderm. The first two have been reported as inductors of mesoderm differentiation (Bernardo *et al.*, 2011) and the PI3K inhibitor facilitates differentiation of hESCs (McLean *et al.*, 2007). As referred before, during embryogenesis, specification of all kinds of mesoderm, including lateral plate mesoderm, follows a posterior-anterior BMP gradient along the primitive streak, so once cells are in that first differentiated state, they can be induced to differentiate towards lateral plate mesoderm (second step of this protocol) with the addition of FGF2 and high BMP4 (FB50) (Cheung *et al.,* 2012).

1.4. Developmental Biology as a Biomedical Science

Developmental biology of the heart as a medical science comprises other emerging disciplines such as regenerative medicine, cell therapy and stem cell biology. Scientific research focused in studying heart development in all these areas has as ultimate objective the restoration of the myocardial function of a sick heart.

In man, about 1 per 100 live-born infants has some form of congenital heart malformation, while the incidence of prenatal fatal congenital malformations is estimated to amount to 5–10% of abortions (Brand, 2003). Moreover, according to the world's health organization (WHO), the leading cause of adult deaths is ischaemic heart disease, amounting to 12.8 % over all deaths. Occlusion of a coronary vessel and the resultant myocardial ischemia rapidly results in myocardial necrosis followed by scar formation. When the ischemic myocardium is reperfused, there is a rapid onset of contraction band necrosis and an intense inflammatory cascade. It has been well established that adult cardiac myocytes do not replicate, thus these pump units are not actually replaced (Jackson *et al.*, 2001).

Thus, there is hope that understanding how the cardiac muscle is induced during embryonic development will have an impact on therapeutic approaches for cardiac regeneration and cell substitution (Brand, 2003).

With this in mind, our laboratory has strive to identify novel genes required for the correct development and differentiation of the vertebrate heart cell lineages by using Affimetrix GeneChip® Chicken Genome arrays (Bento *et al.*, 2011). In this experiment a candidate heart specific gene, that for its own and its product characteristics, was chosen to be further studied. This gene is Ccbe1 (calcium-binding EGF-like domain 1) and it is the central gene studied in this Master Thesis. It had already been reported that mutations in Ccbe1 result in a severe lymphatic system phenotype both in mouse (Bos *et al.*, 2011) and in zebra fish (Hogan *et al.*, 2009) but its role in the heart remains elusive. Moreover, human CCBE1 gene also being associated with Hennekam syndrome, a disorder characterized by abnormal lymphatic system development causing generalized lymphedema, intestinal lymphangiectasias, some of the patients carrying a mutated hCCBE1 gene were also shown to possess congenital heart defects including hypertrophic cardiomyopathy and ventricular septal defects (Connell *et al.*, 2009; Alders *et al.*, 2009).

Besides that CCBE1 has been described as being down regulated in breast and ovarian cancers, by hypermethylation of its promoter (Yamamoto and Yamamoto, 2007; Barton *et al.*, 2010). In addition, it has been proposed that loss of CCBE1 expression may promote ovarian carcinogenesis by enhancing migration and cell survival (Barton *et al.*, 2010), and so being a tumour suppressor gene candidate.

While the importance of m*Ccbe1* for the development of the lymphatic system appears to be indisputable, its role in cardiac development has not been investigated in detail despite the increasing evidence of its potential role in cardiogenesis.

In sum, it is known that Ccbe1 is expressed in cardiac precursors in the mouse, Ccbe1 homozygous mutation cause lethality in both mouse and zebra fish, and the same kind of mutation in humans is responsible for a high incapacitating disease. Thus, the study of Ccbe1 function in heart development could unveil important findings and ESCs are an invaluable toll to ascertain its function in heart development.

1.5. CCBE1 in Heart Development

The screening performed in our laboratory aimed to characterize chick cardiac precursors. (Fig. 1.5.1).

Those cells were isolated from electroporated chick embryos with a construct containing EGFP expression under the control of a 2.5 kb fragment upstream the ATG of chick *Cerberus,* a gene expressed into cell population such as the anterior mesendoderm, lateral plate mesoderm, heart and anterior blood islands or hemangioblasts (*cCer*; *Tavares et al.*, 2007). The genetic profile provided relevant data of the chick heart/hemangioblast precursor lineage identity and led to the detection of 301 uncharacterized genes (119 unknown genes and 182 annotated genes) that were upregulated in the heart/hemangioblast precursors (Bento *et al.*, 2011).

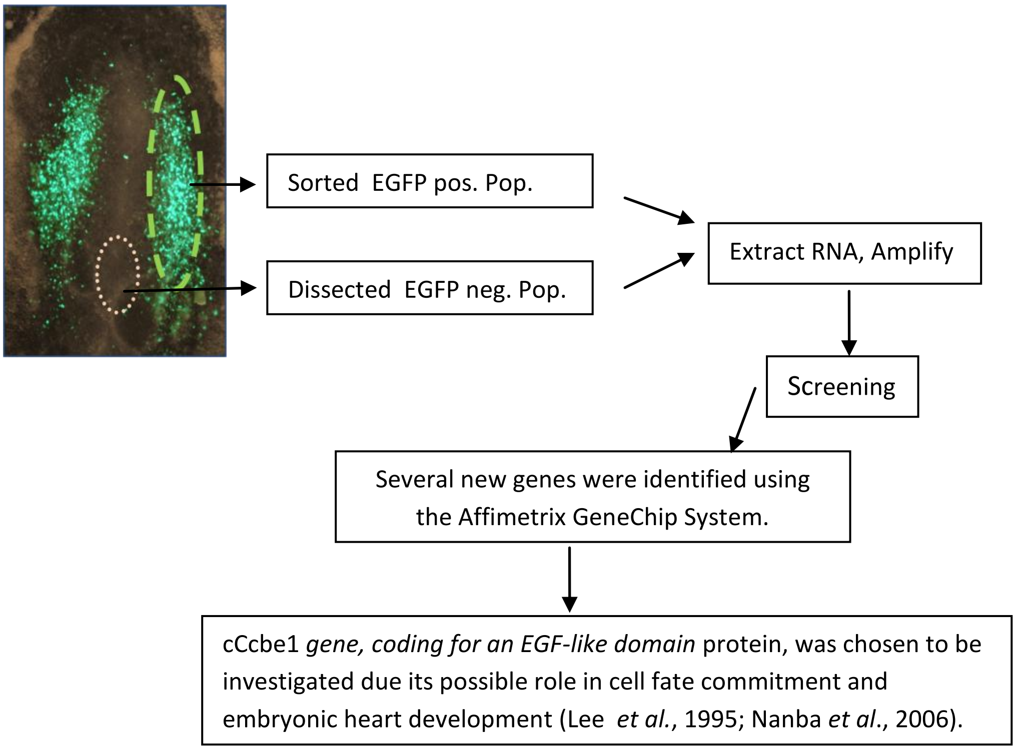


Fig. 1.5.1 - **Differential screening of the E7.5 chick H/HPC population.** The characterized 2.5cCar promoter drives expression of EGFP reporter construct in the H/HPC population (green) (adapted from Bento *et al.*, 2011).

Among the annotated genes potentially involved in heart development, chick *Ccbe1* was found to be upregulated by 7.8 fold in the cardiac progenitors in comparison to the embryonic control cells. c*Ccbe1* gene is conserved across vertebrates with chick *Ccbe1* aminoacid (a.a.)

sequence being 78%, 67% and 79% identical to the mouse, zebrafish and human Ccbe1 protein, respectively (BLAST-NCBI). *Ccbe1* encodes a secreted protein that contains a signal peptide, a collagen domain, a calcium binding EGF-like domain and a RGD sequence. Preliminary expression analysis using whole mount *in situ* hybridization (WISH) further indicated that c*Ccbe1* is preferentially expressed in cardiac mesoderm precursors (Bento *et al*., 2011). Moreover, knockdown of cCcbe1 during early chick development was shown to result in severe heart tube malformations including *cardia bifida* (formation of two tube-like structures) and embryonic heart hypertrophy (M. Bento, unpublished data). *Cardia bifida* is also a phenotype caused by Gata4 or Mesp1 homozygous mutation (Brand, 2003).

In the mouse, m*Ccbe1* was shown to be expressed in the major cardiac lineages, namely FHF, SHF and proepicardium (Facucho-Oliveira *et al*., 2011).

Given that Ccbe1 mutant phenotype in chick, the confirmed presence of CCBE1 in the mouse heart and the fact that patients carrying a mutated h*CCBE1* gene exhibit congenital heart defects (Connell *et al.*, 2009; Alders *et al.*, 2009), CCBE1 appears to be as a novel cardiac gene whose study could unveil new insights of heart development and possibly congenital heart disease. As such, the study of CCBE1 in a human context is of upmost relevance and hESCs are the perfect starting point for such study.

In humans, CCBE1 is located at the chromosome 18 and it is 262.130 kb length (Fig. 1.5.2). hCCBE1 has 11 exons, and the second intron has 216.390 kb. According to *Ensembl* CCBE1 has two coding transcripts, one coding for a protein with 135 aa and the other coding for a larger protein with 406 aa. It is believed that the 406 aa protein is an active form.

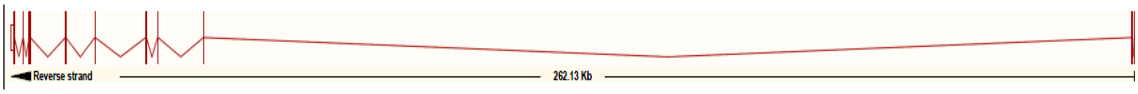


Fig. 1.5.2 – **CCBE1 transcript.** The vertical bars represent the 11 exons and the spaces between them represent the introns (from *Ensembl*, 2012).

As described above, CCBE1 contains an EGF-like domain acting probably as a direct agonist for the EGF receptor. It is therefore thought that, similarly to other EGFR agonists, CCBE1 is synthesized as a transmembrane precursor and must therefore be proteolytically cleaved by matrix metalloproteases (MMP) to release the mature growth factor (Fig. 1.5.3) (Gschwind *et al.*, 2001).

By acting through the EGFR, it is expected that CCBE1 signal is transduced via a number of signaling pathways, namely, the Ras/MEK/ERK [MEK (Mitogen activated protein (MAP) kinase kinase), ERK (Extracellular-signal-Regulated Kinase)] pathway, the PI3K (phosphatidylinositol-3-OH kinase)-dependent signals, the JAK/STAT (Janus kinase/ Signal transducer and activator of transcription) pathway and the PLCγ/DAG/PKC (Phospholipase c-γ/ Diacylglycerol/ Protein kinase-C) (Nyati *et al*., 2006).

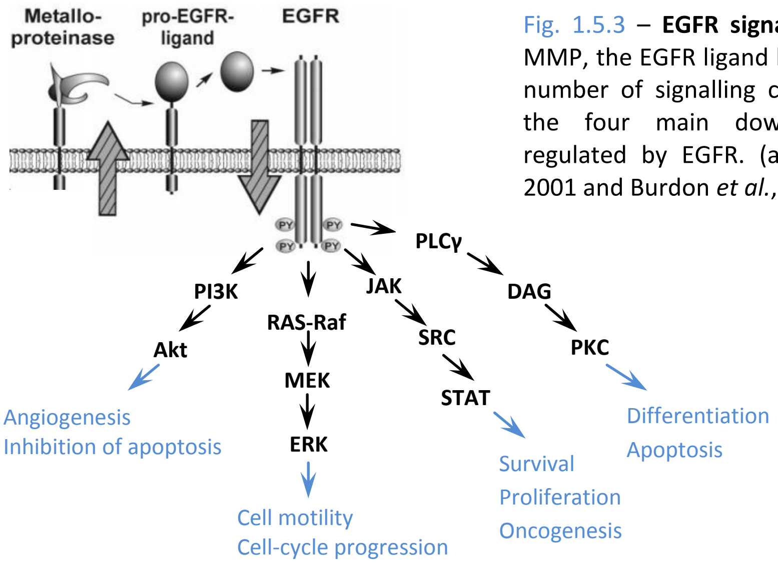


Fig. 1.5.3 – **EGFR signaling.** After being cleaved by a MMP, the EGFR ligand binds to its receptor, initiating a number of signalling cascades. Here are represented the four main downstream signalling pathways regulated by EGFR. (adapted from Gschwind *et al.*, 2001 and Burdon *et al.*, 2002; Nyati *et al*., 2006).

The purpose of this Master study was to extend the knowledge on Ccbe1 role in heart development to the human context using embryonic stem cells. To achieve this aim, hESCs were grown in chemically defined medium and either kept undifferentiated or induced to differentiate towards cardiac progenitor cells. The expression, function and regulation of CCBE1 were accessed in pluripotent and differentiated cells. To aid the characterization of CCBE1 role, CCBE1 knock down lines were generated using shRNAs. Additionally knock down lines were generated for ISL1 and BRA genes which according to this study are two candidate CCBE1 regulators.

2. MATERIALS AND METHODS

2.1. Human Embryonic Stem Cells Culture in Chemically Defined Conditions

Three main things are necessary to keep H9 human embryonic stem cells (hESC) in culture – the right coating, appropriate and timely splitting and maintenance with regular feeds and fresh growth factors.

Before plating the cells, the dishes have to be coated, first with porcine gelatine (Sigma) 0.1% in embryo tested water (Sigma) for 15-60 minutes at room temperature (RT), and secondly, with mouse embryonic fibroblasts medium (MEF medium (Table 2.1.1)) containing 10% fetal bovine serum (Hyclone) over-night (minimum 3h required) at 37^{o}C incubator and 5% CO_2. Plates last up to 7-8 days but can be recoated with MEF medium after that.

Splitting is the critical step. It has to be done carefully, otherwise, cell culture quality can be lost. It has several intermediate steps. hESCs colonies are washed in phosphate *buffered* saline (PBS) without Ca^{2+}/Mg^{2+} (Sigma) for 1-2 minutes at RT (1ml/ well of 6 wells plate). This causes the detaching of the cells from the growing surface and from each others.

Next, cells are scraped off the dish in Chemically Defined Medium (CDM) (Table 2.1.1) with a 5 ml pipette and dissociated into clumps by pipetting up and down 1-4 times. If the cells are going to be cultured in pluripotency medium, clumps should have a big size (20-30 cells/ clump), if the objective is to perform a differentiation experience clumps should be smaller (10-15 cells/ clump).

After scraping, the cellular suspension is pipetted into a 15ml falcon tube and left for 5-10min until clumps gravitate to the bottom of the tube. The top medium (“supernatant”), which contains differentiated cells, is discarded by aspiration.

Cells are resuspended in CDM containing 10 ng/ml Activin A (R&D Systems 338-AC), 12 ng/ml FGF2 (R&D Systems 233 FB) and 14µg/ml insulin (pluripotency/ hES medium) and clumps are plated at low density (100-300 clumps/well) on the FBS-coated 6 wells plate (previously washed with PBS) in CDM enriched with the growth factors described above. Usually cells are plated in a 1:6 to 1:9 dilution factors.

It is important not to move the plates during the first 12-24h, so that colonies can attach properly. Cells are left in culture during 5-7 days, before the next split, until the colonies reach a large size (4 to 6 times bigger than colonies grown on feeder) and the culture becomes confluent. During this time, cells are fed every 48h. If many dead cells are observed, the culture has to be washed with PBS, before adding fresh medium and more frequent feeds need to be done.

For differentiation into cardiac precursors, Bovine Serum Albumin (BSA) was substituted by PVA (Polyvinyl Alcohol) in the CDM. Polyvinyl Alcohol is a BSA substitute, used in differentiation medium, because there is no batch variation.

To obtain cardiac progenitors, a protocol based in two different defined media is used. 12-24h after the splitting, pluripotency medium is changed to FLyB medium, which consists of CDM-PVA enriched with FGF2 (20 ng/ml, zebrafish, recombinant, gift of Marko Hyvönen), LY294002 10 µM (Sigma), and BMP4 10ng/ml (R&D Systems) for 36h. FLyB induces early mesoderm differentiation. The second stage of differentiation protocol, which induces lateral plate mesoderm differentiation, consisted of treating cells with CDM-PVA + FGF2 20 ng/ml + BMP4 50 ng/ml for the next 3.5 days. All experiments were repeated twice on different passages of cells to ensure that the patterns of gene expression described were reproducible.

Human ESC culture conditions for CCBE1 characterization experience

Cells were grown in 12 wp and collected for RNA extraction or fixed for immunocytochemistry at nine different time points: day zero (d0) corresponding to undifferentiated cells, cultured in hES medium, and eight time points for differentiated cells (d1, 36h, d2, 60h, d3, d4, d5, d6), along the FLyB+FB50 differentiation protocol. For each time point three biological replicates were collected.

Culture conditions for CCBE1 knock down experience

Cells were grown in 12wp and collected for RNA extraction at four different time points: day zero (d0) corresponding to undifferentiated cells, and three time points of differentiated cells (36h, d3, d5), along the FLyB+FB50 differentiation protocol. For each time point three biological replicates were collected.

Medium	Components	Quantity	Concentration	Company
MEF	Advanced DMEM F12	450ml		Invitrogen, 12634028
	FBS	50ml	10%	Biosera, S04253S181S
	L-Glutamine	5ml	2mM	Invitrogen, 25030024
	β-Mercaptoethanol	350µl	10mM	Sigma, M6250-100ML
CDM BSA/PVA	IMDM	250ml		Invitrogen, 21980065
	F-12	250ml		Invitrogen, 31765068
	BSA or PVA	2.5g or 0.5g	5mg/ml or 1mg/ml	Europa bioproducts EQBAC62 lot BAC62-624
	Conc. Lipids	5ml	1%	Invitrogen, 11905031
	Insulin	700µl	14µg/ml	Roche 1376497
	Transferrin	250µl	15µg/ml	Roche 652202
	Monothioglycerol	20µl	450 µM	Sigma M6145
	Pen/Strep (optional)	5ml		Invitrogen, 15140122
Gelatine	Gelatine	0.5g	1mg/ml	Sigma, G1890-100G
	Water for Embryo Transfer	500ml		Sigma, W1503-500ML

Table 2.1.1 – **Media composition.**

2.2. RNA extraction, cDNA synthesis and Semi-quantitative PCR

RNA extraction was performed using a RNeasy Mini Kit (Qiagen). All samples were treated with RNase-Free DNase (Qiagen). RNA concentration was measured in a Nanodrop (Thermo Scientific).

To synthesize cDNA from RNA, a *Maxima® First Strand cDNA Synthesis Kit (Fermentas)* was used. This kit consists of three components – Maxima Enzyme Mix, which contains Maxima Reverse Transcriptase and Thermo Scientific RiboLock RNase Inhibitor; 5X Reaction Mix, which contains reaction buffer, dNTPs, oligo $(dT)_{18}$ and random hexamer primers; water nuclease-free. For each sample, a reaction mix with a 10 µl final volume was prepared. The components were added as indicated bellow:

5x Reaction Mix – 2 µl
Maxima Enzime Mix – 1 µl
Templete RNA – volume equivalent to 500 ng
Water – to 10 µl

Reverse transcription was performed following the program described bellow:

Primers Annealing - 10 minutes at 25°C
Polymerization - 15 minutes at 50°C
Enzyme inactivation - 5 minutes at 85°C

All samples were treated with RNaseH and diluted 30 times. The ssDNA was directly used in semi-quantitative real time Polymerase Chain Reaction (qPCR).

qPCR mixtures were prepared using a Fast SYBR® Green Master Mix (Applied Biosystems, 4385614). This product contains SYBR Green I Dye (fluorescent intercalating dye which binds to the double stranded DNA), AmpliTaq®Fast DNA Polymerase UP (Ultra Pure), Uracil-DNA Glycosylase (UDG, prevents the reamplification of carryover PCR products by removing any uracil incorporated into single- or double stranded amplicons), ROX™dye Passive Reference (allows for correction of well-to-well variation due to pipetting inaccuracies and fluorescence fluctuations), nucleotides (dNTPs) and optimized buffer components. qPCR reaction mixes were prepared to a 10 µl final volume, as indicated:

Syber Mix: 5 µl
Forward Primer: 0.2 µM
Reverse Primer: 0.2 µM
Template DNA: 3 µL
Water: 1.6 µl

Human primers for qPCR were design using a primer design software (Primer3) and their sequences are shown on Table 2.2.1.

Gene	Forward	Reverse
BRA	TGCTTCCCTGAGACCCAGTT	GATCACTTCTTTCCTTTGCATCAAG
CCBE1	GAGATGGTTCTAAGGGGAGA	ATGTCAGCCAGCATAAGTAGCA
GATA4	TCCCTCTTCCCTCCTCAAAT	TCAGCGTGTAAAGGCATCTG
ISL1	AGATTATATCAGGTTGTACGGGATCA	ACACAGCGGAAACACTCGAT
KDR	TTTTTGCCCTTGTTCTGTCC	TCATTGTTCCCAGCATTTCA
MESP1	AGCTTGGGTGCCTCCTTATT	TGCTTCCCTGAAAGACATCA
MESP2	GCAGTGTACCAGGGTCTCTCT	ACTGTGGCTCCAGCACCT
NANOG	CATGAGTGTGGATCCAGCTTG	CCTGAATAAGCAGATCCATGG
NKX2.5	CAAGTGTGCGTCTGCCTTT	CAGCTCTTTCTTTTCGGCTCTA
OCT4	AGTGAGAGGCAACCTGGAGA	ACACTCGGACCACATCCTTC
PBGD	ATTACCCCGGGAGACTGAAC	GGCTGTTGCTTGGACTTCTC
SOX2	TGGACAGTTACGCGCACAT	CGAGTAGGACATGCTGTAGGT

Table 3.2.1 – Forward and reverse sequences of primers for qPCR.

qPCR Reactions were performed in a 7500 Fast ABI Instrument, according to the following program:

AmpliTaq®Fast DNA Polymerase, UP Activation – 20 seconds at 95°C
Desnaturation – 3 seconds at 95°C
Annealing/ Extension – 30 seconds at 60°C
40 cycles
Each sample was run in duplicate and normalised to Porphobilinogen Deaminase (PBGD) in the same run. Error bars on all qPCR graphs represent standard deviation from three independent biological replicates.

The $2^{-\Delta CT}$method (Livak *et al.*, 2001) was used to analyze the relative changes in gene expression from real-time PCR experiments, in which

$$2^{-\Delta CT} = 1 \text{ (PBGD primers efficiency)} + 1 \text{ (CCBE1 primers efficiency)}^{-\text{(CT gene of interest- CT housekeeping gene)}}$$

2.3. Generation of Human ESC knockdown lines

Stable knockdowns (KD) of CCBE1, ISL1 and BRA were carried out with pLKO.1-shRNA vector (Thermo Scientific, Sigma) by Lipofectamine 2000 transfection. A Scrambled pLKO.1-shRNA vector (Thermo Scientific, Sigma) was used as a control. Stable clones were screened by PCR and the percentage of knockdown was determined by comparison to expression in the scrambled control transfected lines.

Vectors came in glycogen stock (*E.coli*). DNA was amplified and purified using a Plasmid Maxi Kit (Qiagen) and following manufacturer's instructions.

To expand the culture, cells were picked from the glycogen stock and grew in 2ml LB medium supplemented with the appropriate selective antibiotic (ampicilin, 50ng/ml). After growing over-night at 37°C, the starter culture was diluted 1/125 into selective LB medium and grown again at 37°C for 12 hours. Several steps were followed, namely: bacterial cells resuspention; cell lysis; genomic DNA, proteins and cell debris neutralization by a precipitated material formation. Then, the plasmid DNA was washed and eluted using a QUIAGEN-tip which contains a column with chemical affinity to DNA. After that, a precipitation step was performed by adding isopropanol to the eluate, followed by 70% ethanol washing. Finally, each vector was redissolved in TE (Tris-HCl, EDTA) buffer.

2.4. Transfection using Lipofectamine 2000

Prior to transfection cells were plated in small colonies and at low density. Transfection was performed two days after the plating.

It was used 1 well of a 6-well plate per vector. In this experiment five different CCBE1 (TRCN0000055473; TRCN0000055474; TRCN0000055475; TRCN0000055476; TRCN0000055477 (Thermo Scientific)) (Table 2.4.1) and five ISL1 (TRCN0000014893; TRCN0000014894; TRCN0000014895; TRCN0000014896; TRCN0000014897 (Thermo Scientific)) vectors were used to transfect cells. BRA KD line was performed with only one vector (TRCN0000005481 (Sigma)), previous tested by Tiago Faial.

Vector	Target sequence in the corresponding DNA (5'-3')	Gene location
TRCN0000055473	CCGAGTGCTGTGTACTTGTTA	4th exon
TRCN0000055474	CCATGAGAAGTCTGAGAACAT	6th exon
TRCN0000055475	GCTACTTATGCTGGCTGACAT	11th exon
TRCN0000055476	GAAGCCATACTGTCTGGATAT	4th and 5th exons
TRCN0000055477	GTTCCCTTTACCTCAGGAATT	11th exon
Scrambled		-

Table 2.4.1 – **CCBE1 shRNA vectors sequences and gene pair location.**

Transfection protocol per each well:

Transfection preparation

Two different solutions were prepared; one with 10 µl lipofectamine in 250 µl OptiMEM, and the other had 4 µg plasmid DNA diluted in 250 µl OptiMEM. After 5 min at room temperature (RT), both solutions were mixed and incubated for 20 min, to allow for lipid-DNA complexes to form.

Cell transfection

First of all, cells were prepared for transfection, i.e. wells (12 of a 6wp, in total) were washed with PBS and covered with 1ml OptiMEM.

After that, DNA-lipofectamine complexes were added drop by drop to each well. To homogenise the medium, plates were mixed gently by rocking the plate back and forth.

Cells were incubated at 37°C in a CO_2 incubator. 10 hours later, the transfection mixture was removed and hES medium was added.

Selection antibiotic - puromycin (1µg/ml) - was used to supplement hES medium, 48h later and thereafter.

Picking

From the day puromycin is added cell death starts to take place and about 95% of the cells die within the next couple of days. Surviving cells were left to grow during approximately 5 days.

6 colonies from each well were picked to 6 different wells of a 12wp. At the end, there were 30 CCBE1 KD clones, 30 ISL1 KD clones, 6 BRA KD clones and 6 Scrambled.

After 3 days, each clone was split in a 1:2 proportion. There were 2 wells per each clone, one for RNA extraction and qPCR analysis and the other one to keep in culture.

2.5. Immunocytochesmistry

Cells were grown in 12wp. When ready, medium was removed from plates and cells were fixed for 10 min in ice cold 4% ParaFormolAldeyde (PFA)/PBS. Fixative was removed and cells were washed three times in PBS.

To allow antibody entrance and avoid unspecific binding, cells were permeabilized in PBST (PBS +0.01 % Triton X100) for 10 min and blocked for 1 h in blocking buffer (PBST + 5 % donkey or goat serum)at room temperature.

Primary antibody incubations were performed in blocking buffer (200 µl/well), over-night at 4°C, in a humidified chamber as follows: rabbit anti-CCBE1 (1:75, Sigma), goat anti-NANOG

(1:200, R&D Systems), goat anti-BRACHYURY (1:150, R&D Systems), and goat anti-Isl1 (1:100, ABCAM). On the next day, cells were washed three times for 5 minutes in PBST.

Fluorescently labelled secondary antibodies (IgG, 1:400) were added for 1 h at room temperature as follows: Alexa 488 donkey anti-goat and Alexa 594 goat anti-rabbit (Invitrogen). From this step on, plates were kept in the dark.

Three washes in PBST for 5 minutes were performed. DAPI was added to the last wash (1:10000, Sigma). Cells were washed again twice in PBST.

Staining was observed in a fluorescence microscope.

2.6. Chromatin Immunoprecipitation

This protocol has four main steps, crosslinking and sonication, immunoprecipitation, bead washing and samples purification. Before start the immunoprecipitation, it is necessary to have enough number of cells, so hESCs were grown in three 10 cm plates until the culture become clearly confluent (2×10^7 to 5×10^7 cells). Then, they were collected as described below:

1) Cells were crosslinked with 10mM Dimethyl 3,3′-dithiopropionimidate dihydrochloride (DTBP, Sigma, D2388) and 2.5mM 3,3′-Dithiodipropionic acid di(N-hydroxysuccinimide ester) (DSP, Sigma, D3669) in PBS at RT for 15mins.

2) Further crosslinking was done on plates: 125 µL of 40% formaldehyde (final conc. 1%) added to 5 mL PBS, incubate rocking for 15min at RT.

3) 312.5 µL of 2 M Glycine (final conc. 0.125M) were added and cells incubate for a further 5-10min to neutralise formaldehyde.

4) Plates were washed twice with 5mL ice cold PBS and detached cells by scraping in 3 mL cold PBS plus protease inhibitors mix (1x) and PMSF (0.4 mM) and pooled in a 50mL falcon tube.

5) Cells were next span for 6 min at 1,200 rpm and frozen at -80°C until ready to continue the protocol.

When ready, the samples were thawed and prepared for sonication, as follows:

1) Resuspend pellet in 2mL Cell Lysis Buffer (Add protease inhibitors, 10µL/mL PMSF) and incubate on ice for 10 min.

2) Spin at 1,800 rpm at 4C for 5min.

3) Resuspend pellet (nuclei) in 1.25mL Nuclei Lysis Buffer (add PI, PMSF) and incubate on ice for 10min.

4) Add 0.75mL of IP dilution Buffer (add PI, PMSF) and keep samples on ice.

Sonication (Misonix 4000) was performed using a microtip probe, 60% power, 15secs on/45secs off. This procedure originated DNA fragments about 1000-2000bp length.

Proteins, lipids and cellular debris were despised by spinning the samples at 14,000 rpm for 5min at 4°C. Supernatant, which contains crosslinked DNA, was kept and transferred to a new tube and 3.5 mL IP dilution buffer, supplemented with PI and PMSF. Solutions were mixed gently. 300 µL of sample was removed and frozen (Input).

To continue to immunoprecipitation, samples were dispensed into an appropriate number of tubes, accordingly to the different IP performed. Each sample was incubated shaking/rotating at 4°C overnight, with 5 µg of antibody (the same used in ICC).

Magnetic beads were used to catch Ab, which binds to transcription factors, which in turn are bound to DNA fragments.

First, beads (100 µl per IP) were washed three times with 1 ml Block Solution. Beads were resuspended in 100 µl Block Solution, per IP.

100µl magnetic beads were added to each IP, which were incubated shaking/rotating for 1h.30min at 4°C.

Next, beads were washed with three different buffers, and afterwards, protein complexes need to be eluted from the beads. This wash process was performed with samples always kept on ice and it was performed as described:

1) Collect the beads using a magnetic rack; the complex protein-DNA of interest should now be bound to the beads.

2) Add 500µL of IP Wash 1, collect the beads as above. Repeat washing with IPW1.

3) Add 500µL of IP Wash 2, collect the beads as above.

4) Wash beads twice with 500 µL TE.

5) Add 150µL Elution Buffer, put in a themomixer 30min at 65°C with vortexing (1,200 rpm).

6) Collect the beads as above and transfer supernatant to a new tube and repeat the elution. Final volume should be 300µL per IP.

The next and last step consisted in samples purification, removing of crosslinks, degrading RNA and proteins, and DNA precipitation. Samples were treated as follows:

1) Treat all samples (including Input) with 100mM DTT for 30mins at 37°C, add 15µL 5M NaCl and leave ON at 65°C to reverse chemical crosslinking.
2) Add 8µL RNase A (1 mg/mL stock), incubate at 45°C 2 h.
3) Add 4µL Proteinase K (20 mg/mL stock) and incubate at 55°C 2 h.
4) Add 300µL phenol/chloroform, vortex vigorously and spin at 14,000 rpm for 5min (RT).
5) Add 30µL 3M NaAc pH5.2 and 30µg glycoblue, vortex, add 750µL 100% EtOH and incubate at –80C for at least 30min to precipitate the DNA.
6) Spin at 14,000 rpm for 30 min at 4°C.
7) Wash pellet with cold 70% EtOH and spin at 14,000 rpm for 5min at 4°C.
8) Remove EtOH and dry pellet.
9) Add 30µL sterile dH_2O to IP samples and 100 µL to Input samples.

Buffers used in this protocol are presented in Table 2.6.1.

Finally, samples were analysed by qPCR and the amplification results were compared to those of the input. Input samples were diluted twice prior to being analysed. ChIP samples were run in duplicate and normalised to Input in the same run. Glyceraldehyde-3-phosphate dehydrogenase (GAPDH) was used as housekeeping gene.

Human CCBE1 primers for ChIP-qPCR were designed using the primer design software Primer3.

Genomic Region	Forward	Reverse
5'region	TAGTTGGAGGTGGTTTTGTGC	GTTGACAGCCCATGCTAGCGC
Mid region	AAGGGCTAAGGTCATTTTCTCC	AAAAGAGGCAAAATGGAACAGA
3'region	CATCTGGCGTTCTCTCTATGG	TCCCTACCTCTTCTTCCCAAA
Mid region	GTGGCAGGAGTGTAAGATCACC	CAAGGAAGAGCTGAGTGGAAAG
3'region	CCTCCAATTAGCACCTCATCA	TTCGCAAAAAGAATGCTCAGT

Table 2.6.1 – **Primers sequences for ChIP-qPCR.** Primers used in BRA/ISL1 ChIP after 60h of differentiation are shown is red and the oligonucleotides used in BRA/NANOG ChIP for pluripotent cells are shown with a blue mark.

qPCR reaction mixes were prepared to 10µl final volume, as indicated:

Syber Mix: 5 µl
Forward Primer: 0.2 µM
Reverse Primer: 0.2 µM
Template DNA: 3 µL
Water: 1.6 µl

Reactions were performed in a 7500 Fast ABI Instrument, according to the following program:
AmpliTaq®Fast DNA Polymerase, UP Activation – 20 seconds at 95°C
Desnaturation – 3 seconds at 95°C
Annealing/ Extension – 30 seconds at 60°C
45 cycles

Buffers composition is shown in Table 2.6.2.

Buffers	**Composition**
Cell lysis buffer	10mM Tris pH 8.0, 10mM NaCl, 0.2% NP-40
Nuclei Lysis Buffer	50mM Tris pH 8.1, 10mM EDTA, 1% SDS
IP Dilution Buffer	20mM Tris pH8.1, 2mM EDTA, 150mM NaCl, 0.01% SDS, 1% Triton X-100
IP Wash Buffer I	20mM Tris pH8.1, 2mM EDTA, 50mM NaCl, 0.1% SDS, 1% Triton X-100
IP Wash Buffer II	10mM Tris pH 8.1, 1mM EDTA, 0.25M LiCl, 1% NP-40, 1% Deoxycholic acid, sodium salt
Elution Buffer	100mM NaHCO3, 1% SDS
Blocking Solution	0.5% BSA (w/v) in PBS

Table 2.6.2 – **Buffers used in ChIP protocol.**

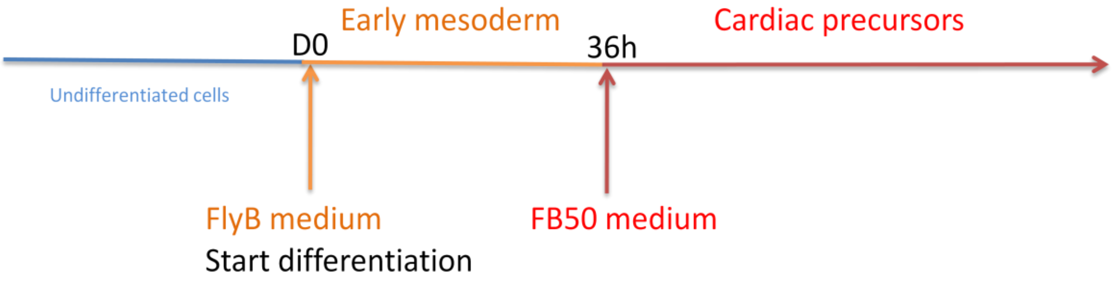
3. RESULTS

3.1. CCBE1 Characterization

In order to establish the expression pattern of CCBE1 in H9 hESCs, cells were cultured in a chemically defined medium (CDM) and collected for qPCR analysis and immunostaining at nine specific time points during cardiac differentiation. For this, pluripotent cells were exposed to growth factors that induce cardiac specification, following a previously described differentiation method (Fig. 3.1.1) (Bernardo *et al*., 2011; Cheung *et al*., 2012). The expression of mesoderm and cardiac specific genes was also analysed to confirm the cells had differentiated as expected and ascertain which genes follow the expression pattern of CCBE1. RNA was extracted from each sample (27 in total), and later reverse transcribed into cDNA to determine the expression levels of various genes by qPCR analysis (Fig. 3.1.2 and Fig. 3.1.3); primers sequences can be found in Table 2.2.1 (*Materials and Methods*).

Subsequently, real time PCR was performed and the expression levels of each gene were normalised to that of Porphobilinogen Deaminase (PBGD) for each sample. Error bars on all qPCR graphs represent standard deviation from three independent biological replicates.

This experiment was repeated twice on different passages of cells to ensure that CCBE1 expression pattern was reproducible.



Pluripotency medium
CDM-BSA
FGF2 – 12 ng/ml
Activin A – 10 ng/ml

FLyB medium
CDM-PVA
FGF2 – 20 ng/ml
Ly – 10 µM
BMP4 – 10 ng/ml

FB50 medium
CDM-PVA
FGF2 – 20 ng/ml
BMP4 – 50 ng/ml

Fig. 3.1.1 – **Differentiation method.** Cells were cultured in CDM-BSA supplemented with Activin A and FGF2. To obtain early mesoderm, cells were cultured in CDM-PVA supplemented with FLyB for 36h, and then with FB50 to obtain cardiac precursors.

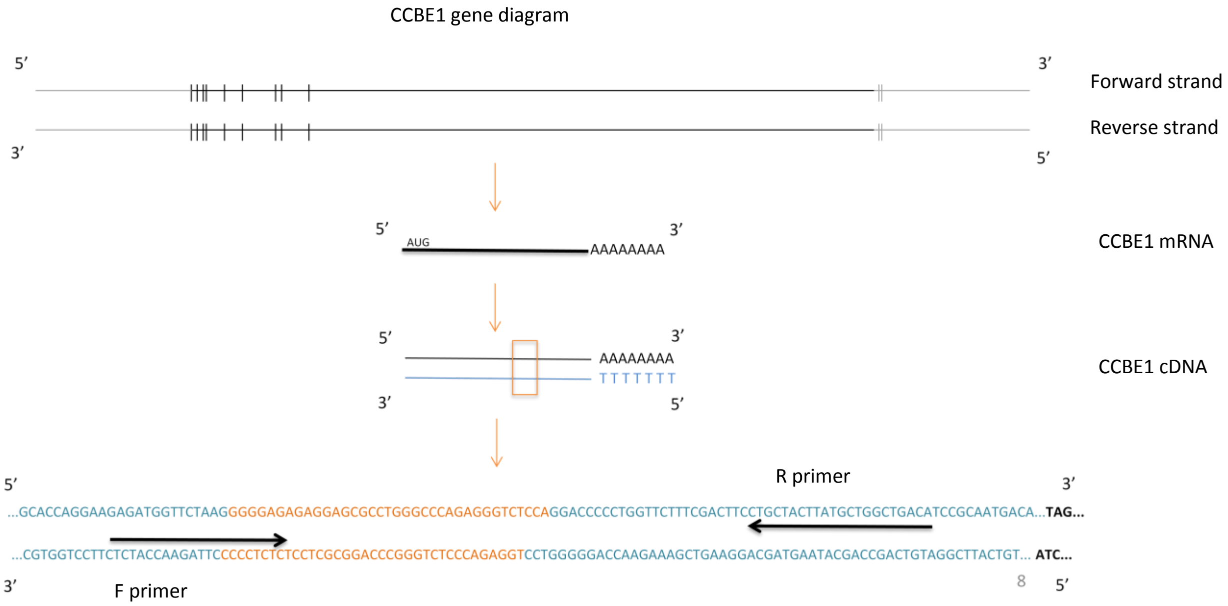


Fig. 3.1.2 - **Gene, transcript and cDNA levels.** At the gene level the 11 exons and the 10 introns are shown. The forward strand is transcribed in mRNA, which was collected and from which cDNA was synthesized. Primers were designed to avoid genomic or immature mRNA amplification and to assure the amplification of all splice variants. Forward primer was designed to anneal with a splice region located within the 9th and 10th exons. Reverse primer was designed 54 pb downstream the forward one, in the 11th exon.

Interestingly, CCBE1 is highly expressed in pluripotent cells its expression is downregulated in early mesodermal cells, and progressively starts to be as differentiation into cardiac precursors occurs. Its expression peaks at day D5 of differentiation and between D5 and D6, CCBE1 mRNA is clearly downregulated.

NANOG, a known pluripotent marker, was highly expressed in pluripotent cells and its expression decreases upon mesoderm differentiation. On the other hand, the early mesoderm genes BRA emerged transiently and peaked at 36h after FLyB treatment. MESP1, the earliest marker of cardiovascular progenitors (*Bondue et al*., 2012), and MESP2, from the same family of transcription factors, started to be expressed by D1 and both had a basal expression from day 3 and thenceforth. Lastly, the expression of the lateral plate markers ISL1, NKX2.5, KDR and GATA4 was increased from D2 of FLyB+FB50 differentiation, peaking between days 4 to 6.

To complement these transcriptional data, cells were fixed and immunostained at three different time points (D0, 36h and D5), with antibodies raised against CCBE1, NANOG, BRA and ISL1.

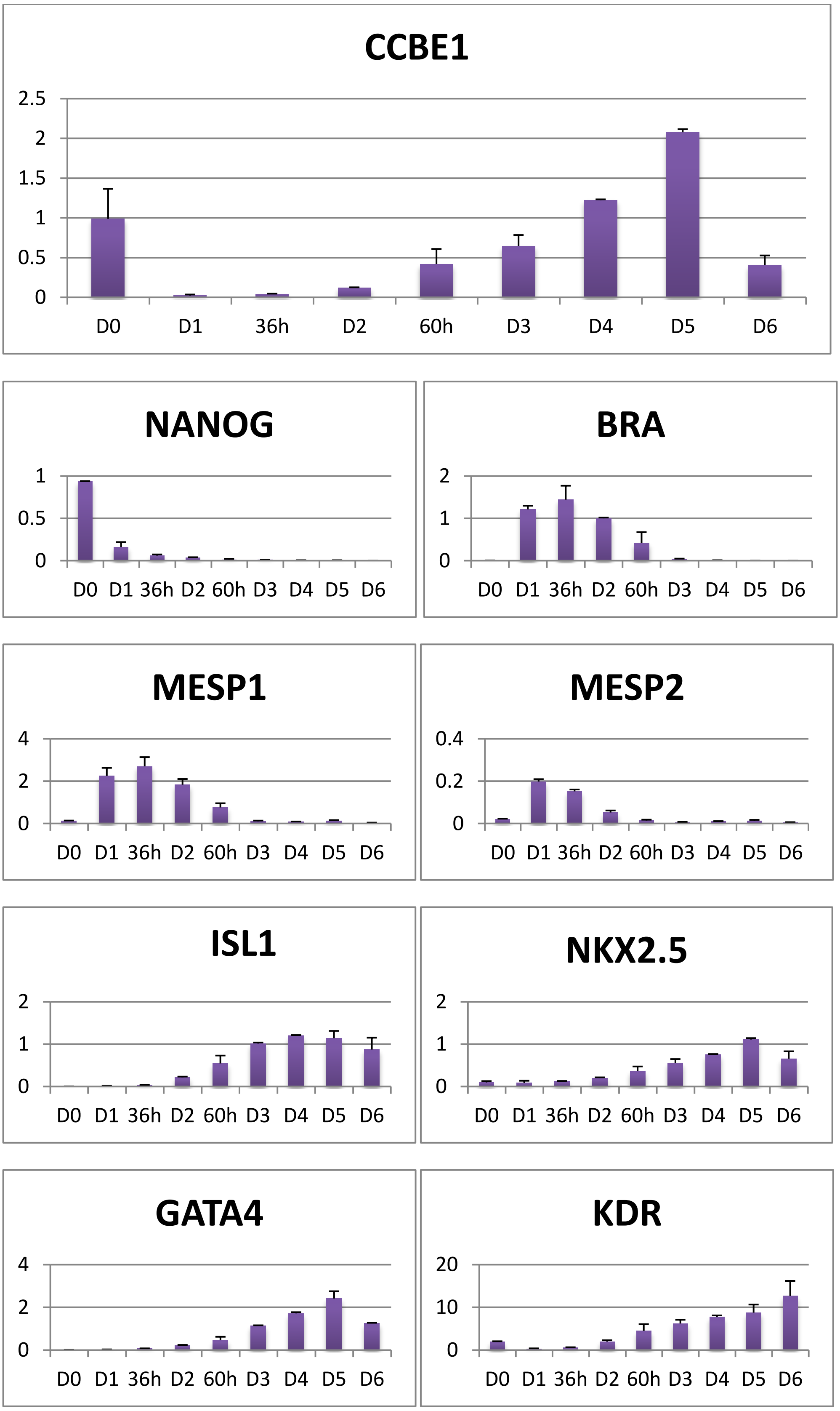


Fig. 3.1.3 – **qPCR graphics.** Cells were cultured as indicated and collected at D0 (pluripotency), and at several time points after differentiation has started – D1, 36h, D2, 60h, D3, D4, D5, D6.

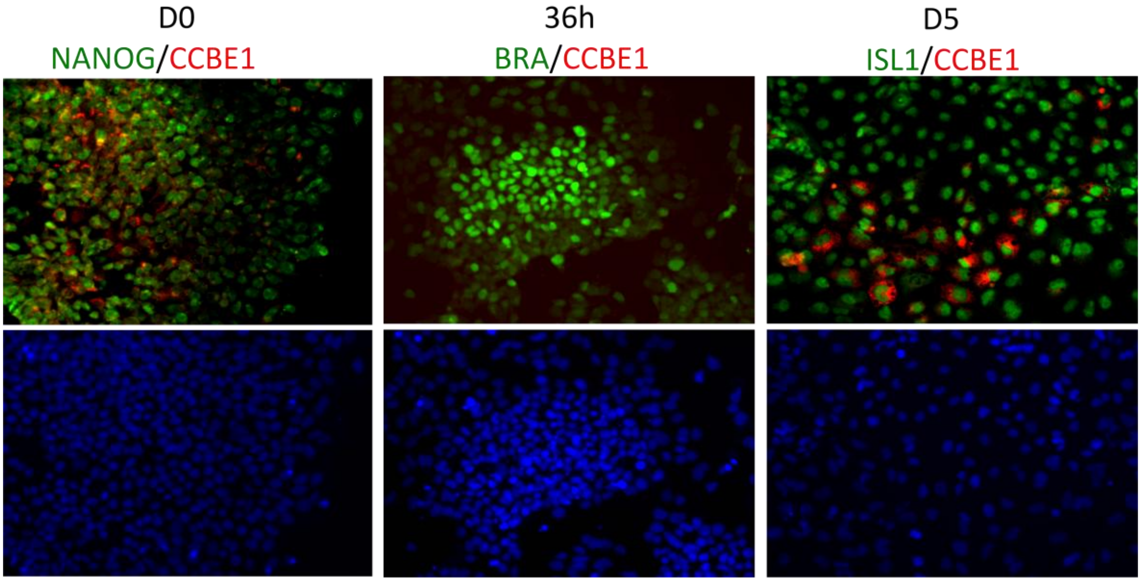


Fig. 3.1.4 - **Immunostaining performed in pluripotent (D0) and differentiated cells (36h, D5).** Samples were co-stained for NANOG (green) and CCBE1 (red) at D0; BRA (green) and CCBE1 at 36h of FLyB differentiation; and ISL1 (green) and CCBE1 at D5 of FLyB+FB50 differentiation. All samples were stained for DAPI, a nuclear marker.

The results show the protein data are in keeping with mRNA data. CCBE1 protein is highly expressed in pluripotent and cardiac precursor cells, and it is downregulated in early mesodermal cells. NANOG has a strong nuclear protein expression in pluripotent cells, with most cells expressing this gene. Some of these also express CCBE1, which expression is observed in the cytoplasm, as expected for this growth factor since CCBE1 is a secreted protein. For the same reason CCBE1, once as a growth factor, CCBE1 can have diffusible proprieties; it is not strongly expressed in the extracellular matrix. By 36h of FLyB treatment, immunostaining confirms CCBE1 is downregulated upon BRA upregulation. CCBE1 expression is then observed at day 5 of FLyB+FB50 treatment, where it colocalizes with ISL1. Curiously, CCBE1 is only expressed in ISL1 positive cells, but only a subset of ISL1+ cells co-express CCBE1.

These results suggest that CCBE1 expression is induced by a pluripotency gene in undifferentiated cells and by a cardiac gene in cardiac progenitor cells. It is also possible that BRA or another early mesoderm gene blocks CCBE1 expression.

3.2. CCBE1 Regulation

Understanding the regulatory mechanisms controlling CCBE1 expression at a DNA level would be highly interesting. . To test the idea that BRA has an important role in this process, possibly a repressive role, since when BRA is expressed CCBE1 is downregulated, a ChIP-seq database for BRA binding sites in 36h FLyB treated cells (Faial *et al.*, in preparation) was screened. Interestingly, several binding sites for BRA were observed in the CCBE1 promoter region. The three main different binding sites observed were: one located downstream of the transcribed region (3’ region), the second one located in the middle (mid) of the gene (middle region), and the third one located more upstream in the very beginning of the gene, just after the ATG (5’ region). Curiously, the last two regions (mid and 5’) are located within the second intron of CCBE1 (Fig. 3.2.1).

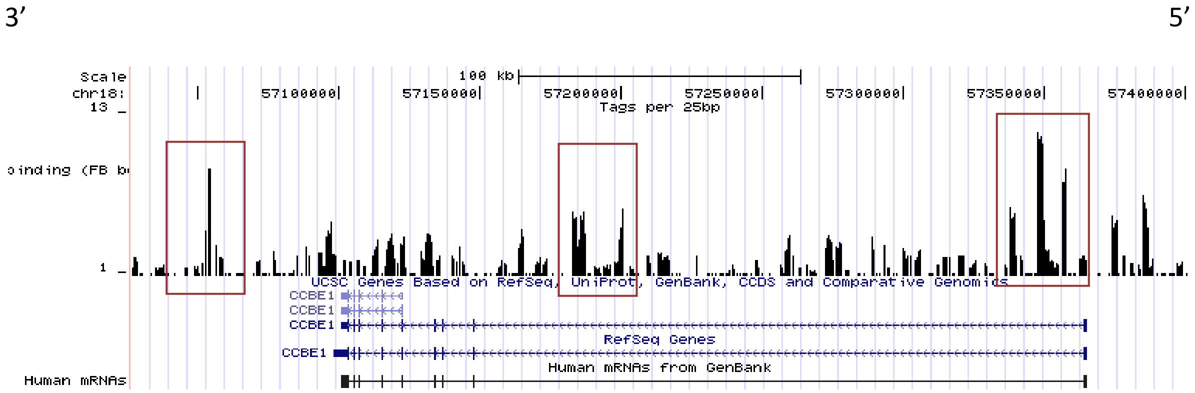


Fig. 3.2.1 – **BRA ChIP-seq after 36h of FLyB treatment.** It is showed that there is BRA binding to three different regions of CCBE1 gene; two within the second intron, and the third in the 3’ downstream region.

These findings confirm that BRA binds and likely regulates CCBE1 expression, given that, CCBE1 expression is the lowest when BRA expression is the highest.

The obvious thought was that BRA has an inhibitory role over CCBE1. Nevertheless, there is no report about BRA having an inhibitory function. Thus, another hypothesis was formulated: BRA, rather than inhibiting, activates CCBE1, but only in the presence of a co-activator. BRA is expressed, albeit at lower levels, in both pluripotent and cardiac precursor cells (C_T(D0)= 31.59; C_T(60h)= 26.05), which would be in keeping with a requirement of BRA for CCBE1 induction. If the hypothesis raised is true, then BRA would need a partner in both contexts in order to induce CCBE1 expression. Such partners could be a pluripotency marker gene in the pluripotent context, such as NANOG, and a cardiac precursor gene, in cardiac progenitor cells, like ISL1.

Interestingly, there are reported binding sites for NANOG along the CCBE1 gene (Fig. 3.2.2). As such, to test the hypothesis that BRA and NANOG interact to induce CCBE1 expression, Chromatin Immunoprecitation (ChIP) was performed in pluripotency conditions using antibodies against NANOG and BRA.

In this report, three main different binding sites are observed, one located downstream the 3’ region, the second one located in the middle (mid) of the gene (middle region), and the third one in the CCBE1 beginning (5’ region).

Amongst some of the NANOG binding regions (consensus sequence *CATT*) there were several BRA consensus sequences (*CACACC*) and as such primers were designed around these regions (Fig. 3.2.3). Three primer pairs were tested, one for each region (5', mid and 3'), but only the primers designed for the middle and 3' regions had an optimal efficiency value, thus the 5' region primers were not used. After conducting the ChIP protocol in undifferentiated hESCs, PCR analysis using the specific primers for each region was preformed. The relative binding of the transcription factors cited was determined by comparison with the IgG negative control. The results clearly show that there is enrichment of both NANOG and BRA binding to the CCBE1 3' regulatory region, when compared to the control (Fig. 3.2.4). With regard to the middle regulatory region, the results suggest that there is no specific binding of either NANOG or BRA due to the high unspecific binding of the IgG control to this region (Fig. 3.2.4).

A candidate partner for BRA in cardiac progenitor cells is ISL1 because: 1. CCBE1 co-localises with ISL1 and 2. BRA and ISL1 have come up as interacting partners in a yeast two-hybrid screen (Bernardo et al. unpublished). It has previously been shown in the lab by immunocytochemistry analysis that BRA and ISL1 co-localise 60h after cardiac differentiation is initiated. Thus, to determine if BRA and ISL1 bind to CCBE1 promoter, cells were grown in the FLyB+FB50 condition for 60h and then cells were collected for ChIP.

ISL1 and BRA consensus regions, *CTAATG* and *CACACC*, respectively, were identified along the CCBE1 promoter region and three pairs of primers were designed so that the amplicon was within them (Fig. 3.2.3).

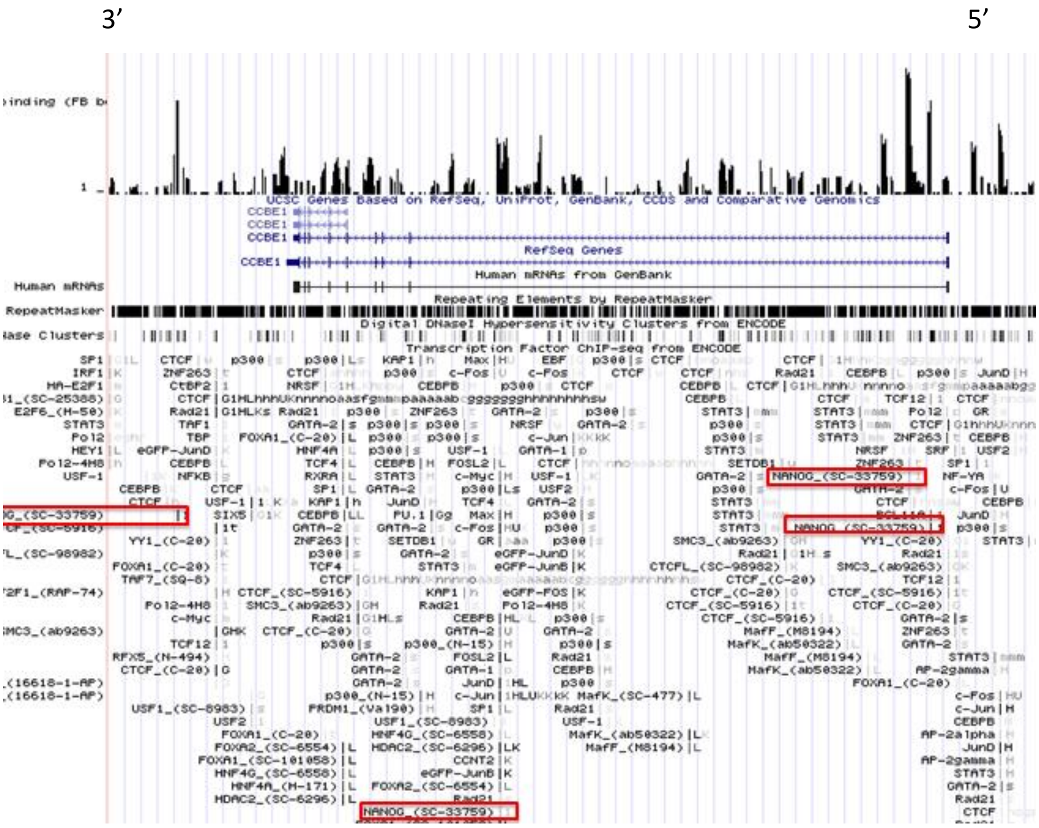


Fig. 3.2.2 – **NANOG ChIP-seq.** NANOG binds in and around the CCBE1 gene, in the same regions as BRA. It is seen many other transcription factors can bind to CCBE1 at this specific time point. Data from UCSC Genome Browser.

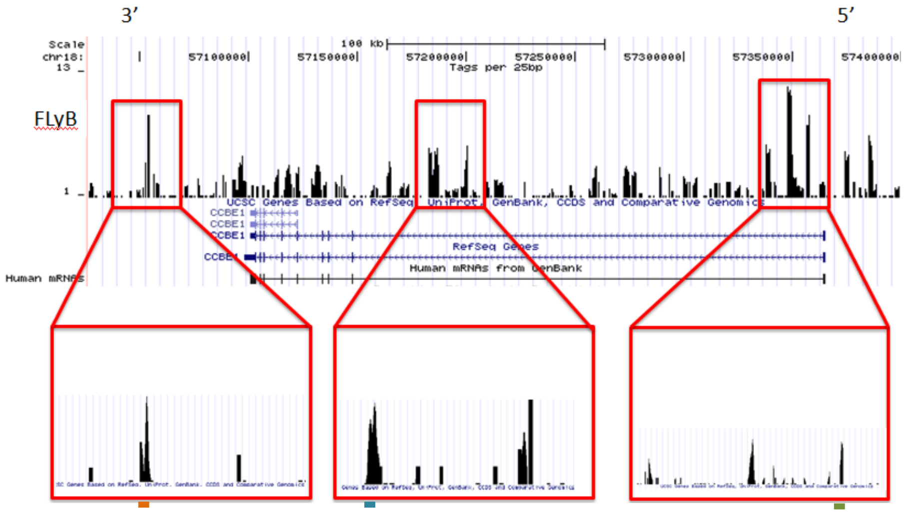


Fig. 3.2.3 – **Zoom in of the three main regulatory regions of the CCBE1 gene.** The small lines above the image represent the genomic regions where primers were designed, both for BRA/ NANOG ChIP and BRA/ ISL1 ChIP. The green one represents the two pairs of primers (one for which ChIP) designed for the 5’ regulatory region. The remaining pairs of primers designed for the middle and 3’ regulatory regions are represented in blue and orange, respectively. The sequences can be consulted in *Materials and Methods*.

The ChIP preformed on cells collected after 60h of FLyB+FB50 treatment showed binding of ISL1 and BRA to the CCBE1 mid regulatory region, but not to the than 3’ or 5’ regulatory regions (Fig. 3.2.5).

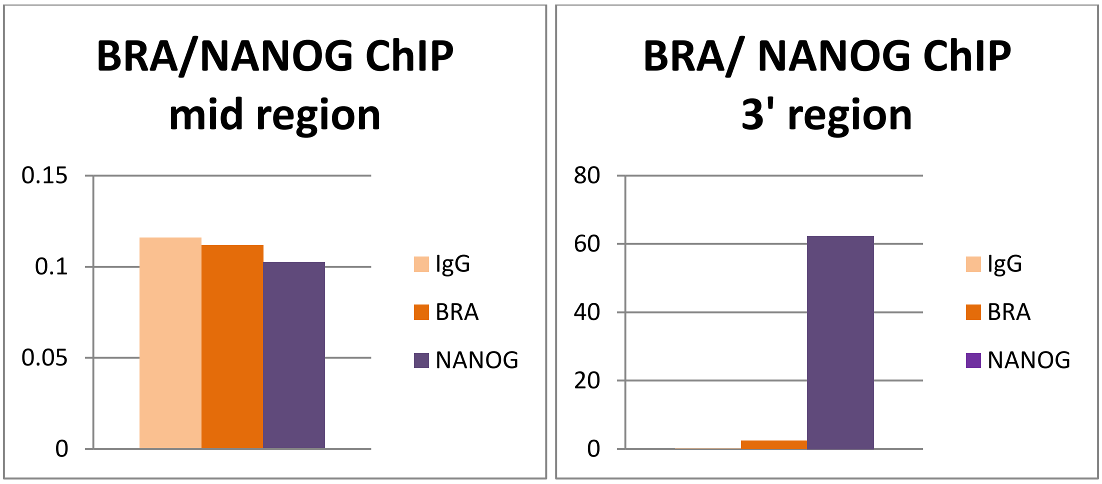


Fig. 3.2.4 – **BRA/ NANOG ChIP pluripotency.** The first graphic represents the NANOG and BRA binding to the CCBE1 middle regulatory region and the second one represents NANOG and BRA binding to the CCBE1 3’ regulatory region. Relative NANOG and BRA binding was determined by PCR and compared with the IgG control.

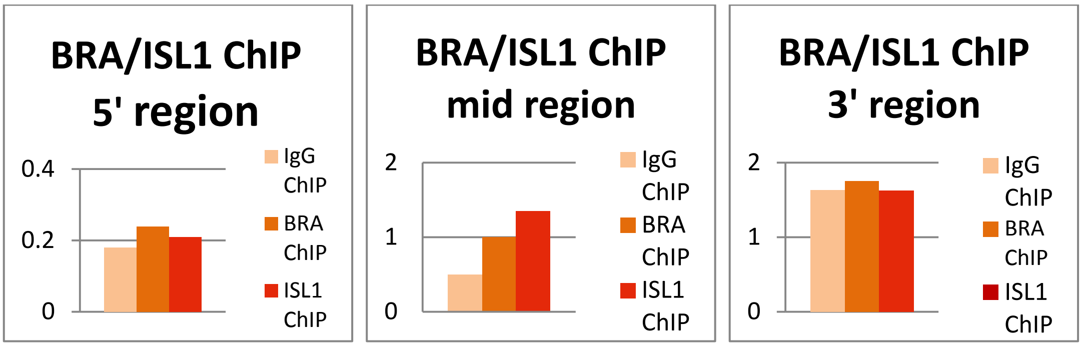


Fig. 3.2.5 – **BRA/ ISL1 ChIP 60h of differentiation.** All the graphics represent ISL1 and BRA binding to the CCBE1 regulatory regions, 5', middle and 3', respectively. Relative NANOG and BRA binding was determined by PCR and compared with the IgG control.

These data support the hypothesis that BRA binds to CCBE1 in different contexts where it interacts with different partners to activate CCBE1. It further suggests that NANOG and ISL1 act as co-activators of CCBE1 in pluripotency and cardiac progenitor cells, respectively. Another interesting finding is that the regulation of CCBE1 expression is done in different regulatory regions according to the differentiation state of the cell and the proteins involved.

To further determine the roles of BRA and ISL1 in the regulation of CCBE1 expression, knock down (KD) cell lines for these genes were generated using Short hairpin RNAs (Fig. 3.2.6).

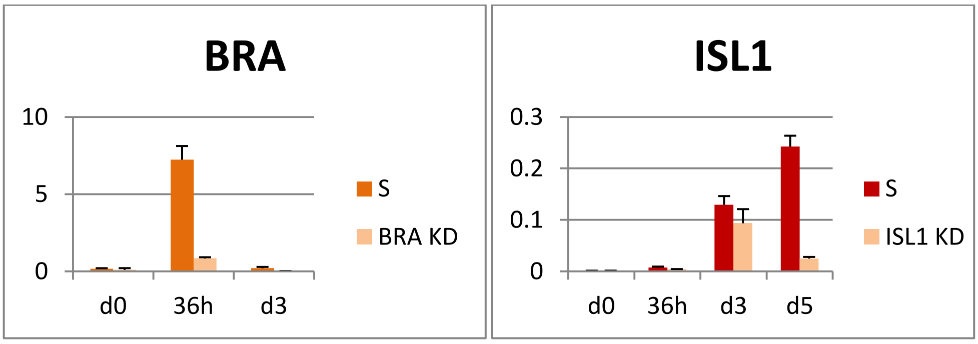


Fig. 3.2.6 – **BRA and ISL1 KD lines.** To establish these cell lines a pLKO.1-shRNA vector was integrated into the genome by Lipofectamine transfection. The best BRA KD and ISL1 KD clones were chosen, with KD percentages of 88.28456 at 36h and 89.69173 at d5, respectively.

Several KD lines were generated for each gene and the clone lines were culture alongside a control line (scrambled) in differentiation medium for a maximum of five days and their mRNA was extracted for semi-quantitative PCR analysis to evaluate the CCBE1 expression (Fig. 3.2.7 and Fig. 3.2.8). The expression of CCBE1 was affected in BRA KD cells at d0 and d3, but not at 36h, when the BRA knock down is strongly evident (Fig. 3.2.7). If BRA worked to repress CCBE1, as was hypothesised, CCBE1 expression should have been upregulated at 36h. Subsequently, this result does not support a repressing role of BRA over CCBE1 expression. Moreover, CCBE1 is clearly downregulated by d3 of FLyB+FB50 differentiation in the KD clones, which could be explained if BRA was an activator of CCBE1. Thus, the results are in keeping with the thesis that BRA activates CCBE1.

Interestingly, BRA KD cells exhibited upregulated levels of CCBE1 expression in pluripotency conditions. As for the ISL1 KD cells, these also exhibited upregulated levels of CCBE1 expression in pluripotency conditions. Furthermore, after three and five days of culture, there was no clear difference between the CCBE1 levels expressed in the scrambled and ISL1 KD cells (Fig. 4.2.8) suggesting that the low levels of ISL1 mRNA do not affect much CCBE1 expression. Despite that, the ISL1 knock down was not as evident at day three as it was at day five of differentiation (there is no significantly difference between ISL1 expression in the KD line and in the scrambled line at day 3). It is precisely around this time point (60h), it was hypothesized, that together with BRA, ISL1 might have an induction effect over CCBE1. So, if this supposition is true, this result would justify the similarity between CCBE1 expression levels in scrambled and ISL1 KD cells.

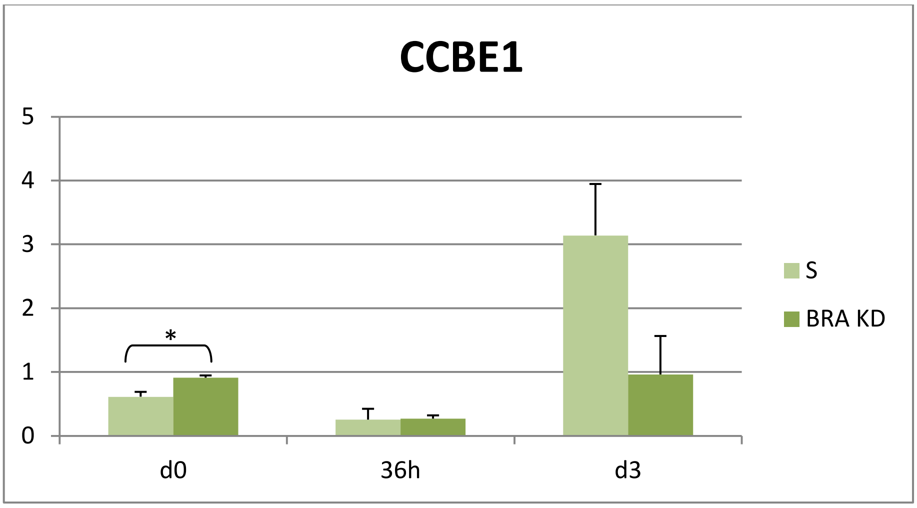


Fig. 3.2.7 – **CCBE1 expression upon BRA KD cells.** T-test was performed where *p ≤ 0.05.

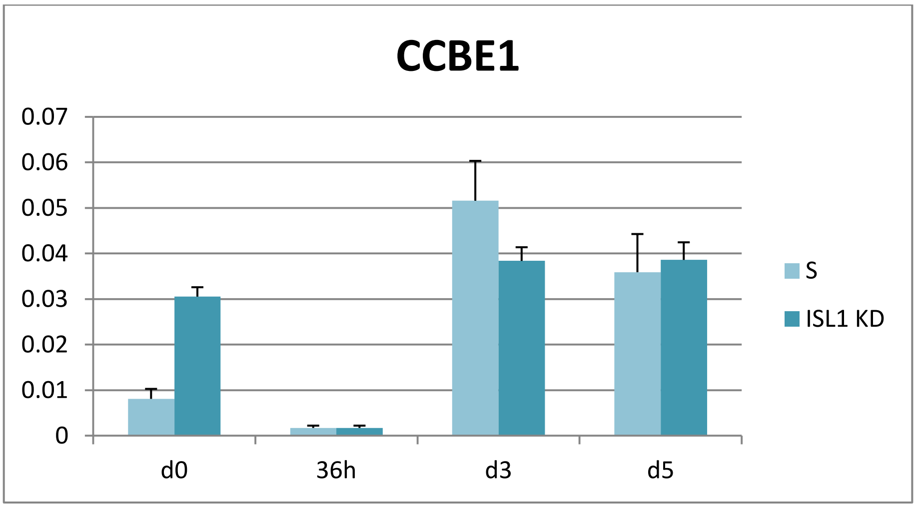


Fig. 3.2.8 – **CCBE1 expression upon ISL1 KD cells.**

3.3. CCBE1 Function

To better understand the role of CCBE1 in both pluripotent and cardiac progenitor cells several CCBE1 KD cell lines were generated. Five different shRNA sequences were used to generate five different lines. Of these, two of them (TRCN0000055474 (KD1) and TRCN0000055477 (KD2)) were most effective at knocking down CCBE1 expression (Fig. 3.3.1).

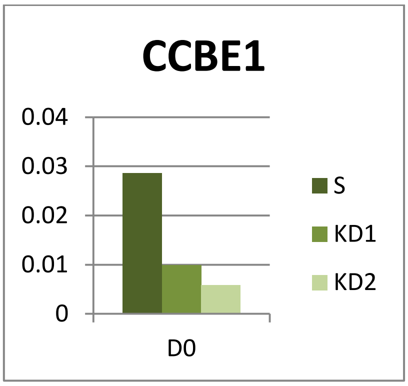


Fig. 3.3.1 – **CCBE1 expression upon CCBE1 KD cells – first collection**. Here are presented the two best KD lines collected at D0, one with 34.59391% of CCBE1 expression (KD1) and the second one with 20.42849%.

Cells that resulted from these transfections were expanded and later studied by real time PCR. Interestingly, the results show that the lack of CCBE1 expression affects the expression of key pluripotency genes (Fig. 3.3.2) and that of some important cardiac progenitor genes (Fig. 3.3.3). Maintaining the CCBE1 KD clones was particularly challenging, which was the first evidence that CCBE1 is required for keeping cells pluripotent. After repeating the experiment a couple of times, it was further confirmed that CCBE1 is required to dowregulate Nanog and OCT4 levels and to induce SOX2 expression (Fig. 3.3.2) suggesting that the pluripotency circuitry needs CCBE1 to be kept balanced. It was unsurprising therefore to observe that, after just a few passages (up to 4), the CCBE1 KD cells lose their KD properties, i.e. the Sh-RNA becomes silenced, or they have an adaptive disadvantage when compared with the wild type ones (Fig. 3.3.4).

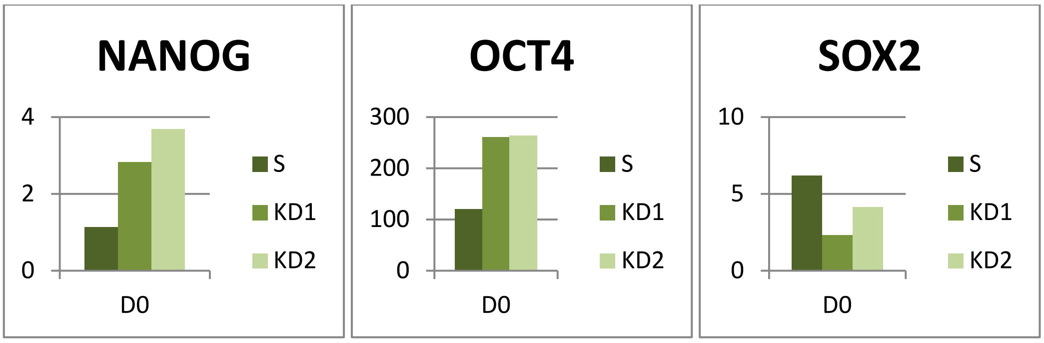


Fig. 3.3.2 – **NANOG, OCT4 and SOX2 expression upon CCBE1 KD cells** – first collection. These data report to the CCBE1 KD clones presented in the Fig. 4.3.1.

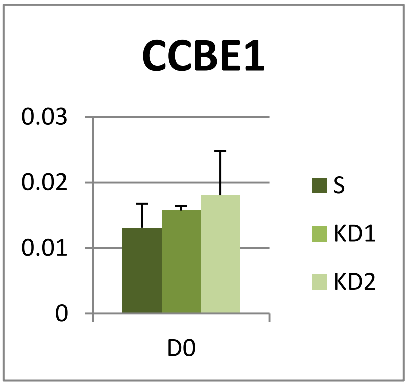


Fig. 3.3.4 – **CCBE1 expression upon CCBE1 KD cells – fourth passage**. The lost of CCBE1 knock down in both clones was demonstrated by a qPCR analysis.

Next, the CCBE1 KD clones were grown in cardiac differentiation medium and the expression of cardiac progenitor genes determined by real time qPCR (Fig. 3.3.2).

The majority of the genes tested was downregulated in both or just in one of the CCBE1 KD clones (clone KD2). The reduced levels of CCBE1 expression lead cells to dowregulate the expression of BRA, NKX2.5, GATA4 and KDR. However, MESP1 and ISL1 levels of expression did not appear to be affected by the CCBE1 knockdown and the results for MESP2 were antagonistic; MESP2 expression was upregulated in the KD1 cell line and downregulated in the other KD line.

These preliminary results lead to the conclusion that CCBE1 is quite important to potentiate cardiac differentiation, because when it is not present in physiological levels, expression of early mesoderm and cardiac precursor genes is downregulated. Moreover they suggest that there is a feedback loop between CCBE1 and BRA given that they appear to regulate each other (Figs. 3.2.7 and 4.3.3).

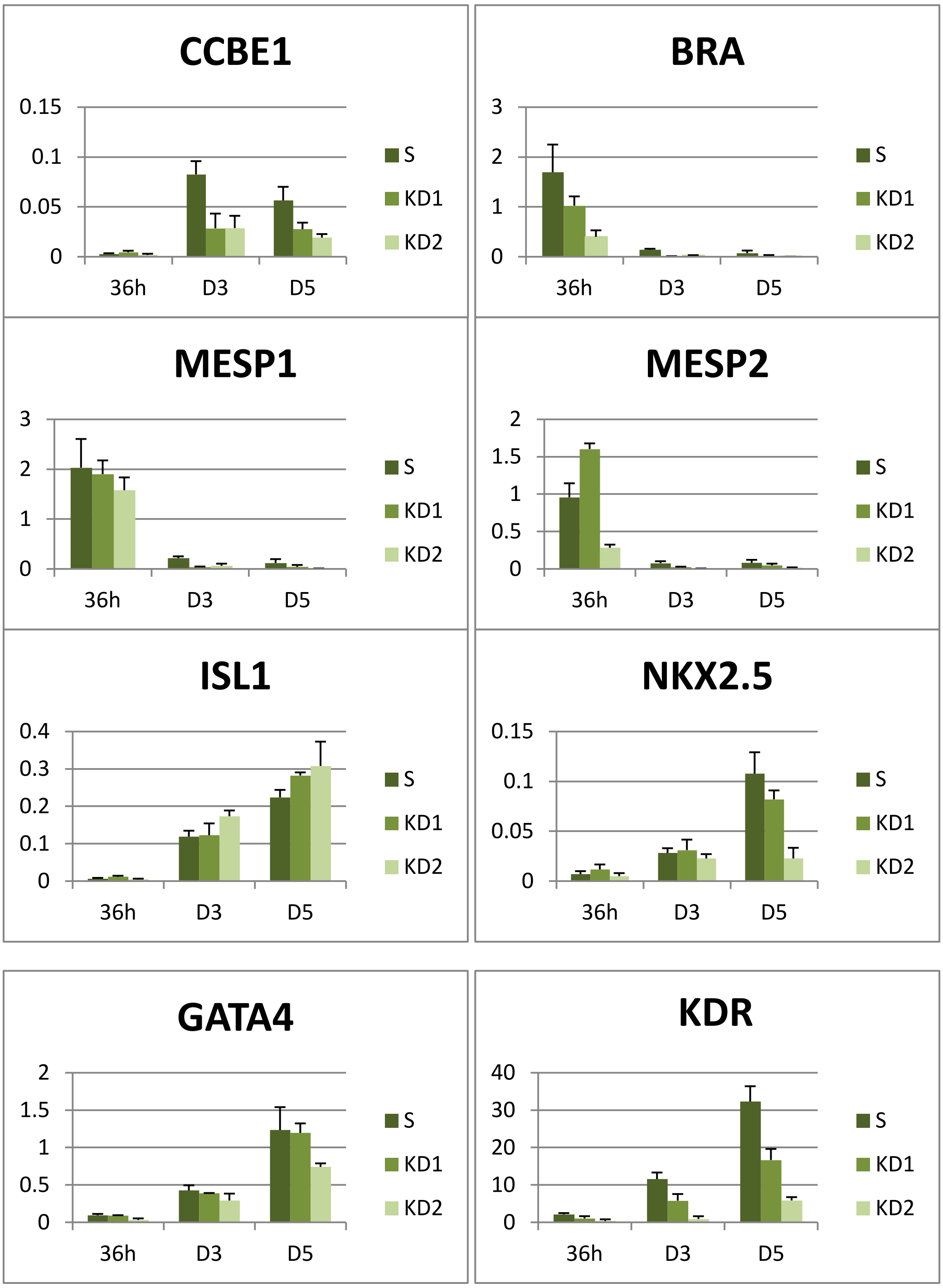


Fig. 3.3.2 – **qPCR graphics - CCBE1 KD cells.** Here are presented the two best KD clones grown during 36h, three and five days of differentiation.

4. DISCUSSION

Much of what is known about the molecular pathways that lead to human cardiovascular disorders has come from studying animal models, particularly genetically modified mouse models. Although these approaches have been useful and it is possible in many cases to translate genetic discoveries from humans to mice, important differences exist between rodent and human cells. Recent advances in stem cell and cardiovascular progenitor biology now raise the possibility of generating human models of human cardiovascular development, physiology, and disease (Musunuru *et al.*, 2010).

In this report, a reproducible protocol in which human ES cells are used as model of early development is described. This method was used in this study to determine the role of the secreted protein CCBE1 in cardiac progenitor cells.

The hESC line H9 was used for this purpose and a two fazed protocol followed in which cells were first pushed towards mesoderm and later towards cardiac progenitor cells. When exposed to the mesoderm medium (FLyB medium) the H9 cells suffer a cellular reorganization, which resembles the process of epithelial to mesenchymal transition (EMT) observed in early gastrulation. More specifically, the H9 hESCs pluripotent colonies, which are a compact monolayer and circular in shape, lose their compact organization and acquire a mesenchymal phenotype when they start to differentiate into the mesodermal lineage. Clearly, there is extracellular matrix (ECM) restructuration, which implies degradation of ECM components. One of such components could be the protein CCBE1, which is shown in this report to be present in pluripotent cells and to suffer drastic alterations of gene expression upon mesoderm induction.

The results presented here suggest that when cells start to differentiate towards mesoderm CCBE1 promoter is not being activated, or is being repressed and its product is degraded, since CCBE1 expression is downregulated.

One possible explanation for this phenomenon is based on BMP signalling, one of the components of the FLyB medium. It has been reported that BMP signalling could induce EMT and disrupt the basal membrane through induction of matrix metalloproteinases (MMP) expression, resulting in ECM degradation and, ultimately, in ingressive cell movement during gastrulation (Ohta *et al.*, 2010). As such, one could raise the hypothesis that BMP could lead to ECM degradation and as such to CCBE1 degradation.

Furthermore, it has been reported that loss of CCBE1 expression enhances migration of ovarian cancer cells in vitro and its overexpression in breast cancer cells inhibits cell migration (Barton *et al.*, 2010). Moreover, proliferation assays performed in our laboratory show that Ccbe1 knockout fibroblasts exhibit enhanced proliferation rates (Ana Perestelo, unpublished data). All in all, this suggests that CCBE1 is usually not expressed or downregulated during cell migration, and therefore it is unsurprising to see that upon early mesoderm differentiation in culture CCBE1 expression levels are low.

As opposed to what is observed upon mesoderm differentiation, CCBE1 is highly expressed in both pluripotent and cardiac precursors cells. Since its protein structure contains collagen and RGD domains, EGF-like and Ca^{2+} binding domains, it can be assumed that CCBE1 has an important role in cell adhesion and migration. However, the role of CCBE1 in either cell type is not known. As such, CCBE1 knockdown lines were generated to address its specific function (if any) in pluripotency maintenance and cardiac differentiation.

CCBE1 KD cells show downregulation of the cardiac marker genes NKX2.5, GATA4, but not of ISL1. BRA, an early mesoderm inducer, is also affected negatively by the CCBE1 KD. It can therefore be concluded that, when CCBE1 expression is bellow its normal levels, hESCs cannot differentiate properly into cardiac mesoderm. These preliminary results raise the hypothesis that CCBE1 has an important function in cardiac mesoderm specification and therefore in heart development, which is in keeping with what has been described in mouse embryos (Facucho-Oliveira *et al.,* 2011) and in mES cells (Tiago Justo and Facucho-Oliveira*,* unpublished data).

These results could be explained in two different ways. One is that the loss of CCBE1 during the *in vitro* EMT could preclude correct differentiation to occur or, on the other hand, since the CCBE1 KD impairs the cells in their pluripotent state, it could be that the CCBE1 KD compromises the cells' capacity to differentiate, due their 'ground' state being unstable.

CCBE1 signal is possibly transduced through Ras/MEK/ERK pathway once it has an EGF-like binding domain. This signalling pathway has a pro-differentiation effect and is antagonistic to mES cell self-renewal (Burdon *et al.*, 2002). Since CCBE1 is highly expressed in pluripotent cells, if indeed CCBE1 signal is transduced via the Ras/MEK/ERK pathway, one could speculate that the overall balance of CCBE1 and other pluripotency maintenance genes might determine the efficiency of embryonic stem cell self-renewal. To further understand the effect of CCBE1 in pluripotency and its role in the switch from the pluripotent state to a differentiated state, expression of pluripotent marker genes was analyzed in CCBE1 KD lines. Importantly, the successful CCBE1 KD lines lost their KD state in a few passages after being generated, which shows that the shRNA vector was rapidly silenced or excised, or in turn, CCBE1 KD cells could have died prematurely and been replaced by a minority of non-targeted puromycin resistant cells that lingered in the culture, suggesting that CCBE1 plays an important role in pluripotency.

NANOG, OCT4 and SOX2 expressions were affected by the CCBE1 knockdown; i.e. the first two transcription factors were upregulated and SOX2 was downregulated.

A recent report suggests distinct lineage specification roles for NANOG, OCT4, and SOX2 (Wang *et al*., 2012). Loss- and gain-of-function studies coupled with transcriptome profiling have shown that while SOX2 represses mesendoderm differentiation, high levels of OCT4 suppress ectoderm and promote mesendoderm specification in the presence of BMP4. Furthermore, it has been reported that NANOG inhibits ectoderm differentiation (Wang *et al*., 2012) and promotes mesoderm differentiation (Yu *et al.*, 2011). Overall, both KD and OE data strongly suggest that, in hESCs, NANOG, OCT4, and SOX2 function as differentiation repressors.

Based on the results obtained it can be suggested the loss of CCBE1 expression in pluripotent cells difficult ectodermal cellular differentiation and potentiates mesendodermal differentiation by rising the expression of NANOG and OCT4 and dowregulating the expression of SOX2.

To further understand the role of CCBE1 in the pluripotent context, complementary experiments need to be performed. The preliminary results presented here suggest that CCBE1 depletion impairs their ability to proper differentiate into cardiac mesoderm, but nothing is known about what CCBE1 depletion would cause upon ectoderm and endoderm differentiation. It has been reported that transcription factors required for establishment and maintenance of the pluripotent also promote signalling that drives progression from this state (Wray *et al.*, 2010). Interestingly though when Fgf4, whose expression is regulated by Sox2 and Oct4, is depleted in pluripotent mES cells they lose their ability to differentiate into mesoderm and neuroectoderm. Curiously, this factor is also important in cardiac development (Brand, 2003) and uses the same kind of receptor as CCBE1 probably does (tyrosine kinase receptor).

In sum, the results obtained from the analysis of the CCBE1 KD clones reveal that CCBE1 protein plays an essential role in mesoderm specification, regulating positively both early and later mesoderm markers. This is in keeping with data from mouse and mESCs (Facucho-Oliveira, data not shown). The data which suggest that CCBE1 has also an important function in keeping pluripotent cells balanced is novel and further experiments need to be performed in order to confirm the preliminary results present here.

Another important question raised in this study relates to CCBE1 expression pattern and its regulation. CCBE1 is expressed in two completely different cellular contexts: in pluripotency and in cardiac mesoderm, while being absent from early mesoderm cells. The expression pattern of CCBE1 was determined by mRNA analysis and validated by accessing its protein expression by immunocytochemistry. Therefore, different transcription factors are activating CCBE1 in these two different states or a repression signal is being produced in early mesoderm, the intermediate state between the pluripotent state and the cardiac mesoderm state. This hypothesis was tested using two different approaches: chromatin immunoprecipitation (ChIP) and knocking down the candidate genes thought to be CCBE1 regulators.

Three transcription factors were chosen as candidate CCBE1 regulators: NANOG for the pluripotent cellular context; BRA for the early mesoderm context; and ISL1 for the cardiac mesoderm differentiation context. NANOG was chosen because ChIP-seq data published on the UCSC Genome Browser shows binding of this transcription factor to CCBE1 gene. Similarly, BRA was chosen because when it is expressed CCBE1 is downregulated and also because ChIP-seq data of early mesoderm cells show that BRA binds to CCBE1 (Faial et al., manuscript in preparation). Lastly, ISL1 was chosen because it is a crucial gene in cardiac development which expression co-localises with that of CCBE1. This study revealed that BRA and NANOG bind to the 3’ regulatory region of CCBE1 in pluripotency; while BRA and ISL1 bind to the CCBE1 middle region in cardiac progenitor cells. Based on these data it is possible to propose that CCBE1 is differentially regulated according to the cell state. These two regulatory mechanisms are likely to depend on BRA interacting with NANOG or ISL1 to bind to two different regions of the CCBE1 gene. Further analysis needs however to be performed to validate these preliminary data.

This study went a little step further in order to determine if indeed BRA, ISL1 and NANOG regulate CCBE1 by generating and analysing BRA and ISL1 KD lines. As for the BRA KD, results show that CCBE1 expression is unchanged at 36h of FLyB differentiation. This result suggests that BRA does not have a repressive effect over CCBE1. Interestingly though, at D3 of differentiation, when cardiac progenitor specification takes place along this differentiation protocol, it is evident that CCBE1 is downregulated in the BRA KD line. Given that BRA expression level is relatively low at this time point, this result could be indirectly related with the low BRA expression levels at 36h, which could prevent BRA downstream genes from being activated that in turn activate CCBE1. Alternatively, the lack of BRA could directly prevent the induction of CCBE1 in cardiac progenitor cells by not being present and thus not binding to activate CCBE1.

With regard to the pluripotent state, CCBE1 mRNA expression levels were higher both in BRA and in ISL1 KD lines. There is a report showing that BRA has a role in pluripotency by upregulating the expression of NANOG (Belmonte *et al.*, 2006). Thus, with lower levels of BRA expression, those cells would easily differentiate, which could explain the CCBE1 higher expression level. However, as for the ISL1 KD clone, CCBE1 result remains unenlightened. Thus, a more efficient ISL1 KD line need to be generated, to present lower expression levels at day three of differentiation and subsequently the CCBE1 data from this clone would be more consistent and reliable.

As for the ISL1 KD line results, they reveal that CCBE1 mRNA levels are not significantly affected during cardiac differentiation. This result suggests that ISL1 does not activate CCBE1 despite there being binding of ISL1 to CCBE1 gene. As such, it appears that neither ISL1 nor CCBE1 have acted as activators of one another. In fact, although mCcbe1 transcripts are present in both the FHF and the SHF, higher levels of CCBE1 are expressed by the FHF progenitors, which are ISL1^{low}, in contrast with the SHF, characterized by high levels of ISL1 transcripts (Facucho-Oliveira *et al.*, 2011).

All these experiments overlooking CCBE1 characterization, regulation and function in hESCs, confirm that CCBE1 plays not only an important role in cardiac specification but also in pluripotency maintenance. HESCs hold huge promise in modern regenerative medicine and as a model for studying early heart development.

Concluding, understanding heart development in molecular and cellular terms might have an impact on clinical management of congenital malformations and might help to efficiently generate stem cell-derived myocardial tissue for substitutive therapies. Here a novel gene, CCBE1, involved in cardiac progenitor specification was studied and showed to be an important early mesoderm and cardiac mesoderm regulator.

5. FUTURE PERSPECTIVES

With this study many questions about CCBE1 expression in hES cells were enlightened, however many others remain up in the air.

To complement the KD data and elucidate CCBE1 function both in pluripotency and along cardiac differentiation, three further studies should be performed: the study of the CCBE1 knock out, a CCBE1 over expression study and the analysis of CCBE1 signalling. For the first two studies vectors will need to be generated in order to disrupt CCBE1 expression, both of them should be inducible to distinguish between CCBE1 function in pluripotency and in cardiac progenitors. As for the third study, a purified recombinant protein will need to be generated and added to the differentiation medium. Moreover, it would also be beneficial to perform an experiment in which the knock down would be inducible, so that at different stages of the differentiation process CCBE1 expression could be switched off, to exclude any influence of instability from pluripotent cells.

Another important approach to study CCBE1 function in pluripotency is to stimulate differentiation in additional hESCs lineages with wild type and modified CCBE1 expression, followed by characterization of these.

An important topic that was not addresses in this work was the recognition of the specific receptor and signalling pathway of this growth factor. It is paramount to identify the receptor (or receptors, once it has several domains) to which CCBE1 bind, the proteins involved in the signalling cascade(s) induced by CCBE1 and the target genes which these signalling cascade induce and/or repress, to further understand and mechanism of CCBE1 function.

Regarding CCBE1 regulation it was proposed that BRA together with NANOG or ISL1 has an induction action. To be sure BRA forms a dimer with NANOG or ISL1, an immunoprecipitation assay needs to be performed. Subsequently, it needs to be proven that these dimers have the capacity to activate the CCBE1 promoter, even if preliminary ISL1 KD data do not support part of this hypothesis. Thus, other transcription factors should be tested, taking into account ChIP-seq data bases.

Finally, an attractive result is that BRA changes its position between 3 prime and middle regions as cells differentiate, so it would be interesting to perform Chromatin immunoprecipitation for time points between D0 and 60h hours and before 60h.

6. REFERENCES

Alders, M. *et al.* Mutations in CCBE1 cause generalized lymph vessel dysplasia in humans. *Nat Gene*t 12, 1272-1275 (2009).

Barton, C. Collagen and calcium-binding EGF domains 1 is frequently inactivated in ovarian cancer by aberrant promoter hypermethylation and modulates cell migration and survival. *British Journal of Cancer* 102, 87 – 96 (2010).

Beckingham, K. *et al.* Calcium-binding proteins and development. *BioMetals* 11, 359-373 (1998).

Belmonte, JC. *et al.* Maintenance of embryonic stem cell pluripotency by Nanog-mediated reversal of mesoderm specification. *Nature Clin Pract Cardiovasc Med.* 3, S114-S122 (2006).

Bento, M. *et al*. Identification of differentially expressed genes in the heart precursor cells of the chick embryo. *Gene Expr Patterns* 11, 437-447 (2011).

Bernardo, A.S. *et al.* BRACHYURY and CDX2 mediate BMP-induced differentiation of human and mouse pluripotent stem cells into embryonic and extraembryonic lineages. *Cell Stem Cell* 9, 144–155 (2011).

Bondue, A. *et al.* Mesp1 : A Key Regulator of Cardiovascular Lineage Commitment. *Circ Res.* 107(12):1414-1427 (2010).

Bos, F.L., *et al*. 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 (2011).

Boullin, J. and Morgan, J. *et al*. The development of cardiac rhythm. *Heart* 91(7), 874-875 (2005).

Boyer, L.A. *et al.* Core transcriptional regulatory circuitry in human embryonic stem cells. *Cell* 122, 947–956 (2005).

Brand, T. Heart development: molecular insights into cardiac specification and early morphogenesis. *Developmental Biology* 258, 1–19 (2003).

Buckingham, M. *et al.* Building the mammalian heart from two sources of myocardial cells. *Nature Rev. Genet.* 6(11), 826-835 (2005).

Burdon, T. *et al.* Signalling, cell cycle and pluripotency in embryonic stem cells. *Trends in Cell Biol.* 12 (9), 432-438 (2002).

Chambers and Tomlinson. The transcriptional foundation of pluripotency. *Development* 136, 2311-2322 (2009).

Cheung, C. *et al*. Generation of human vascular smooth muscle subtypes provides insight into embryological origin–dependent disease susceptibility. *Nature Biotechnology* 30, 165-173 (2012).

Connell, F. *et a*l. Linkage and sequence analysis indicate that CCBE1 is mutated in recessively inherited generalized lymphatic dysplasia. *Hum Gene*t 127, 231-241 (2010).

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

Geens, M. *et al.* Human embryonic stem cell lines derived from single blastomeres of two 4-cell stage embryos. *Hum Reprod.* 24(11), 2709-2717 (2009).

Gilbert, S. Developmental Biology, 7th Edition, chap. 6, 11, 15. *Sinauer Associates, Inc.*, (2003).

Gschwind, A. Cell communication networks: epidermal growth factor receptor transactivation as the paradigm for interreceptor signal transmission. *Oncogen*e 20, 1594 – 1600 (2001).

Harvey, R. Patterning the vertebrate heart. *Nature Rev. Genet*. 3(7), 544-556 (2002).

Jackson, K. *et al.* Regeneration of ischemic cardiac muscle and vascular endothelium by adult stem cells. *J Clin Invest*. 107(11), 1395–1402 (2001).

Jaenisch, R. and Young, R. Stem Cells, the Molecular Circuitry of Pluripotency and Nuclear Reprogramming. *Cell* 132, 567–582 (2008).

Jura, M. Catalytic control in the EGF Receptor and its connection to general kinase regulatory mechanisms. *Mol Cell.* 42(1), 9–22 (2011).

Kehat, I. *et al*. Human embryonic stem cells can differentiate into myocytes with structural and functional properties of cardiomyocytes. *J Clin Invest.*108(3), 407-414 (2001).

Keller, R. Cell migration during gastrulation. *Curr Opin in Cell Biol* 17, 533–541 (2005).

Kouskoff *et al*. Sequential development of hematopoietic and cardiac mesoderm during embryonic stem cell differentiation. *PNAS* 102 (37), 13170–13175 (2005).

Lawson *et al.* Clonal analysis of epiblast fate during germ layer formation in the mouse embryo. *Development* 113, 891-911 (1991).

Lee, K.F. *et al.* Requirement for neuregulin receptor erbB2 in neural and cardiac development. Nature. 378(6555), 394-8 (1995).

Livak *et al*. Analysis of Relative Gene Expression Data Using RealTime Quantitative PCR and the $2^{-\Delta\Delta CT}$ Method. *Methods* 25(4), 402-408 (2001).

McLean, A.B., *et al*. Activin A efficiently specifies definitive endoderm from human embryonic stem cells only when phosphatidylinositol 3-kinase signaling is suppressed. *Stem Cells* 25, 29-38 (2007).

Musunuru *et al.* Stem Cell Models of Cardiac Development and Disease. *Annu. Rev. Cell Dev. Biol.* 26, 667-687 (2010).

Nanba, D. *et al.* Loss of HB-EGF in smooth muscle or endothelial cell lineages causes heart malformation. *Biochem Biophys Res Commun.* 350(2), 315-321 (2006).

Nichols, J. Introducing embryonic stem cells. *Curr Biol.* 11(13), R503-R505 (2001).

Nyati, M. *et al*. Integration of EGFR inhibitors with radiochemotherapy. *Nature Reviews Cancer* 6, 876–885 (2006).

Ohta, S. *et al.* The cessation of gastrulation. *Cell Adhesion & Migration* 4(3), 440-446 (2010).

Redkar, A. *et al*. Fate map of early avian cardiac progenitor cells. *Development* 128, 2269–2279 (2001).

Schultheiss et al. A role for bone morphogenetic proteins in the induction of cardiac myogenesis. *Genes Dev.* 11, 451-462 (1997).

Schultheiss and Alsan. Regulation of avian cardiogenesis by Fgf8 signaling. *Development* 129, 1935-1943 (2002).

Thomson, J. A. *et al*. Embryonic stem cell lines derived from human blastocysts. *Science* 282, 1145–1147 (1998).

Tavares, A.T. *et al*. Cerberus is a feedback inhibitor of Nodal asymmetric signaling in the chick embryo. *Development* 134 (11), 2051–2060 (2007).

Tonegawa, A. *et al*. Mesodermal subdivision along the mediolateral axis in chicken controlled by different concentrations of BMP-4. *Development* 124(10), 1975-1984 (1997).

Vallier, L. *et al*. Activin/Nodal and FGF pathways cooperate to maintain pluripotency of human embryonic stem cells. *J. Cell Sci.* 118, 4495–4509 (2005).

Vincent, D. and Buckingham, E. How to make a heart: the origin and regulation of cardiac progenitor cells. *Curr Top Dev Biol*. 90, 1-41 (2010).

Wang, Z. *et al*. Distinct Lineage Specification Roles for NANOG, OCT4, and SOX2 in Human Embryonic Stem Cells. *Cell Stem Cell* 10, 440–454 (2012).

Wells, A. EGF receptor. *Inter J of Biochem & Cell* Biol 31, 637-643 (1999).

Wolpert, L. Principles of Development, 3rd edition, chap. 3. *Oxford University Press* (2007).

Wouters, M. *et al*. Evolution of distinct EGF domains with specific functions. *Protein Science* 14, 1091–1103 (2005).

Wray, J. et al. The ground state of pluripotency. *Biochem Soc Trans* 38(4), 1027-1032 (2010).

Yáñez, M. *et al.* Calcium binding proteins. *Adv Exp Med Biol.* 740, 461-82 (2012).

Yang, X. *et al.* Cell Movement Patterns during Gastrulation in the Chick Are Controlled by Positive and Negative Chemotaxis Mediated by FGF4 and FGF8. *Developmental Cell* 3, 425–437 (2002).

Yu P., *et al*. FGF2 sustains NANOG and switches the outcome of BMP4-induced human embryonic stem cell differentiation. *Cell Stem Cell* 8(3):326-334 (2011).

Zhang, X. *et al*. Derivation of human embryonic stem cells from developing and arrested embryos. *Stem Cells.* 24(12), 2669-2676. (2006).

http://www.stanford.edu/class/history13/earlysciencelab/body/heartpages/heart.html (15-07-2012).

http://www.ensembl.org/Homo_sapiens/Transcript/Summary?db=core;g=ENSG00000183287;r=18:57102442-57364574;t=ENST00000439986 (08/06/2012).

http://www.who.int/mediacentre/factsheets/fs310/en/index.html (01/09/2012).